



TECHNISCHE
UNIVERSITÄT
DARMSTADT

Anwendung von Data Mining zu statistischen Auswertungen und Vorhersagen im Sport

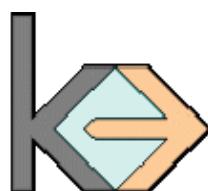
Diplomarbeit

von Vladimir Awerbuch

Betreuung

Prof. Dr. Johannes Fürnkranz

Fachbereich: Informatik



Fachgebiet: „Knowledge Engineering“

30. September 2009

Hiermit versichere ich, die vorliegende Diplomarbeit ohne Hilfe Dritter und nur mit den angegebenen Quellen und Hilfsmitteln angefertigt zu haben. Alle Stellen, die aus den Quellen entnommen wurden, sind als solche kenntlich gemacht worden. Diese Arbeit hat in dieser oder in ähnlicher Form noch keiner Prüfungsbehörde vorgelegen.

Darmstadt, 30. September 2009

Wladimir Awerbuch

Inhaltsverzeichnis

Kapitel 1: Einleitung

1.1. Überblick über Vorhersagesysteme	1
1.1.1 Das „A“ und „O“ eines Tippspiels	1
1.1.2 Zwei Beispiele der Tippspielgestaltung	3
1.1.3 ODDSET, das staatliche kommerzielle Projekt	6
1.1.4 Vorhersagen über einen Fußball Manager	8

Kapitel 2: Sammlung und Erstellung von Attributen

2.1. Datenquellen	11
2.1.1 „Fußball Studio“.....	11
2.1.2 Kicker Saisonhefte und sonstige Quellen	18
2.2. Beschreibung der Attribute	19
2.2.1 „Fußball Studio“-Attribute	20
2.2.2 Gruppenattribute	20
2.2.3 Manuell hinzugefügte Attribute	26
2.2.4 Aus anderen Attributen kombinierte oder neu erstellte Attribute.....	27
2.2.5 Differenzen	28

Kapitel 3: Fachliche Grundlagen

3.1. Beschreibung der verwendeten Attribute	30
3.1.1 Einleitung	30
3.1.2 Naive Bayes	31
3.1.3 JRip	34
3.1.4 J48	35
3.1.5 SVM	36
3.1.6 Ensembles	38
3.2. Erklärung der Verfahren zur Featureauswahl	39
3.2.1 Einleitung	39
3.2.2 Feature-Ranking	39
3.2.3 Wrapper	41

3.3. Erklärung der verschiedenen Evaluierungsarten	43
3.3.1 Cross-Validation	45
3.3.2 Supplied Test Set	45
Kapitel 4: Analyse der Attribute	
4.1. Einleitung und zu beantwortende Fragen	46
4.2 Analyse der Gruppenattribute	46
4.2.1 Versus	47
4.2.2 Form	48
4.2.3 AvgPlatz	51
4.2.4 Abstieg	52
4.3. Ergebnisüberblick vor der großen Analyse	54
4.4. Hauptanalyse der kompletten Attributmenge	56
4.4.1 Naive Bayes	57
4.4.2 JRip	61
4.4.3 J48	65
4.4.4 SVM	68
4.4.5 Random Forest	72
4.4.6 InfoGain	75
4.4.7 Fazit und Gesamtergebnisüberblick	78
4.5. Rückschlüsse aus der Analyse	83
4.5.1 Gruppenattribute unter Berücksichtigung der Analyse-Ergebnisse.....	84
4.5.2 Betrachtung der Ausprägungen	87
4.5.3 Diskussion über die schwachen Attribute	88
4.6. Auswahl der Datensätze	89
4.7. Vergleich der Datensätze	91

Kapitel 5: Praktische Tests an den Datensätzen

5.1. Einleitung	93
5.1.1 Auswahl von Tippstrategien	94
5.1.2 Erläuterung des Teilnehmerfeldes und Ausprägungsformats	96
5.2. Benchmarks: „Datensätze vs Tippstrategien“	97
5.2.1 Erläuterung der Punkteverteilung	97
5.2.2 Benchmarks: Betrachtung nach Saisons	98
5.2.3 Benchmarks: Betrachtung nach unterschiedl.Trainingssetlängen.....	111
5.2.4 Gesamtüberblick und Zusammenfassung	115
5.2.5 Auswahl der Datensätze für die Saisonvorhersage.....	118
5.3. Vorhersage der Saisontabelle mittels selbsterstellter Datensätze	118
5.4. Vorhersage einer Saison anhand von großen Trainingssets	131
5.5. Qualitative Diskussion	135
5.5.1 Diskussion über Klassifizierer	135
5.5.2 Betrachtung der Attribute	145
5.5.3 Betrachtung der Datensätze und Tippstrategien	146
5.6. Vorhersage der aktuellen Saison	148
Fazit	150
Literaturquellen	153
Anhang	155

Danksagung

Hiermit würde ich mich gerne bei Professor Dr. Johannes Fürnkranz bedanken, der mir ermöglicht hat, eins der für mich spannendsten Themenkomplexe überhaupt zu meiner Diplomarbeit zu machen und mich während meiner Diplomarbeit bestens betreute.

Ein besonderer Dank gilt meinen Eltern und meiner Frau Olga für die große moralische Unterstützung während meines gesamten Studiums und der Diplomarbeit.

Ebenso möchte ich mich bei meinen Freunden und Bekannten (insbesondere Manuel Ladebeck und Gabriele Olbricht) für das Korrekturlesen bedanken.

Zu guter Letzt würde ich mich gerne bei Gerald Köhler von Brightfuture für hilfreiche Informationen zur Fußball Manager-Matchprognose, sowie Holger Mallmann für die Freigabe der Exportdaten seines Programms „Fußball Studio“ bedanken.

Vorwort

„Das Wesentliche des Fußballs ist die Unvorhersehbarkeit des Spieldurchgangs“ (Joseph Blatter, FIFA-Präsident [Loy08])

„Das Reflex, das Unerklärliche erklären zu wollen, ist so alt wie das Denken (Adrian Kreye, Zeitungsjournalist [Loy08])

Tipps und Vorhersagen sind in der Sportwelt sehr beliebt. Überall versuchen fachliche oder auch selbsternannte Experten durch Vorhersagen Ihre Meinung zu dem aktuellem Sportgeschehen kundzutun. Genaue Vorhersagen und Tipps werden sehr oft belohnt, entweder „fachlich“ durch größere Anerkennung des Expertenwissens oder auch finanziell, durch beispielsweise einen Gewinn bei ODDSET. **[Oddset]**

Es war für mich schon immer ein Thema vom sehr großen Interesse Fußballspiele sowohl analytisch zu betrachten, als auch ebendiese möglichst genau vorherzusagen. Um diesem Interesse nachzugehen und durch mein eigenes seit Jahren durchgeführtes Tippspiel **[Tipp]** zusätzlich motiviert, habe ich mich zu dieser Diplomarbeit entschlossen. Eine Motivation dazu waren auch folgende Fragestellungen:

- Wie genau kann man Spiele mittels Sammlung vieler Daten aus der Vergangenheit vorhersagen?
- Ist ein Datensatz bzw. eine Attributensammlung mit großem Fußballwissen einem Fußballlaien mit einer guten Tippstrategie überlegen bzw. kann ein Computer gefüttert mit vielen Daten genauere Vorhersagen treffen als ein Fußballfan der „mit dem Herzen fühlt“?
- Welche Daten tragen mehr und welche weniger zu einer guten Vorhersage bei?
- Wie genau sind die Tabellen, die man mit Hilfe eines Computers mit einer größeren Datensammlung, vorhersagen kann?
- Welcher der verwendeten Klassifizierer sagt am besten voraus? Welcher ist am realistischsten und am menschlichsten?

Im Mittelpunkt der Diplomarbeit steht aber die Frage: „Kann ein größerer Datensatz und die Anwendung von Data Mining-Verfahren zu einer genaueren Vorhersage der Tendenzen einer Sportart verhelfen?“ Als betrachtete Sportart wurde über die ganze Diplomarbeit hindurch Fußball gewählt und als Analyseliga eine der besten und fairen Ligen der Welt, die erste deutsche Bundesliga.

Zum Schluss meines Vorwörteres erläutere ich noch kurz meine Erwartungen an die Diplomarbeit. Es wird nicht möglich sein als Ergebnis dieser Diplomarbeit etwas Exaktes zu bekommen, dazu ist Fußball, wie viele andere Sportarten, zu unberechenbar. Denn, wie bei vielen Sportarten, kann auch bei Fußball ein Schuss an den Innenpfosten oder ein (z.B. über eine Papierkugel) versprungener Ball über Sieg oder Niederlage entscheiden. Somit spielen Glück und Zufall, wie bei den meisten anderen Sportarten eine große Rolle. Auf der anderen Seite ist es natürlich nicht zu übersehen, dass gerade aus dem Zufallscharakter des Fußballs seine Spannung, seine manchmal kaum noch zu steigernde Dramatik und damit letztendlich seine Faszination bezieht **[Loy08]**.

Ich hoffe somit mit meiner Diplomarbeit einige sehr interessante Aussagen treffen zu können, meine oben gestellten Fragen zumindest in groben Ansätzen zu beantworten und diese Antworten mit Ergebnissen zu belegen.

Kapitelüberblick und geplante Vorgehensweise

Bevor ich aber beginne, gibt es noch einen kurzen Leitfaden zu meiner Diplomarbeit. Meine Diplomarbeit umfasst 5 Kapitel. In Kapiteln 1-3 meiner Arbeit werden alle Details und Begriffe erläutert, die zur Vorbereitung dieser Diplomarbeit notwendig waren. Im Kapitel 4 findet eine Analyse des Gesammelten statt, bevor im Kapitel 5 auf Basis der gesammelten Ergebnisse im vorherigen Kapitel praktiziert wird. Nun möchte ich aber noch zu jedem Kapitel eine etwas detailliertere Erläuterung abgeben:

Im 1. Kapitel werde ich einen kurzen Überblick über einige Vorhersagesysteme geben, sowie einen kurzen Exkurs in die Welt der Tippspiele und Wettbüros wagen. Hier werden sowohl einige mir bekannte Tippspiele als auch eine Alternative (Vorhersagen mittels eines Fußball Managers) vorgestellt.

2. Kapitel beschäftigt sich mit Attributen. In diesem Kapitel möchte ich sowohl alle verwendeten Attribute erklären, als auch die Quellen aufzeigen aus denen diese Attribute „gewonnen“ wurden. Auch wird in diesem Kapitel etwas genauer erklärt, wie diese Attribute aus ebendiesen Quellen extrahiert wurden.

Im 3. Kapitel werden dem Leser die notwendigen fachlichen Grundlagen vermittelt. In diesem Kapitel habe ich mir als Ziel gesetzt dem Leser die Funktionsweise von verwendeten Klassifizierern, Lerntechniken und Evaluierungsarten kurz zu erklären.

Im 4. Kapitel startet die Analyse der Attribute. Hier werden am Anfang Fragen aufgestellt, die ich im Verlauf des Kapitels zu beantworten versuche. Ebenso beschäftigt sich dieses Kapitel mit der Analyse der Trainingssets mit unterschiedlichen Größen mittels verschiedener Lerntechniken. Ferner wird versucht mittels Anwendung verschiedener Verfahren im Rahmen der Analyse schwache Attribute auszufiltern, starke Attribute vorzumerken und auf Basis dessen neue Datensätze zu erstellen. Ziel dieses Kapitels mit einer guten Attributmenge Datensätze zu erstellen auf denen man die praktischen Tests aus dem 5.Kapitel anwenden kann. Ein weiteres Ziel besteht darin, die neu erhaltenen Datensätze auf die Verbesserung gegenüber dem Komplett datensatz zu prüfen.

Im 5. Kapitel werden sowohl die im vorherigen Kapitel erhaltene Datensätze, als auch der komplette Anfangsdatensatz verschiedenen praktischen Tests unterzogen. In diesem Kapitel baue ich mittels Vorhersagen eine Endtabelle von einigen unterschiedlichen Saisons auf und vergleiche diese mit den Tabellen der Realität. Ebenso kommen diese Datensätze zu einem Einsatz gegen verschiedene „einfache“ Vorhersageverfahren. Ein Highlight dieses Kapitels sind sicherlich auch die Vorhersagen der neueren vollendeten Saisons der 1. Bundesliga auf Basis der Daten von allen bisherigen Bundesligasaisons (ab 1963) Das Kapitel abschließen wird eine qualitative Diskussion der Ergebnisse für verschiedene Klassifizierer und eine Vorhersage für die zum Zeitpunkt der Diplomerstellung unvollendete Saison der 1.Bundesliga (2009/2010) auf Basis der Rückschlüsse qualitativer Diskussion.

Zum Abschluß der Diplomarbeit wird ein Fazit gezogen und die Erweiterungsmöglichkeiten sowie alternative Herangehensweisen beschrieben. Hier werden einige Ratschläge erteilt „wie“ und „mit welchen weiteren Attributen“ man die Genauigkeit der Vorhersagen in eventuellen zukünftigen darauf aufbauenden Arbeiten noch weiter verbessern kann.

Kapitel 1: Einleitung

Wie bereits im Vorwort erwähnt, steht im Mittelpunkt der Diplomarbeit die Frage: „Kann die Anwendung von Data Mining-Verfahren zu einer genaueren Vorhersage der Tendenzen einer Sportart verhelfen?“ Um diesen Sachverhalt jedoch genauer untersuchen zu können, sollte hierfür jedoch zuallererst die Frage „Wie funktioniert überhaupt eine Vorhersage?“ beantwortet werden. Die hierfür notwendigen Informationen werden nachfolgend in einem einleitenden Überblick über verschiedene Vorhersagesysteme gegeben.

1.1 Überblick der Vorhersagesysteme

In diesem Unterkapitel finden Sie einen kleinen Überblick über einige der verschiedenen Vorhersagesysteme. Natürlich ist es nicht möglich alle vorhandenen Systeme in einem kleinen Unterkapitel zu erfassen aber ich werde hier versuchen einen kurzen, möglichst breitgefächerten, Einblick mittels einiger Beispiele zu geben.

Dabei sehe ich mir zuallererst ein generelles Standardmuster der Tippspiele an, gehe somit am Anfang auf Ihre Unterschiede und Gemeinsamkeiten ein. Als Zweites wird dann auf 2 Variationen des Tippspiels eingegangen, in denen zunächst ein einfaches (beispielhaftes) und danach ein komplizierteres (eigenes) Tippligaprojekt betrachtet werden. Danach gehe ich zu einer kurzen Betrachtung eines staatlichen kommerziellen Projekts über. Zum Abschluss des Kapitels sehe ich mir mit der Möglichkeit einer Matchprognose über einen Fußball Manager (FM2009-Matchprognose von Brightfuture/Electronic Arts) eine komplett alternative Vorhersagemöglichkeit an

1.1.1. Das „A“ und „O“ eines Tippspiels

Zuerst betrachte ich nun die Basis eines Tippspiels. Das Ziel eines Tippspiels ist es, Sportereignisse (in unserem Fall Fußballspiele der 1.Bundesliga) möglichst genau und vor allem genauer als seine Mitkonkurrenten vorherzusagen. Je näher man am Originaltipp ist, desto wahrscheinlicher ist der Gesamtsieg. Dabei unterscheiden sich die Tippspiele in mehreren Faktoren. Weiter unten folgt die Erläuterung einiger sehr wichtigen Faktoren in einem Tippspiel mit den am häufigsten vorkommenden Beispielen.

1. Art der zu tippenden Ereignisse

- die höchste Fußballliga eines jeweiligen Landes (z.B. 1. Bundesliga)
- eine (möglicherweise niedrigere) Liga/KO-Runde mit der sich aber alle Teilnehmenden identifizieren (z.B. einerseits Kreisliga A oder aber auch andererseits eine Welt-/Europameisterschaft)
- (ausgewählte/alle) Spiele aus mehreren Ligen gleichzeitig (z.B. alle Spiele der 1. und 2. Bundesliga)

2. Anzahl und Auswahlmöglichkeiten der Teilnehmer

- unbegrenzte Teilnehmerrunde (d.h. jeder kann mitmachen, ohne jegliche Beschränkungen)
- eine von (Tippspielleiter) ausgewählte Gruppen von Personen (z.B. alle Bürokollegen)
- feste Teilnehmeranzahl (z.B. 18 Tipper nicht mehr oder weniger, wenn weniger, dann kann die Liga nicht anfangen, nach dem 18. Teilnehmer wird die Anmeldung geschlossen)

3. Anmelde bzw. Mitmacheinsatz

- kein Einsatz
- (z.B. finanzieller) Einsatz

4. Anzahl der zu tippenden Spiele an einem Spieltag

- eine feste Zahl für alle (z.B. 9, ein Spieltag der Fußball-Bundesliga)
- verschiedene Anzahl je Tipper (wird im Beispiel des Kapitels 1.2.2 näher erläutert)

5. Was tippt man?

- genaues Ergebnis (nach 90 oder mehr Minuten)
- Tendenz (i.A. festgelegt als „0“ für Unentschieden, „1“ für Heimsieg, „2“ für Auswärtssieg)

6. Punkteverteilung fürs Tippen

Hier kann man sehr stark in die Tiefe gehen, Punkteverteilung kann es geben für:

- X Punkte für Spiel komplett erraten (Mit der Vorhersage das Ergebnis genau getroffen Bsp: Tipp 2:1-Ergebnis 2:1)
- Y für Tordifferenz und richtige Tendenz (Die Tordifferenz des Ergebnisses genau getroffen aber das Ergebnis selbst ist falsch, Bsp: Tipp 2:1-Ergebnis 4:3, in beiden Fällen Tordifferenz = +1)
- Z für Tendenz (Tendenz genau getroffen, Ergebnis komplett falsch, Bsp: Tipp 1:0 Ergebnis 2:0)
- W für richtig erratene Tore bei falscher Tendenz (Bsp: Tipp 1:0-Ergebnis 1:2)
- A Zusatzpunkte für einen Bonustipp (Bsp.: Tipp vor der Beginn der Bundesliga von allen Teilnehmern, „wer wird Meister der zu tippenden Liga“)

Ein kleiner Sonderfall hierbei ist das „falsch in der Höhe erratenes Unentschieden“ (Bsp: Tipp 1:1-Ergebnis 0:0). Dieses wird von vielen Tippspielen oft unterschiedlich gewertet, einige werten es als Tendenz (Argument: „Die Tordifferenz bei Unentschieden ist sowieso immer gleich, da gibt es nichts zu erraten“) und einige als Tordifferenz (Argument: „Die Tordifferenz ist in beiden Fällen 0, also damit automatisch erraten“)

Wie tief hier die Tippspiele ins Detail gehen hängt sehr oft vom Tippspielleiter, bzw. dem von ihm gewünschten Auswertungsaufwand ab. Die meisten Spiele begnügen sich mit den ersten 3 oberen Stichpunkten bei der Punkteverteilung. Es gibt aber auch detailliertere Tippspiele, die die Punkteverteilung für Heim, Auswärts und Unentschieden-Tipps unterscheiden. Zur Verdeutlichung der Unterschiede beider Extremformen werde ich im nächsten Unterkapitel 2 Beispiele geben. Generell gilt als Faustregel für die Punkteverteilung $X > Y > Z > W$. A hängt sehr von der Dauer der Liga ab, bei längerer Dauer ist A meist höher als X, bei einer kürzeren Dauer kann A auch niedriger als X sein.

Ebenso ein wichtiger Bestandteil der Punkteverteilung ist das Umgehen mit Nichttippern. Die meisten Ligen geben in diesem Fall 0 Punkte für diesen Spieltag, was dem Nichttipper automatisch die Chance auf den Gesamtgewinn nimmt. Es gibt aber auch andere Möglichkeiten, die den Tippspieler nicht so hart bestrafen, eine Möglichkeit hierfür wird später im Beispiel 1.2.2 gezeigt.

7. Art der Belohnung für den Sieger

- moralische Belohnung (Ruhm und Ehre)
- finanzielle Belohnung (z.B. die Summe der Anmeldeeesätze)
- sachliche Belohnung (z.B. der Sieger des Tippspiels erhält einen Pokal oder einen Preis)

Diese 7 Punkte sind mit die wichtigsten Details, die ein Tippspielleiter bei der Gründung eines Tippspiels überlegen muss. Natürlich gibt es auch weitere Details die zu überlegen sind und hier unerwähnt blieben. Da hier jedoch nur eine kurze Übersicht gezeigt werden sollte, gehe ich im nächsten Kapitel bereits zum praktischen Teil über.

1.1.2. Zwei Beispiele der Tippspiel-Gestaltung

In diesem Unterkapitel werden dem Leser 2 Beispiele präsentiert. Das erste Beispiel, soll ein einfaches und in der Realität sehr oft vorkommendes Standardbeispiel eines Tippspiels sein, das zweite Beispiel zeigt mein eigenes Projekt. [Tipp] Es zeigt ebenso, dass die Punkteverteilung auch etwas komplizierter sein kann, was aber auf den Spielspaß keinen unbedingten Einfluss haben muss.

Einfaches Beispiel eines Tippspielprojekts

Ein Forumbetreiber veranstaltet ein Tippspiel zur 1. Bundesliga, hierbei kann jeder der im Forum angemeldet ist (freie Anmeldung im Forum möglich) teilnehmen.

Folgende Regeln werden beachtet:

- Getippt werden jeweils alle Spiele eines Spieltags
- ein Nichttipper erhält 0 Punkte
- Punkteverteilung: Richtiger Tipp 3 Punkte, Richtige Tordifferenz 2 Punkte, Richtige Tendenz 1 Punkt
- Ein falsch in der Höhe getipptes Unentschieden zählt als richtige Tendenz (1 Punkt)
- Gewinner ist der Tipper mit den meisten Punkten am Saisonende, der Sieger erhält einen Pokal

Beispiel

Angenommen der Spieltag hätte 9 Spiele, z.B. ein Spieltag der 1.Bundesliga. Die Ergebnisse und die Tipps sind folgendermaßen ausgefallen (siehe Tabelle 1.1):

Spielnr./Teilnehmer	Ergebnis	Spieler A	Spieler B
Spiel 1	1:1	2:1=0 Punkte	0:0=1 Punkt
Spiel 2	0:3	2:1=0	3:0=0
Spiel 3	1:0	2:1=2	1:0=3
Spiel 4	0:2	2:1=0	2:3=1
Spiel 5	5:0	2:1=1	2:1=1
Spiel 6	1:1	2:1=0	1:1=3
Spiel 7	1:1	2:1=0	0:2=0
Spiel 8	3:3	2:1=0	0:1=0
Spiel 9	3:0	2:1=1	4:0=1
Gesamt:	---	4 Punkte	10 Punkte

Tabelle 1.1: Tippvergleich einfaches Tippspiel

Rangliste nach diesem Spieltag ist also (bei 2 Teilnehmern):

Platz 1: Spieler B = 10 Punkte

Platz 2: Spieler A = 4 Punkte

Eigenes komplizierteres Projekt-„Tippliga“

Dieses von mir, nach langer Arbeit und Feintuning, entwickelte Projekt, soll zeigen wie stark man die Regeln bei einem Tippspiel kombinieren kann.

In der „Tippliga“ gelten folgende Regeln:

- Es nehmen in einer Liga begrenzt 18 Spieler teil. Es gibt 2 Ligen, das heißt das Tippspiel ist für genau 36 Spieler ausgelegt. Gespielt wird nach Bundesliga-System, also 2x jeder gegen jeden, einmal „Heim“ und einmal „Auswärts“. „Heim“ bedeutet, dass der Spieler alle Spiele eines Spieltags tippt. „Auswärts“ bedeutet, dass der Tipper ein Spiel weniger tippt, er aber dieses ausgelassene Spiel auswählen kann.

3 Schritte der Punktberechnung

Schritt 1: Die Punktevergabe

Punktevergabe unterscheidet Punkte für Heim- und Auswärts- sowie Unentschiedentipps:

Für ein komplett erratenes Spiel gibt es

Bei Heimtipp 9 Punkte, Auswärts- und Unentschiedentipp 10 Punkte.

(Grund: In der Regel sind Heimsiege am einfachsten zu erraten, tippen auf Auswärtssiege und Unentschieden birgt ein höheres Risiko und sollte mehr belohnt werden)

Beispiel Tipp 2:1-Ergebnis 2:1->9 Punkte, Tipp 1:2-Ergebnis 1:2->10 Punkte

Für eine richtig erratene Tordifferenz gibt es

Bei Heimtipp 8 Punkte, bei Auswärtstipp 7 Punkte

(Grund: Die Heimsiege fallen meist höher aus als die oft knappen Auswärtssiege, daher Heimsieg schwerer zu erraten->mehr Punkte)

Beispiel Tipp 4:3-Ergebnis 2:1->8 Punkte, Tipp 1:2-Ergebnis 0:1->7 Punkte

Für eine richtig erratene Tendenz

Heimtipp 4 Punkte, Auswärtstipp 5 Punkte, Unentschieden 5 Punkte

(Grund: Wie bei „komplett erraten“)

Beispiel Tipp 2:1-Ergebnis 4:1->4 Punkte, Tipp 1:2-Ergebnis 0:2->5 Punkte

Für die richtig erratene Toranzahl bei falscher Tendenz

Heimtor=2 Punkte, Auswärtstor=1 Punkt

(Grund: Wie bei „Tordifferenz“)

Beispiel Tipp 2:1-Ergebnis 2:5->2 Punkte, Tipp 1:2-Ergebnis 3:2->1 Punkt

Schritt 2: Umwandlung der Punkte in Tore nach folgender Tabelle:

20-26 Punkte - 1 Tor

27-32 Punkte - 2 Tore

33-38 Punkte - 3 Tore

usw. in Fünfer-Schritten

Diese Umwandlung ermöglicht auch Unentschieden in der Liga und schafft einen gewissen Flair einer realen Fußballliga, indem die Resultate so wie in einer Bundesliga vorkommen. Dieses Flair kann, meinen Erfahrungen nach, die Motivation enorm steigern.

Schritt 3: Die erzielten Tore beider gegeneinander an diesem Spieltag spielenden Teilnehmer werden gegenübergestellt.

Der Sieger erhält 3 Punkte, der Verlierer 0, bei einem Unentschieden gibt es je einen Punkt (ebenfalls zur Auswertung siehe Bild 1.1).

Beispiel (siehe Tabelle 1.2)

Es ist der 1.Spieltag und es spielen Spieler A gegen Spieler B gegeneinander. Spieler A hat Heimrecht. Spieler B lässt als Auswärtsspiel-Spiel 4 aus.

Spielnr./Teilnehmer	Ergebnis	Spieler A	Spieler B
Spiel 1	1:1	2:1=1 Punkt	0:0= 5 Punkte
Spiel 2	0:3	2:1=0	3:0=0
Spiel 3	1:0	2:1=8	1:0=9
Spiel 4	0:2	2:1=0	X:X
Spiel 5	5:0	2:1=4	2:1=4
Spiel 6	1:1	2:1=1	1:1=10
Spiel 7	1:1	2:1=1	0:2=0
Spiel 8	3:3	2:1=0	0:1=0
Spiel 9	3:0	2:1=4	4:0=4
Gesamt:	---	19 Punkte=0 Tore	32 Punkte=2 Tore

Tabelle 1.2: Tippvergleich eigenes komplizierteres Projekt

Somit ist das Ergebnis des Spiels: Spieler A-Spieler B 0:2. Spieler B erhält 3 Punkte in der Tabelle und eine Tordifferenz 2:0 und Spieler B 0 Punkte und eine Tordifferenz von 0:2.

	5:1	2:3	2:1	3:0	2:2	4:0	1:1	1:1	2:3	↙
<u>ChristianH</u>	3:1 = 4	0:2 = 5	2:2 = 2	0:1 = 0	1:1 = 4	1:2 = 0	0:2 = 0	0:3 = 0	0:0 = 0	15 = 0
<u>Wladi</u>	3:2 = 4	2:1 = 2	3:1 = 4	0:2 = 0	2:0 = 2	1:2 = 0	2:1 = 1	2:0 = 0	1:1 = 0	13 = 0
<u>PferdImAnzug</u>	1:2 = 0	1:2 = 7	2:1 = 9	1:2 = 0	1:2 = 1	1:2 = 0	1:1 = 9	1:2 = 2	2:1 = 2	30 = 2
<u>Olva</u>	3:1 = 4	1:2 = 7	2:1 = 9	1:1 = 0	1:0 = 0	0:2 = 0	2:1 = 1	0:2 = 0	1:2 = 7	28 = 2
<u>JCT</u>	1:1 = 1	0:2 = 5	1:3 = 0	1:2 = 0	1:0 = 0	1:2 = 0	2:0 = 0	1:1 = 9	2:1 = 2	17 = 0
<u>Ella</u>	2:1 = 4	2:2 = 2	0:1 = 1	0:0 = 1	1:0 = 0	3:2 = 4	2:2 = 4	0:1 = 1	1:1 = 0	17 = 0
<u>Gaby</u>	2:3 = 0	1:3 = 5	2:3 = 2	0:3 = 0	1:2 = 1	0:3 = 0	1:1 = 9	0:2 = 0	2:1 = 2	19 = 0
<u>RealChrisOF</u>	2:0 = 4	1:1 = 0	1:2 = 0	1:1 = 0	2:0 = 2	1:2 = 0	1:0 = 2	0:2 = 0	2:2 = 2	10 = 0
<u>TCK</u>	2:1 = 4	2:1 = 2	2:1 = 9	2:1 = 4	1:2 = 1	2:1 = 4	2:1 = 1	1:2 = 2		27 = 2
<u>Florian</u>	1:1 = 1	0:3 = 5	1:1 = 1	1:2 = 0	1:2 = 1	1:2 = 0		1:2 = 2	1:1 = 0	10 = 0
<u>dgb</u>	2:1 = 4	0:1 = 7	1:1 = 1	1:1 = 0	1:2 = 1	1:2 = 0	1:1 = 9	1:2 = 2		24 = 1
<u>Nebi</u>	2:3 = 0	1:2 = 7	3:2 = 8	2:0 = 4	1:0 = 0	3:1 = 4		1:2 = 2	2:2 = 2	27 = 2
<u>Sport-Bild</u>										
<u>Alkeks</u>	1:2 = 0	2:1 = 2	3:1 = 4	1:0 = 4		2:0 = 4	2:1 = 1	3:1 = 1	2:3 = 10	26 = 1
<u>SleepyFox</u>										
<u>Masakari</u>	2:1 = 4	0:2 = 5	1:2 = 0	1:1 = 0	2:1 = 2	0:1 = 0	2:1 = 1	1:3 = 2		14 = 0
<u>Wolf</u>	1:0 = 4	0:1 = 7	1:2 = 0	0:1 = 0	0:1 = 0	0:1 = 0	1:1 = 9	0:1 = 1		21 = 1
<u>Jako</u>										

Ergebnisse der Spiele

Spiele der Tippliga			
#	Heim	Gast	Ergebnis
1	Jako	TCK	0:27 = 0:2

Bild 1.1: Auswertungsbildschirm im Tippliga-Adminbereich

Ein Nichttipper erhält bei diesem System 0 Punkte am entsprechenden Spieltag, damit es aber nicht zu einem Unentschieden verkommt, muss sein Gegner noch 20 Tipppunkte schaffen und damit 1 Tor erzielen um dieses Spiel zu gewinnen. Somit werden die Nichttipper hier nicht so stark bestraft, dass sie, wie bei anderen Tippspielen jegliche Chancen auf einen Gesamtsieg verlieren.

Eine weitere Besonderheit des Systems, ist dass ein Tipper mit etwas mehr Punkten nicht zwingend als Sieger vom Platz geht (z.B. 25-23 Punkte, 1:1 Tore). Dies kann man mit einem Spiel der Bundesliga vergleichen, wo eine Mannschaft überlegen war, jedoch verpasst hat, Ihre Chancen in Tore umzumünzen.

Der Sieger ist wer nach 34.Spieltag auf dem ersten Tabellenplatz steht und damit die meisten Punkte erzielt. Ferner sind beide Ligen auf 2 hintereinander geordneten Ebenen. Dies bedeutet, dass nur der Beste der 1.Liga Tippspielmeister werden kann und es in der 1.Liga 3 Absteiger gibt. In der 2.Liga steigen die ersten 3 Teilnehmer in die 1. auf. Zu gewinnen gibt es bei diesem Tippspiel nichts, der Sieger erhält also „nur“ Ruhm und Ehre.

1.1.3 ODDSET, das staatliche kommerzielle Projekt

(dieses Unterkapitel wurde auf Basis der Informationen auf dem offiziellen Oddset-Internetauftritt gestaltet [Oddset])

Die Sportwette ODDSET wird vom deutschen Lotto und Totoblock seit Februar 1999 angeboten. Damals wurde die ODDSET Kombi-Wette von der Staatlichen Lotterieverwaltung in Bayern eingeführt. Bereits ein Jahr später wurde ODDSET bundesweit angeboten. Mittlerweile zählt das Projekt zu den größten staatlichen Wettanbietern in Europa.



Bild 1.2: Oddset-Logo

Als einziges staatliches Wettangebot in Deutschland, setzt ODDSET sich als Aufgabe, den Teilnehmern eine offene, sichere und seriöse Möglichkeit zu bieten, Glücksspiel zu betreiben. Die Seriosität drückt sich in vielerlei Hinsichten aus. Ein großes Anzeichen hierfür ist sicherlich, dass im Gegensatz zu vielen anderen nichtstaatlichen Tipp- und Spielgemeinschaften bei ODDSET die Garantie eines seriösen und vor allem manipulationsfreien Sportwettenangebots (und nicht die Gewinnmaximierung) im Vordergrund steht. So klärt ODDSET die Interessenten in erster Linie über Risiken des Glückspiels auf, was in unter anderem an vielerlei Spielsuchtwarnungen bereits beim Internetauftritt der Sportwette deutlich sichtbar ist. Auch eine Aufklärung über die Krankheit „Spielsucht“ gehört dazu.

Nun aber zur Beschreibung des Angebots der staatlichen Sportwette. Bei ODDSET können diverse Ereignisse aus dem aktuellen Sportgeschehen (u.a. Fußball, Basketball, Eishockey, Formel1 uvm.) getippt werden. Die Quoten einer ODDSET-Wette sind dabei von vorneherein fixiert, so dass der Teilnehmer die Quote bereits vor der Abgabe des Wettscheins kennt. Ferner darf die Quote von ODDSET nicht mehr geändert werden, wenn sie bereits einmal festgesetzt wurde. Dies hat sowohl Vor- als auch Nachteile. Ein großer Vorteil ist natürlich die Sicherheit des Teilnehmenden, denn die Quoten sind bekannt und Manipulation daran damit ausgeschlossen. Die Nachteile drücken sich unter anderem sehr oft in Nichtberücksichtigung kurzfristiger Umstände aus. Von diesem Nachteil sind sowohl die Teilnehmer (wenn sich bei einem starken Verein, in letzter Sekunde einer oder mehrere gute Spieler verletzen), als auch der Veranstalter betroffen (wenn ein sehr starker Verein beispielsweise mit einer zweiten oder gar dritten Mannschaft anreist, weil das Spiel für ihn zweitrangig ist und die Schonung der Spieler Vorrang hat). Die Quoten werden ODDSET intern festgesetzt, dabei wird logischerweise einem Favoriten eine sehr niedrige Quote (meist im Bereich <1,5), einem meist schwer zu erratendem Unentschieden eine höhere und einem Außenseiter eine

sehr hohe Quote (Oft im Bereich 6-7.x) festgesetzt. Die Berechnung des Gewinns aus diesen Quoten wird etwas weiter unten erwähnt.

Die Wettarten

Generell lässt sich ODDSET in 2 Wettarten (mit vielen Unterwettarten) unterteilen. Diese heißen „Kombi-Wette“ und „TOP-Wette“.

1. Kombi-Wette

Bei der „Kombi-Wette“ darf man sich 3-10 (in Ausnahmefällen 2) Spiele aus einer Liste (siehe Bild 1.3) aussuchen und die Tendenz von diesen vorhersagen. Dabei steht auch hier „1“ für Sieg, „0“ für Unentschieden und „2“ für Niederlage. Nur wenn man alle gewetteten Spiele errät, erhält man den Gewinn. Die Quoten für jedes der erratenen Spiele multiplizieren sich dann mit dem Einsatz aufeinander.

Spielplan 22		Dienstag, 26.05.2009 - Montag, 01.06.2009			Stand: 26.05.2009, 15:44:22 Uhr		
		Bitte wählen Sie Ihre Voraussagen im Spielplan.			Filter: alles		
Abgabe-	Liga	Nr. Heim	Gast	Quote 1	Quote 0	Quote 2	Info
schluss		12.05.09 Madrid	ClevenAdvances	1,70	2,70	1,00	
Mi. 27.05.2009							
17:25	FIN	13 JJK Jyväskylä	IFK Mariehamn	2,25	2,80	2,25	i
17:55	FIN	14 FC Honka Espoo	FF Jaro	1,50	3,00	4,00	
20:40	CHAMP	15 FC Barcelona*	Manchester Utd. Z	2,35	2,75	2,35	i
20:40	CHAMP	16 Torsumme*	Barca / ManU	2,80	1,85	2,85	i
20:40	HCCHAM	17 FC Barcelona+1*	Manchester Utd.	1,35	3,30	4,85	
20:40	HCCHAM	18 FC Barcelona*	Manchester Utd+1	4,85	3,30	1,35	i
20:40	MLS	19 Columbus Crew	San Jose Earth.	1,50	3,00	4,00	i
20:40	COPLIB	20 FC Caracas (VEN)	Gremio P.A.(BRA) Z	2,35	2,85	2,15	i
20:40	COPLIB	21 Cruzeiro (BRA)	FC SaoPaulo(BRA) Z	1,80	2,80	3,00	i
20:40	BRAPOK	22 Internacional	Coritiba	1,40	3,00	5,00	i
20:40	BRAPOK	23 Vasco da Gama	Corinthians	2,25	2,80	2,25	i
20:40	NBA	24 L. A. Lakers	Denver Nuggets	1,35	9,85	2,50	i
20:40	NHL	25 Detroit RedWings	Chicago B'hawks	1,50	3,75	3,20	i
Do. 28.05.2009							
17:25	FIN	26 FC Inter Turku	Vaasa PS	1,30	3,55	6,00	i
III							
Spielform: Normalschein		Einsatz: 5,00 €		Gesamteinzahl: €			
Spiele: 0		Quote: -		Mögl.Gewinn: -			

Bild 1.3: Kombi-Wette bei ODDSET

Beispiel

Ein Spieler setzt 10 Euro ein. Auf „1“ im Spiel 13 auf dem Bild 1.3 (Quote 2,25), auf „0“ im Spiel 14 (Quote 3,00) und auf „1“ im Spiel 15 (Quote 2,35). Alle 3 Spiele fallen genau so aus wie gesetzt wurde, somit erhält der Teilnehmende folgenden Gewinn:

10 Euro (Einsatz)*2,25 (Quote Spiel 1)*3 (Quote Spiel 2)*2,35 (Quote Spiel 3)= 158,63 Euro.

Ferner bietet ODDSET viele Untervarianten der Kombiwette an, dessen Betrachtung ich in meiner Diplomarbeit aus Zeit- und Platzgründen leider außen vor lassen muss (u.a. eine Systemwette bei der man Fehltipps ausgleichen kann, oder Handicapwette bei der man Ergebnisse der Spiele leicht verändern kann und vieles mehr).

2. Top-Wette

Die Top-Wette ist eine Einzelwette. Dies bedeutet, dass man bereits dann gewinnt, wenn man mit seiner Vorhersage bei einem der Ereignisse richtig liegt. Hier kann man maximal auf 4 Ereignisse gleichzeitig wetten. Ebenso sollte man bei Top-Wetten zwischen Sieger-, Ergebnis- und Sonderwetten unterscheiden. Bei einer Siegerwette sagt man den Gewinner eines Top-Ereignisses voraus (z.B. Sebastian Vettel-Sieger beim nächsten Formel1-Rennen aber auch Sieger im Hinspiel der Fussball-Bundesligarelegation). Bei einer Ergebniswette kann man in einem Sportereignis auf ein genaues Ergebnis wetten (Beispielsweise auf ein 1:0 im Endspiel der Champions League, siehe Bild 1.4). Bei einer Sonderwette schließlich kann man einige spezielle Wetten abschließen (wie z.B. Absteiger aus der 1.Bundesliga).

Spielplan 22		Dienstag, 26.05.2009 - Montag 01.06.2009	Stand: 26.05.2009, 15:48:52 Uhr		
Bitte wählen Sie Ihre Voraussagen im Spielplan.					
Nr.	TOP-Ereignis		Abgabeschluss	Info	
	Fußball: Champions League - Finale in Rom (TOP-Spiel)				
(-)	03 Champions League	FC Barcelona-Manch.Utd	Mi. 27.05.2009 20:40	i	
<input type="checkbox"/>	01 1:0	6,50	<input type="checkbox"/> 09 4:2	40,00	
<input type="checkbox"/>	02 2:0	10,00	<input type="checkbox"/> 10 and.Heimsieg	22,00	
<input type="checkbox"/>	03 2:1	8,00	<input type="checkbox"/> 11 0:0	8,00	
<input type="checkbox"/>	04 3:0	22,00	<input type="checkbox"/> 12 1:1	6,00	
<input type="checkbox"/>	05 3:1	17,00	<input type="checkbox"/> 13 2:2	11,00	
<input type="checkbox"/>	06 3:2	19,00	<input type="checkbox"/> 14 3:3	35,00	
<input type="checkbox"/>	07 4:0	50,00	<input type="checkbox"/> 15 and.Remis	85,00	
<input type="checkbox"/>	08 4:1	40,00	<input type="checkbox"/> 16 0:1	6,50	
				<input type="checkbox"/> 24 and.Gastsieg	20,00
Fußball: Champions League - Finale in Rom					
(+)	04 Copa Libertadores	Cruzeiro - Sao Paulo	Mi. 27.05.2009 20:40	i	
(-)	05 Relegation 1./2.Bundesliga	Cottbus-Nürnberg (TOP)	Do. 28.05.2009 17:55	i	
<input type="checkbox"/>	01 Heimsieg	2,00	<input type="checkbox"/> 02 Unentschieden	2,70	
				<input type="checkbox"/> 03 Auswärtssieg	3,00
Fußball: Relegation 1.Bundesliga / 2.Bundesliga (TOP-Spiel)					
(+)	06 Relegation 1./2.Bundesliga	Cottbus - Nürnberg	Do. 28.05.2009 17:55	i	
				Gesamteinsatz: 0,00 €	

Bild 1.4: Top-Wette bei ODDSET

Somit bin ich nun am Ende des Tipp- und Wettkomplex angekommen. Allerdings würde ich noch gerne eine komplett alternative Vorhersageart in meiner Diplomarbeit erwähnen und zwar die Vorhersagen mittels eines Fußball Managers.

1.1.4 Vorhersage über einen Fußball Manager

(Dieses Unterkapitel wurde mit Hilfe der Informationen und mit ausdrücklicher Erlaubnis des BrightFuture-Mitbegründers und Konzeptautors der Fußball Manager-Serie [FM09] von Electronic Arts, Gerald Köhler, gestaltet)

Fußballmanager sind unter den Fußballfans sehr beliebt. Mittlerweile ist der Realismus eines Fußball Managers jedoch so hoch, dass sich sogar die Personen, die unmittelbar selbst in das Fußballgeschehen involviert sind, mit Managern auseinandersetzen. So war beispielsweise ein Fußballmanager so detailliert und umfangreich, dass ein englischer Erstligaclub sich ein exklusives Recht auf dessen immense Datenbank sicherte, um diese später zu eigenen Scoutingzwecken nutzen zu können.

Ich wende mich aber nun, ebenfalls einem weltberühmten, sehr beliebten, größtenteils in Deutschland entwickelten und jährlich in vielen Ländern der Welt erscheinenden Fußball Manager „Fußball Manager 2009“ von Bright Future/Electronic Arts zu. Dieses Computerspiel hat seit der neusten Version eine sehr interessante Funktion der „Matchprognose“(siehe Bild 1.5). Diese bietet

die Möglichkeit ein Spiel mittels eigens aufgestellten realen Mannschaften zu berechnen. Hierzu kann man von verschiedenen Ländern und verschiedenen Ligen 2 Mannschaften auswählen, die Formation, sowie jeden einzelnen aufgestellten Spieler festlegen. Außerdem kann man von jedem Spieler die Fitness, Frische, Moral in Zahlen von 1 bis 100 und die Form in Zahlen von 1 bis 20 angeben und damit die aktuelle Form und Verfassung des Spielers mit ins Programm transferieren. Schließlich ist es sogar möglich das Spiel in mehreren Darstellungsmöglichkeiten unter anderem real in einem 3DStadion mit mehreren Perspektiven oder „zuhause“ in einem Videotext anzusehen.



Bild 1.5: Matchprognose-Startbildschirm bei Fussball Manager 2009

Hier eine sehr grobe Beschreibung der Ergebnisberechnung:

Der Berechnung zugrunde liegt eine über 30000 Spieler umfassende Datenbank. Jedem Spieler sind mehrere Stärken zugewiesen (siehe Bild 1.6), aus diesen Stärken (entsprechend der Aufstellung) wird von allen aufgestellten Spielern eine Durchschnittsstärke berechnet und zu Stärke des Vereins summiert. Bei dieser Berechnung werden circa 200 Werte pro Spieler mit ins Betracht gezogen. Ferner werden weitere Faktoren wie unter anderem Heimstärke, Moral, Fitness, Form einkalkuliert und gegeneinander gesetzt und schließlich wird mit einer Berücksichtigung des Zufalls ein Ergebnis berechnet. Alle Berechnungen erfolgen „live“ d.h. taktischen Änderungen des Spielers mitten im Spiel werden genauso mit einbezogen, wie die Auswechslungen sowie weitere Spielereignisse wie z.B. Karten oder Verletzungen. Hinzu kommt auch, dass jeder einzelne Spieler seine Entscheidung (abhängig vom Gegenspieler) theoretisch bis über 100 Mal pro Sekunde überdenken kann (d.h. es wird jede 1/100 Sekunde geprüft ob eine Entscheidung zu treffen ist und ob der Spieler etwas an seinem aktuellen Zustand ändern sollte).

Der Spieler kann also entsprechend und schnell auf das Verhalten seines Gegenspielers und auf das aktuelle Spielgeschehen reagieren.



Bild 1.6: Spielerinformation mit Übersicht einiger Fähigkeits-Werte aus dem Fußball Manager 2009

Somit ergibt sich aus der Summe aller Faktoren der Matchprognose des Fußball Managers 2009 eine realistische Spielvorhersage, die man durchaus als eine Orientierung für den kommenden Spieltag verwenden kann. Ziel dieser Matchprognose „eine möglichst realistische Berechnung eines Fußballspiels zu erreichen“ wird also durchaus erfüllt, auch wenn einiges von subjektiven Eingaben (wie z.B. Form eines Spielers) abhängt.

Hiermit bin ich nun am Ende des 1.Kapitels angelangt. Im nächsten Kapitel beschreibe ich die Attribute, Datenquellen sowie erläutere kurz mein Vorgehen bei der Extraktion von ebendiesen Attributen aus den Datenquellen.

Kapitel 2: Sammlung und Erstellung von Attributen

Dieses Kapitel beschäftigt sich in erster Linie mit den verwendeten Attributen und Datenquellen aus denen diese Attribute stammen.

Dabei unterteilt sich dieses Kapitel in 2 größere Teile:

Im ersten Teil gehe ich auf die Datenquellen ein. Hierbei konzentriere ich mich vorwiegend auf 2 Quellen: „Das Fußball Studio“ (weiter unten nur noch als „Fußball Studio“ bezeichnet) und Kicker-Saisonhefte. Natürlich beschränken sich die Attributquellen nicht nur auf diese beiden, denn es gab ebenso einige Internetseiten aus denen man nützliche Informationen für einige Attribute herausfiltern bzw. schöpfen konnte. **[FDat]** Diese werde ich jedoch aus Zeit und Platzgründen in diesem Kapitel nicht näher beleuchten.

Der zweite Teil des Kapitels beschäftigt sich mit Attributen selbst, hier werden die Attribute in 5 Typen aufgeteilt und nacheinander beschrieben. Diese Unterteilung in 5 Typen, wird auch im weiteren Verlauf der Diplomarbeit von enormer Bedeutung sein.

2.1 Datenquellen

2.1.1 „Fußball Studio“

(Dieses Unterkapitel wurde auf Basis des Programms „Das Fußball Studio“ **[Fst]** von Volker Mallmann gestaltet. Dies geschah mit ausdrücklicher Erlaubnis des Autors)

In diesem Unterkapitel gehe ich auf das Programm „Fußball Studio“ von Volker Mallmann/VMLogic ein. Dabei werde ich nach einer Kurzbeschreibung der Programm möglichkeiten, auf den Export der Daten, sowie auf die Umwandlung der Daten in das für Weka notwendige ARFF-Format eingehen.

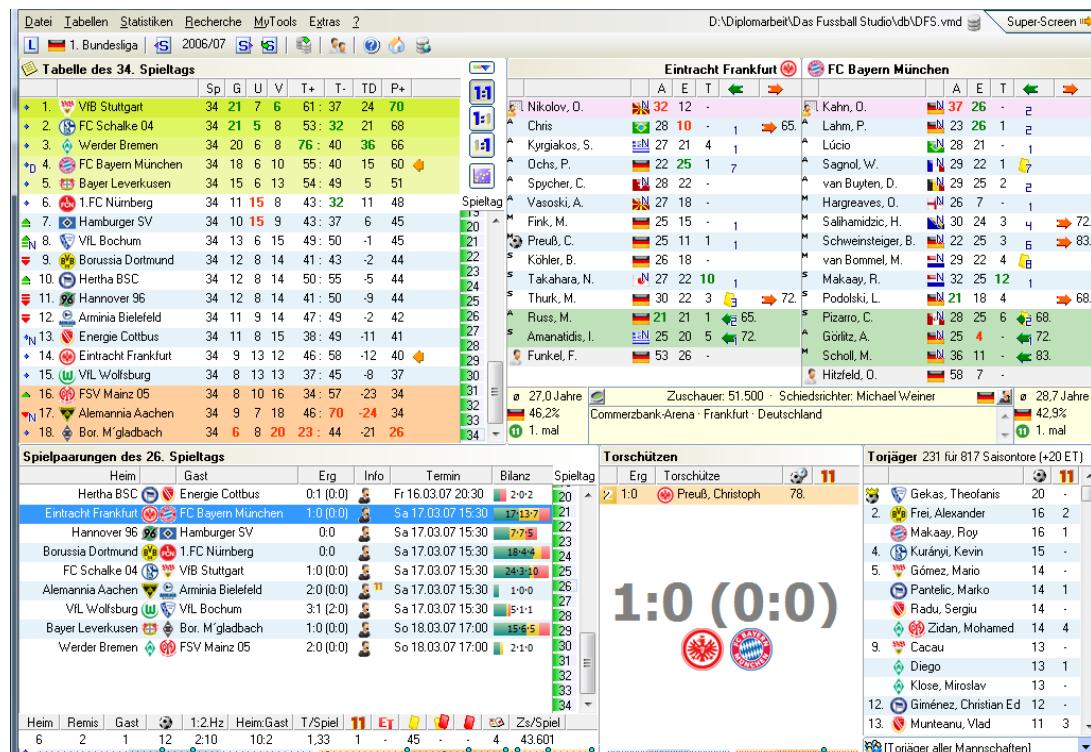


Bild 2.1: Benutzeroberfläche von „Fußball Studio“

Kurze Beschreibung des Programms

„Fußball Studio“ (siehe Bild 2.1) ist einerseits ein sehr leistungsfähiges und andererseits kostenloses Programm zur Verwaltung und Auswertung von Fußball-Ligen. Der immens wichtige Punkt für meine Diplomarbeit ist, dass dieses Programm eine Bundesliga-Datenbank für die 1.Bundesliga seit der ersten Saison im Jahre 1963 besitzt. Ferner hat „Fußball Studio“ sämtliche Statistiken wie u.a. Ergebnisse, Tabellen, Kader, Stadien, Torjäger und wird nach jedem Bundesliga-Spieltag auf den aktuellsten Stand gesetzt. Schließlich hat das Programm eine Exportfunktion, die aber leider stark beschränkt ist und somit ein Eingreifen des Programmautors und eine kleine Spende für die Programmentwicklung unumgänglich machte.

Export der Daten

Einleitung und grobe Beschreibung

Wie bereits im vorherigem Kapitel erwähnt, hat „Fußball Studio“ eine Exportfunktion. In diesem Unterkapitel gehe ich etwas detaillierter auf diese Funktion ein.

Die Exportfunktion [**FStExp**] des Programms (siehe Bild 2.2) verwendet Exportvorlagen. Wenn also eine Exportvorlage ins Programm geladen wird, wird das Programm nach bestimmten Schlüsselwörtern durchsucht und sobald eins dieser Schlüsselwörter gefunden wird, werden an dieser Stelle mittels Variablen/Parameter ausgewählten Daten aus der Datenbank eingesetzt. Eine Exportvorlage kann mit jedem beliebigen Editor erstellt werden. Weiter unten werde ich, die für meine Diplomarbeit notwendige Strukturen der Exportvorlage beschreiben.

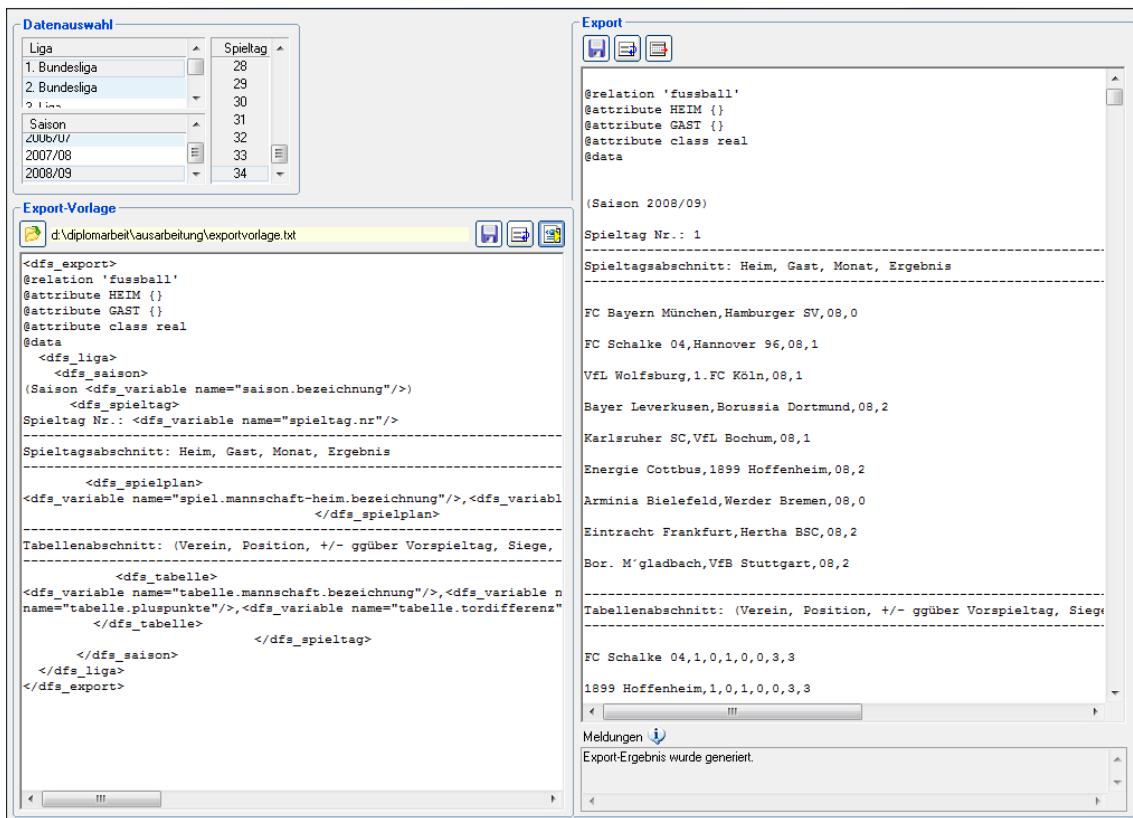


Bild 2.2: Exportbildschirm von „Fußball Studio“

Grundstruktur einer Exportvorlage [FStExp]

Grundsätzlich sind die Bereiche der Exportvorlage mit einem öffnenden und schließenden Tag definiert und ineinander verschachtelt. Zwischen den vorgegebenen Zeilen kann beliebiger Text eingegeben werden, dieser wird dann von der Exportfunktion unverändert ins Ergebnis mit übernommen.

Die Struktur der Exportvorlage sieht folgendermaßen aus:

```
<dfs_export>
  <dfs_liga>
    <dfs_saison>
      <dfs_spieltag>
        <dfs_tabelle>          (optional, kann auch mit dfs_spielplan getauscht werden)
        </dfs_tabelle>
        <dfs_spielplan>        (optional, kann auch mit dfs_tabelle getauscht werden)
        </dfs_spielplan>
      </dfs_spieltag>
    </dfs_saison>
  </dfs_liga>
</dfs_export>
```

Eine kleine Besonderheit, die bevor man auf die Parameter und Variablen eingeht noch erwähnt werden sollte, ist die Schleifefähigkeit von dfs_spieltag, dfs_tabelle und dfs_spielplan. Wenn man bei den Parametern mehr als ein Spieltag auswählt, werden die Daten in diesen Bereichen mehrfach ausgegeben. Somit können mehrere Spieldateien und Tabellen auf einmal exportiert werden.

Variablen und Parameter (ausgewählter Auszug) [FStVP]

Variable/Vorlage	Beschreibung	Beispiel
<dfs_saison>		
saison.bezeichnung	Gibt das Jahr der Saison an	2008/2009 (Saison)
<dfs_spieltag>		
spieltag.nr	Gibt die Nummer des Spieltags wieder	9. (Spieltag)
<dfs_spielplan>		
spiel.mannschaft-heim(gast)bezeichnung	Gibt den Namen des Heim(Gast)vereins beim Spieldatum aus	Eintracht Frankfurt
spiel.start	Gibt den Spielstartzeitpunkt je nach Eingabeformat aus (Kompletter Format: dd.mm.yyyy hh:mm (einzelne Fragmente daraus möglich))	25.05.2009 15:30 oder nur „25“ bzw. „05“
spiel.toto-wert	Gibt die Ergebnistendenz aus (Klassenattribut)	„0“, „1“ oder „2“
<dfs_tabelle>		
tabelle.mannschaft.bezeichnung	Gibt den Mannschaftsnamen in der Tabelle aus	Eintracht Frankfurt
tabelle.platzierung	Gibt den Tabellenplatz der obengenannten Mannschaft	13. (Platz)
tabelle.siege(.unentschieden,.niederlagen)	Gibt die Anzahl der Siege (Unentschieden, Niederlagen) der obengenannten Mannschaft in der Tabelle aus	7 (Siege)
tabelle.pluspunkte	Gibt die Anzahl der Punkte der obengenannten Mannschaft aus	55 (Punkte)
tabelle.tordifferenz	Gibt die Tordifferenz der obengenannten Mannschaft aus	-3

Tabelle 2.1: (ausgewählte) Variablen und Parameter der Exportvorlage von „Fußball Studio“

Somit wurde von mir mittels oben genannter Variablen, die unten angegebene Exportvorlage erstellt. Ein beispielhaftes Ergebnis dieser Exportvorlage ist direkt danach auf der Seite 15 zu finden. Da Tabelle und Spieltag nur in 2 unterschiedlichen Bereichen ausgegeben werden könnten, musste ich die Ausgabe in einen Tabellenbereich und in einen Spieltagsbereich trennen. Beide Bereiche werden später bei der Umwandlung in Weka's ARFF Format zusammengefügt (siehe Seite 17).

Die Exportvorlage

```
<dfs_export>
@relation 'fussball'
@attribute HEIM {}
@attribute GAST {}
@attribute class real
@data
  <dfs_liga>
    <dfs_saison>
      (Saison <dfs_variable name="saison.bezeichnung" />)
        <dfs_spieltag>
          Spieltag Nr.: <dfs_variable name="spieltag.nr" />

-----
Spieltagsabschnitt: Heim, Gast, Monat, Ergebnis
-----

<dfs_spielplan>
<dfs_variable name="spiel.mannschaft-heim.bezeichnung"/>,<dfs_variable
name="spiel.mannschaft-gast.bezeichnung"/>,<dfs_variable name="spiel.start"
format="mm" />,<dfs_variable name="spiel.toto-wert"/>
</dfs_spielplan>

-----
Tabellenabschnitt: (Verein, Position, +/- ggüber Vorspieltag, Siege, Unentschieden,
Niederlagen, Tabellenpunkte, Tordifferenz
-----

<dfs_tabelle>
<dfs_variable name="tabelle.mannschaft.bezeichnung"/>,<dfs_variable
name="tabelle.platzierung"/>,<dfs_variable name="tabelle.platzierung-
delta"/>,<dfs_variable name="tabelle.siege"/>,<dfs_variable
name="tabelle.unentschieden"/>,<dfs_variable
name="tabelle.niederlagen"/>,<dfs_variable
name="tabelle.pluspunkte"/>,<dfs_variable name="tabelle.tordifferenz" />
</dfs_tabelle>
  </dfs_spieltag>
</dfs_saison>
</dfs_liga>
</dfs_export>
```

Beispielhaftes Ergebnis der Exportvorlage

```
@relation 'fussball'
@attribute HEIM {}
@attribute GAST {}
@attribute class real
@data
(Saison 2008/09)

Spieltag Nr.: 1
-----
Spieltagsabschnitt: Heim, Gast, Monat, Ergebnis
-----
Bayern München,Hamburger SV,08,0
FC Schalke 04,Hannover 96,08,1
VfL Wolfsburg,1.FC Köln,08,1
Bayer Leverkusen,Borussia Dortmund,08,2
Karlsruher SC,VfL Bochum,08,1
Energie Cottbus,1899 Hoffenheim,08,2
Arminia Bielefeld,Werder Bremen,08,0
Eintracht Frankfurt,Hertha BSC,08,2
Bor. M'gladbach,VfB Stuttgart,08,2
-----
Tabellenabschnitt: (Verein, Position, +/- ggüber Vorspieltag, Siege, Unentschieden,
Niederlagen,Tabellenpunkte,Tordifferenz
-----
FC Schalke 04,1,0,1,0,0,3,3
1899 Hoffenheim,1,0,1,0,0,3,3
VfB Stuttgart,3,0,1,0,0,3,2
Hertha BSC,4,0,1,0,0,3,2
Borussia Dortmund,5,0,1,0,0,3,1
VfL Wolfsburg,6,0,1,0,0,3,1
Karlsruher SC,7,0,1,0,0,3,1
Arminia Bielefeld,8,0,0,1,0,1,0
Werder Bremen,8,0,0,1,0,1,0
Hamburger SV,8,0,0,1,0,1,0
Bayern München,8,0,0,1,0,1,0
Bayer Leverkusen,12,0,0,0,1,0,-1
1.FC Köln,13,0,0,0,1,0,-1
VfL Bochum,14,0,0,0,1,0,-1
Bor. M'gladbach,15,0,0,0,1,0,-2
Eintracht Frankfurt,16,0,0,0,1,0,-2
Energie Cottbus,17,0,0,0,1,0,-3
Hannover 96,17,0,0,0,1,0,-3
```

Umwandlung der Daten in die für Weka notwendige Form

In diesem Unterkapitel wird beschrieben, wie die in vorherigem Kapitel exportierten Daten aus dem „Fußball Studio“ in das ARFF-Format von Weka umgewandelt wurden. Dabei gehe ich zuerst kurz auf das ARFF-Format von Weka ein, um danach die Schritte der ARFF-Umwandlung aufzuzählen.

ARFF-Format

Ein Weka-Input File besitzt ein ARFF-Format (siehe hierzu Anhang A). Nachfolgend erläutere ich dieses Format etwas detaillierter. **[WIFR05]** **[ARFF]**

Eine ARFF-Datei (Attribute-Relation File Format) besteht aus ASCII-Text, der eine Liste von Instanzen beschreibt. Diese Liste teilt einen Satz von Attributen. ARFF-Dateien unterteilen sich in 2 Hauptbereiche, einen Header und einen Datenteil.

Header

Im Header wird zuerst der Name der Relation aufgeführt, danach folgt eine Attributliste mit jeweiligem entsprechenden Attributentyp.

Somit sieht ein ARFF-Header beispielsweise so aus:

```
@relation 'fussball'  
@attribute HEIM{energiécottbus,hannover96,bayernmuenchen,borussia dortmund,vflwolfsburg,...}  
@attribute GAST{energiécottbus,hannover96,bayernmuenchen,borussia dortmund,vflwolfsburg,...}  
@attribute POSHEIM real  
@attribute PLUSMINUSHEIM real  
...
```

Daten-Teil (Data)

Direkt nach dem Header folgt der Daten-Teil. Im Daten-Teil stehen die Werte/Informationen der im Header enthaltenen Attribute. Der Daten-Teil wird durch ein @data eingeleitet, genauer gesagt sobald @data in einer ARFF-Datei auftaucht, endet der Header und der Datenteil beginnt.

Somit sieht ein Daten-Teil beispielsweise so aus:

```
... (Header-Teil)  
@data  
energiécottbus,hannover96,18,0,4,...  
bayernmuenchen,borussia dortmund,1,0,13,...  
...
```

Umwandlung der Daten von „Fußball Studio“ ins ARFF-Format

Nun war es mein Ziel, das auf Seiten 14-15 bereits erwähnte Ausgabeformat der Exportfunktion von „Fußball Studio“ in das ARFF-Format von Weka umzuwandeln.

Hierfür waren folgende Schritte notwendig:

1. Vereinheitlichung der Vereinsnamen ins Weka-Format
2. Eliminierung von überflüssigen Informationen (und Leerzeichen)
3. Hinzufügen von Attributinformationen im ARFF-Header
4. Vereinigung vom Werten aus Spieltags- und Tabellenabschnitt im ARFF-Daten-Teil
5. Umbenennung der Output-Datei ins ARFF-Format.

Den Sinn und die Implementierungsart dieser Schritte gehe ich nun etwas detaillierter durch.

1. Vereinheitlichung der Vereinsnamen ins Weka-Format

Dieser Schritt war notwendig um einerseits Vereinsnamen in eine einheitliche und vor allem Weka-lesbare Form zu bringen und andererseits einige Problemfälle in der Namengebung zu eliminieren.

Es wurden hier unter anderem:

- Die Leerzeichen zwischen den Vereinsnamen entfernt, (da Weka doppelte Vereinsnamen als 2 verschiedene Vereine interpretiert), ferner wurde auch der Einfachheit halber überall die Kleinschreibung verwendet (Bsp. Eintracht Frankfurt->eintrachtfrankfurt)
- Umlaute und ß entfernt (ä->ae, ö->oe, ü->ae, ß->ss)
- Sonderzeichen aus den Vereinsnamen entfernt (Bsp. Punkte: fcst.pauli->fcstpauli, oder Striche: rot-weissessen->rotweissessen)

-Berücksichtigung der Vereinsnamenänderungen (Bsp. MSV Duisburg hieß bis Januar 1967 Meidericher SV. Aus Gründen der Einfachheit wurde von mir während der gesamten Zeit (also auch in den Jahren 1963-1966) der Name MSV Duisburg verwendet)

2. Eliminierung von überflüssigen Informationen (und Leerzeichen)

In dieser Phase wurden lediglich die überflüssigen Informationen, die in der Exportvorlage aus Gründen der Übersicht eingetragen wurden, sowie überflüssige Leerzeichen entfernt (Beispielsweise: Information über Spieltagsabschnitt und Tabellenabschnitt; siehe Exportvorlage auf Seite 15).

3. Hinzufügen von Attributinformationen im ARFF-Header

Wie bereits auf den vorherigen Seiten erwähnt, wurden im Header Informationen über alle Attribute hinzugefügt. Dies geschah über `~s` Befehl von Perl. Beispielsweise der Befehl: „`$line=~s/@attribute POSHEIM real\n@attribute PLUSMINUSHEIM real...`“ fügte die Attribute Posheim und Plusminusheim mit dem Attributtyp real in den Header der Datei ein.

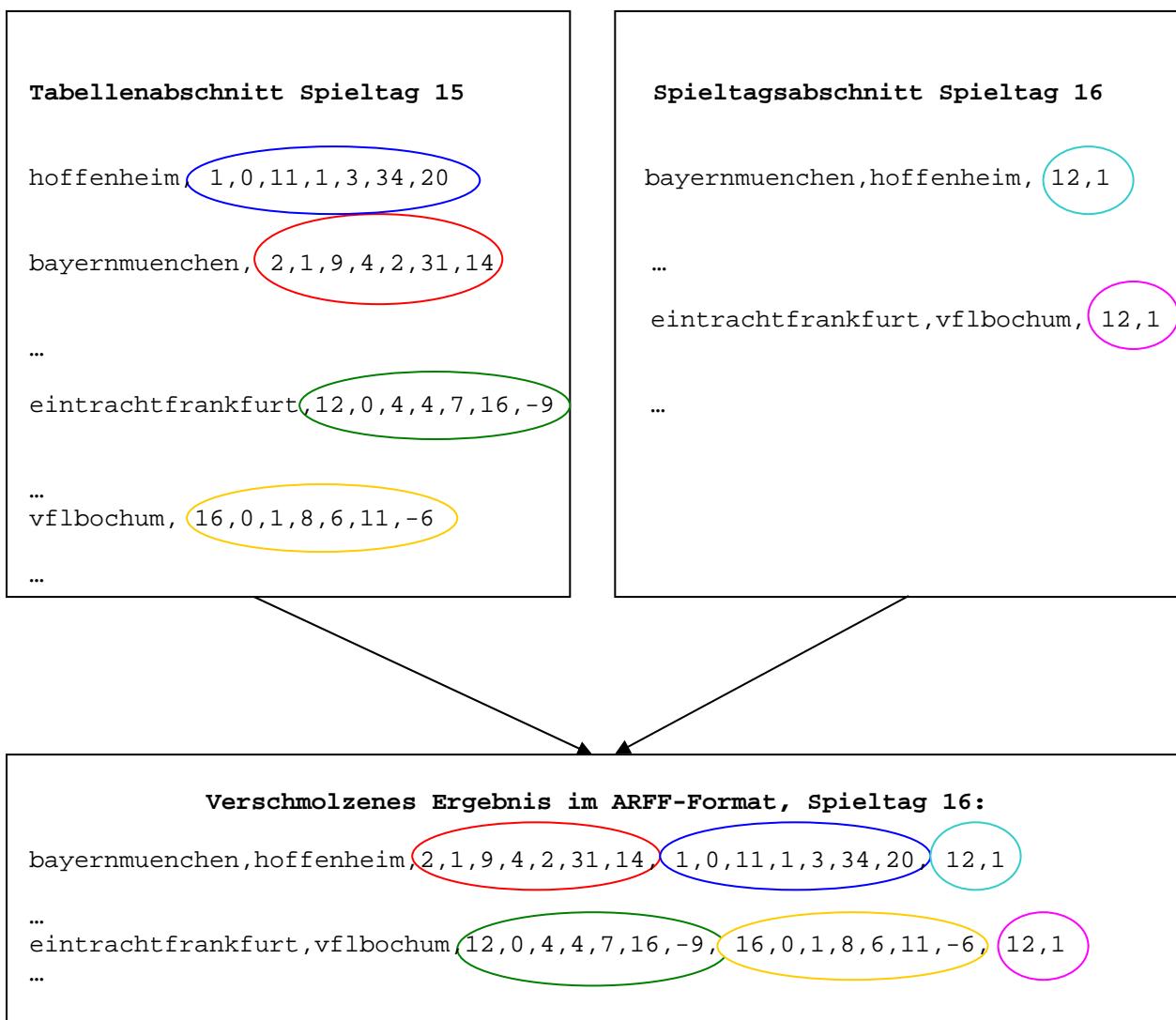


Bild 2.3: Verschmelzung der „Fußball Studio“-Exportdaten ins ARFF-Format

4. Vereinigung vom Werten aus Spieltags- und Tabellenabschnitt im ARFF-Data

Danach wurden die Spieltags- und Tabellenabschnitte verschmolzen. Dabei wurde der Tabellenabschnitt des Spieltages X-1 in den Spieltagsabschnitt X vor dem Klassen- und Monatsattribut importiert (zur Verdeutlichung siehe Bild 2.3). Im Sonderfall des 1.Spieltags werden

die Daten vom 34.Spieltag des Tabellenabschnitts der letzten Saison importiert (Aufsteiger erhalten einen Defaultwertesatz von 16(Platz),0,...,0).

5. Umbenennung der Output-Datei in ARFF-Format

Zum Schluss wurde die Output-Datei in output20xx/xy.ARFF umgewandelt um damit in Weka einlesbar zu werden. Dabei entspricht xx dem Startjahr und xy dem Endjahr der Saison.

Es liegt nun fertige ARFF-Datei mit den exportierten Daten aus dem „Fußball Studio“ vor. Im nächsten Unterkapitel wird nun auf eine weitere Datenquelle, auf „Kicker Saisonheft“ eingegangen.

2.1.2 Kicker-Saisonhefte und sonstige Quellen

Um manuelle Daten eintragen zu können, brauchte ich eine Quelle, bei der die Daten direkt vor einer Saison vorlagen. Der Grund hierfür war, dass Änderungen während einer Saison, nur sehr schwer für alle Vereine in gleichem Umfang zu finden und auch in gleichen Maßen zu berücksichtigen sind. Hierfür waren Kicker-Saisonhefte [Kicker] (siehe Bild 2.4) mit den Vereinsdaten direkt vor jeder Saison die ideale Lösung. Die einzige Beschränkung war, dass die Saisonhefte nur von Saisons 2002-2008 vorlagen, was aber für eine umfangreiche Untersuchung der manuellen Attribute (siehe Kapitel 4) mehr als ausreichend war.



Bild 2.4: Kicker-Saisonhefte

Die Datenextraktion aus einem Kicker-Saisonheft wurde folgendermaßen umgesetzt: Daten wurden in den Kicker-Saisonheften manuell markiert (siehe Bild 2.5) und extrahiert. Die extrahierten Daten wurden in eine externe Datei eingetragen, die dann vom Hauptprogramm eingelesen wurde. Im Hauptprogramm wurde dieses Attribut jeweils einmal für Heim- und für Auswärtsmannschaft eingefügt und mit anderen Attributen in die ARFF-Datei von Weka exportiert. Die manuell eingetragenen Attribute sind an dem Dateinamen addxy im Quellcode des Hauptprogramms zu erkennen (siehe die Attributliste im Anhang B oder Kapitel 2.2.3).

29 Dreher, Bernd	02.11.1966	D	1,87	80	1990	11 A-Länderspiele
1 Kahn, Oliver	15.06.1969	D	1,88	91	1994	Karlsruher SC (1975-94), 86 A-Länderspiele
22 Rensing, Michael	14.05.1984	D	1,88	86	2000	TuS Lingen (1988-00), 18 U-21-Länderspiele
6 Demichelis, Martin	20.12.1980	ARG/ ITA	1,84	78	2003	River Plate Buenos Aires (1998-03), Clube Renato Cesarini (1995-98), Complejo Deportivo (bis 1995), 3 A für Argentinien
32 Hummels, Mats	16.12.1988	D	1,91	88	10/95	eigene Jugend, 2 U-21-Länderspiele
25 Ismael, Valerien	28.09.1975	FRA	1,91	81	2005	Werder Bremen (2003-05), Racing Straßburg (2002/03, 1/01-01, 1984-12/97), RC Lens (2001/02, 11/98-12/00), Crystal Palace (1/98-10/98)
23 Jansen, Marcell	04.11.1985	D	1,91	84	2007	Bor. M'gladbach (93-07), SV M'gladbach 1910 (bis 93), 16 A-Länderspiele, 4 U21
21 Lahm, Philipp	11.11.1983	D	1,70	64	2005	VfB Stuttgart (2003-05), Bayern München (1995-03), FT Gern München (bis 1995), 34 A-Länderspiele, 3 U21
30 Leitl, Christian	29.08.1984	D	1,81	78	2006	1. FC Köln (04-06), Bayern München (93-04), Alemannia München (bis 93), 2 U21
3 Lucio	08.05.1978	BRA	1,88	84	2004	Bayer Leverkusen (1/01-04), Internaciona Porto Alegre (1/98-12/00), Clube Regatas Guara (1/95-12/97), Planaltina EC (bis 12/94), 61 A für Brasilien
2 Sagnol, Willy	18.03.1977	FRA	1,82	81	2000	AS Monaco (1997-00), AS St. Etienne (1990-97), 53 A für Frankreich

Bild 2.5: Beispiel der Markierung für Datenextraktion aus einem Kicker-Saisonheft

Der Quellcode der externen Datei lautet beispielweise:

```
%nationalspieler = ();
$nationalspieler{'energiécottbus'} = 0;
$nationalspieler{'borussiadortmund'} = 6;
$nationalspieler{'hannover96'} = 5;
$nationalspieler{'bayernmuenchens'} = 12;
$nationalspieler{'eintrachtfrankfurt'} = 1;
...
...
```

Im Hauptprogramm wurde der Name dieser externen Datei dann importiert (z.B. u.a. do("add1Data")... und \$addnationalheim = \$nationalspieler{\$vereinheim}; \$addnationalgast = \$nationalspieler{\$vereingast};)

Somit wurden damit über 20 zusätzliche Attribute für die Saisons 2002-2008 gewonnen. Zur Überprüfung und Ergänzung dieser Attribute wurden einige Internetquellen in Form von Sportportalen [FDat] benutzt. Eine detailliertere Beschreibung der Attribute finden Sie im nächsten Unterkapitel.

2.2 Beschreibung der Attribute

In diesem Unterkapitel finden Sie eine Beschreibung sämtlicher Attribute. Zuallererst gehe ich auf die Attribute von „Fußball Studio“ ein, um danach mit den (aus den „Fußball Studio“ Daten) eigens erstellten Gruppenattributen fortzufahren. Danach werde ich auf die manuell eingetragene Attribute eingehen um schließlich mit den kombinierten und auf Basis anderer Attribute neuerstellten Attributen das Kapitel sowie Differenzattributen das Kapitel abzuschließen. Der Wertebereich aller Attribute (mit Ausnahme von einigen wenigen) ist „real“. Die Ausnahmen sind die Attribute „Heim“ und „Gast“ (hier werden Strings in Form von vorgegebenen Vereinsnamen der betrachteten Saison verwendet), sowie die Variablen „Tabtendenz“ (fest definierter Wertebereich: 1,2,3,4), „Tabvergleich“(1,2), „Violetore“(0,1) und das Klassenattribut „Tendenz“(0,1,2).

2.2.1 „Fußball Studio“-Attribute

In diesem Unterkapitel geht es um die Attribute, die aus dem Programm „Fußball Studio“ extrahiert wurden (siehe Unterkapitel 2.1.1). Diese Attribute sind für die gesamte Dauer der Bundesliga vorhanden (also von Saison 1963-heute) und beziehen sich auf die Ergebnisse des letzten Spieltags. Die Aufzählung der Attribute finden Sie unterhalb in der Tabelle 2.2. Mit „#“ ist in der Tabelle die Attributnummer im Weka ARFF-Datensatz gemeint.

Hauptidee hinter den „Fußball Studio“-Attributen

Diese Attribute spiegeln den Verlauf der bisherigen Saison wieder. Anhand der Tabellendaten kann man u.a. erkennen, wie gut das betrachtete Team momentan ist. Zusätzlich hierzu wurde noch das Attribut Monat exportiert, um eventuelle Formschwankungen einiger Teams in bestimmten Jahreszeiten zu prüfen. Das Attribut Tendenz ist das Klassenattribut.

<u>„Fußball Studio“ Attribute</u>		
gelten für Saisons: 1963-heute		
#	Name	Erklärung
1/2	Heim/Gast	Heim(Gast)mannschaft des zu betrachtenden Spiels
3/10	PosHeim/-Gast	Position der Heim(Gast)mannschaft
4/11	PlusminusHeim/-Gast	Veränderung der Platzierung der Heim(Gast)mannschaft nach dem letzten Spieltag gegenüber der Platzierung vor dem letzten Spieltag
5/12	SiegeHeim/-Gast	Anzahl Siege der Heim(Gast)mannschaft im bisherigen Verlauf der Saison
6/13	RemisHeim/-Gast	Anzahl Unentschieden der Heim(Gast)mannschaft im bisherigen Verlauf der Saison
7/14	NiederlagenHeim/-Gast	Anzahl Niederlagen der Heim(Gast)mannschaft im bisherigen Verlauf der Saison
8/15	PunkteHeim/-Gast	Aktueller Punktestand der Heim(Gast)mannschaft
9/16	TDHeim/-Gast	Aktuelle Tordifferenz der Heim(Gast)mannschaft
17	Monat	Monat in dem das Spiel stattgefunden hat
170	Tendenz (Klassenattribut)	Die Ergebnistendenz des Spiels. Dabei hat diese Attribut 3 Ausgangswerte: "0", "1" und "2". Eine „0“ steht für ein Unentschieden „1“ steht für einen Sieg der Heimmannschaft „2“ steht für einen Sieg der Gastmannschaft

Tabelle 2.2: Aufzählung und Erklärung der „Fußball Studio“-Attribute

2.2.2 Gruppenattribute

In diesem Unterkapitel geht es um die sogenannten Gruppenattribute. Als Gruppenattribute habe ich die Attribute bezeichnet, die in mehrfacher Form unterschiedend nur durch verschiedene Variablenbelegung mit in die Attributmenge aufgenommen wurden. Diese Gruppenattribute sind in jeweils fünffacher Form, also in Fünfer-Gruppen (in je verschiedenen Heim- und Auswärtsausprägungen) vorhanden. Dabei gehe ich in diesem Unterkapitel erst einmal nicht auf die einzelnen Belegungen ein, diese werden später im Kapitel 4 besprochen und geprüft.

Zu den Gruppenattributen zählen: Versusattribute, Formattribute, AvgPlatz und Abstiegsattribute.

Versus-Attribute

Versus-Attribute sind Attribute bei denen die letzten X Aufeinandertreffen der spielenden Teams unter sich betrachtet werden. X wird durch 5 verschiedene Belegungen variiert. Dabei gibt es 2 Versusausprägungen: Versus_Heim und Versus_Gast. (mit Versus gibt es noch eine Differenzausprägung der beiden anderen Ausprägungen, diese wird also bei den Differenzen tabellarisch aufgelistet).

Zur Erklärung und Verdeutlichung nehme ich an es gibt ein Team A aus Aachen und ein Team B aus Berlin. Ich nehme ferner an, dass X=3 sei.

- VersusHeim-Attribut berechnet die letzten 3 Aufeinandertreffen in Aachen, Team A wird dabei jeweils als Heimteam geführt.
- VersusGast-Attribut berechnet die letzten 3 Aufeinandertreffen in Berlin, Team B wird dabei jeweils als Heimteam geführt (also einfach die Umkehrung von VersusHeim).
- Die Differenzausprägung Versus berechnet, wie der Name schon sagt, die Differenz aus VersusHeim und VersusGast. Somit berechnet Versus die letzten 3 Aufeinandertreffen der beiden Teams (Team A bzw. Team B) zuhause und auswärts (d.h. also die letzten 6 Aufeinandertreffen).

Die Berechnung bei Versus-Attributen

- Bei einem Sieg des Heimteams im betrachteten Aufeinandertreffen wird das jeweilige Versusattribut um 3 Punkte erhöht
- Bei einem Sieg des Gastteams im betrachteten Aufeinandertreffen wird das jeweilige Versusattribut um 3 Punkte gesenkt
- Bei Unentschieden bleibt das Versusattribut unverändert

Aufgrund der Ergebnisse im Versusattribut, kann man eine Tendenz der letzten Aufeinandertreffen von betrachteten Mannschaften erkennen. Bei einem Pluswert des Attributs war das Heimteam die letzten Spiele besser, bei einem Minuswert war das Gastteam besser.

Ein Beispiel zur Verdeutlichung

Team A (Aachen) gegen Team B (Berlin)

X=3; (bedeutet, ich betrachte die letzten 3 Aufeinandertreffen)

(alle Angaben der Ergebnisse erfolgen aus Gründen der einfacheren Übersicht aus der Sicht von Team A)

Zuhause bei Team A (Aachen)

Spiel 1: 1:2

Spiel 2: 2:1

Spiel 3: 0:0

Zuhause bei Team B (Berlin)

1:2

1:2

1:2

VersusHeim (in Aachen): $1:2(-3)+2:1(3)+0:0(0)=0$

VersusGast (in Berlin): $2:1(3)+2:1(3)+2:1(3)=9$

Versus (Differenzausprägung): VersusHeim-VersusGast; $0-9=-9$

Damit ist der Versus-Wert -9. Da der Wert ein Minuswert ist, ist ein Vorteil für Team B sichtbar.

Hauptidee hinter den Versus-Attributen

Ein Team, das den Vergleich deutlich gewinnt, kann ein überlegenes Team oder ein sogenannter Angstgegner sein, somit können aus den letzten Ergebnissen in den Spielen gegeneinander viele Rückschlüsse gezogen werden.

Versus-Attribute		
gelten für Saisons: 1963-heute		
#	Name	Erklärung
34/35	Versus1Heim/-Gast	Ergebnisse der Aufeinandertreffen der beiden Teams in den letzten X1-Spielen von Team A (Heimteam) (bzw. Team B (Gastteam))
37/38	Versus2Heim/-Gast	Ergebnisse der Aufeinandertreffen der beiden Teams in den letzten X2-Spielen von Team A (Heimteam) (bzw. Team B (Gastteam))
40/41	Versus3Heim/-Gast	Ergebnisse der Aufeinandertreffen der beiden Teams in den letzten X3-Spielen von Team A (Heimteam) (bzw. Team B (Gastteam))
43/44	Versus4Heim/-Gast	Ergebnisse der Aufeinandertreffen der beiden Teams in den letzten X4-Spielen von Team A (Heimteam) (bzw. Team B (Gastteam))
46/47	Versus5Heim/-Gast	Ergebnisse der Aufeinandertreffen der beiden Teams in den letzten X5-Spielen von Team A (Heimteam) (bzw. Team B (Gastteam))

Tabelle 2.3: Aufzählung und Erklärung der Versus-Gruppenattribute

Form-Attribute

Form-Attribute konzentrieren sich auf die letzten X Spiele eines Teams. Dabei kann ausgewählt werden, ob die Formattribute saisonübergreifend berechnet werden sollen oder man jeweils immer am Anfang einer neuen Saison das Attribut jeweils immer „resetten“ soll. (Eine Überprüfung welche der beiden Optionen besser ist, erfolgt im 4. Kapitel)

Dabei gibt es 4 sichtbare Formattributausprägungen:

„FormHeimHeim“, „FormGastAusw“, „FormHeim“, „FormGast“

(Ebenfalls gibt es noch eine Differenzausprägung „DiffForm“, die bei den Differenzen gelistet ist)

und 2 unsichtbare programminterne FormAttributAusprägungen:

- „FormHeimAusw“ (überflüssig, da es die Form der Heimmannschaft in den letzten Auswärtsspielen betrachtet. Die Leistung aus den x Auswärtsspielen ist für die Form des Heimteams als Einzelattribut eigentlich wenig relevant)

- „FormGastHeim“ (ebenso überflüssig, da es hier um die Form des Gastteams in den letzten x Heimspielen geht. Ebenso als Einzelattribut überflüssig)

Die beiden unsichtbaren Attribute, werden später nur benutzt um die Gesamtform des Heimteams bzw. des Gastteams zu berechnen.

Auch hier nehme ich zur Verdeutlichung ein Team A aus Aachen und ein Team B aus Berlin.

Angenommen es findet ein Spiel in Aachen Team A gegen Team B statt.

-FormHeim berechnet die Form des Heimteams (also hier Team A) in den letzten X Heimspielen (also in Aachen)

-FormGastAusw berechnet die Form des Auswärtsteams (also hier Team B) in den letzten X Auswärtsspielen (also in Berlin)

-FormHeim berechnet die Form des Heimteams (hier Team A) in den letzten X Spielen (Heim und Auswärtsspielen). Berechnet wird es dabei aus der Addition des sichtbaren Attributs FormHeimHeim und des programminternen Attributs FormHeimAusw

-FormGast berechnet die Form des Gastteams (hier Team B) in den letzten X Spielen (Heim und Auswärtsspielen). Berechnet wird es dabei aus der Addition des sichtbaren Attributs FormGastAusw und des programminternen Attributs FormGastHeim

Die Berechnung der Form erfolgt ähnlich zu Versusattribut:

-Bei einem Sieg des betrachteten Teams wird das jeweilige Formattribut um 3 Punkte erhöht

-Bei einer Niederlage des betrachteten Teams wird das jeweilige Formattribut um 3 Punkte gesenkt

-Bei einem Unentschieden bleibt das Formattribut unverändert

Beispiel zur Verdeutlichung

Team A(Aachen) trifft auf Team B (Berlin), das Spiel findet in Aachen statt. Ergebnisse beider Teams sind aus deren jeweiligen Sicht dargestellt.

Ergebnisse TeamA:

0:1 (Auswärts)

1:0 (Zuhause)

3:2 (Auswärts)

0:0 (Zuhause)

Ergebnisse Team B:

3:3 (Zuhause)

1:2 (Auswärts)

1:3 (Zuhause)

0:2 (Auswärts)

Es ergeben sich also folgende Ergebnisse der Formatattribute:

FormHeimHeim: Team A->1:0(3)+0:0(0)=3

FormGastAusw: Team B->1:2(-3)+0:2(-3)=-6

FormHeim: Team A->0:1 (-3)+1:0(3)+3:2(3)+0:0(0)=3

FormGast: Team B->3:3(0)+1:2(-3)+1:3(-3)+0:2(-3)=-9

(DiffForm: FormHeim-FormGast=12)

Hauptidee hinter den Form-Attributen

Ein Team, das mehrere Spiele hintereinander (zuhause, auswärts oder insgesamt) gewinnt oder verliert, kann einen leichter vorherzusagenden Aufwärtstrend oder Abwärtstrend starten.

<u>Form-Attribute</u>		
<i>gelten für Saisons:</i> 1963-heute		
#	Name	Erklärung
51/52	Form1HeimHeim/-GastAusw	Form des Heim(Gast)teams in den letzten X1-Heim(Auswärts)spielen
53/54	Form1Heim/-Gast	Form des Heim/(Gast)teams in den letzten X1-Spielen
56/57	Form2HeimHeim/-GastAusw	Form des Heim(Gast)teams in den letzten X2-Heim(Auswärts)spielen
58/59	Form2Heim/-Gast	Form des Heim/(Gast)teams in den letzten X2-Spielen
61/62	Form3HeimHeim/-GastAusw	Form des Heim(Gast)teams in den letzten X3-Heim(Auswärts)spielen
63/64	Form3Heim/-Gast	Form des Heim/(Gast)teams in den letzten X3-Spielen
66/67	Form4HeimHeim/-GastAusw	Form des Heim(Gast)teams in den letzten X4-Heim(Auswärts)spielen
68/69	Form4Heim/-Gast	Form des Heim/(Gast)teams in den letzten X4-Spielen
71/72	Form5HeimHeim/-GastAusw	Form des Heim(Gast)teams in den letzten X5-Heim(Auswärts)spielen
73/74	Form5Heim/-Gast	Form des Heim/(Gast)teams in den letzten X5-Spielen

Tabelle 2.4: Aufzählung und Erklärung der Form-Gruppenattribute

AvgPlatz-Attribute

AvgPlatz-Attribute berechnen die Durchschnittsplatzierung eines Teams in den letzten X Saisons. Dabei zählt eine Nichtteilnahme in einer Saison wegen eines Abstiegs als ein 19.Platz in dieser Saison (siehe Beispiel unten). Von AvgPlatz gibt es 2 Ausprägungen: AvgPlatzHeim (Durchschnittsplatzierung des Heimteams in den letzten X Jahren) und AvgPlatzGast (Durchschnittsplatzierung des Gastteams in den letzten X Jahren). Auch von AvgPlatz gibt es eine Differenzausprägung „DiffAvg“, auch diese wird (wie auch bei den anderen Gruppenattributen) bei Differenzen gelistet.

Beispiel: X=5

Team A hat in den letzten 5 Saisons folgenden Platzierungen erreicht:

12,16, (zweite Liga=19), (zweite Liga=19), 8

Dadurch werden folgende Durchschnittsplatzierung berechnet: $12+16+19+19+8=14,8$

Diese Durchschnittsplatzierung wird mit der Durchschnittsplatzierung von Team B in den letzten 5 Saison verglichen bzw. für „DiffAvg“ wird hierfür auch eine Differenz gebildet.

Hauptidee hinter den AvgPlatz-Attributen

Ein Team, das in der Vergangenheit bessere Platzierungen und damit eine bessere Durchschnittsplatzierung erreicht hat, tendiert dazu, bessere Ergebnisse zu erzielen.

AvgPlatz-Attribute		
gelten für Saisons: 1963-heute		
#	Name	Erklärung
77/78	Avg1PlatzHeim/-Gast	Durchschnittsplatzierung der Heim(Gast)mannschaft in den letzten X1 Saisons
80/81	Avg2PlatzHeim/-Gast	Durchschnittsplatzierung der Heim(Gast)mannschaft in den letzten X2 Saisons
83/84	Avg3PlatzHeim/-Gast	Durchschnittsplatzierung der Heim(Gast)mannschaft in den letzten X3 Saisons
86/87	Avg4PlatzHeim/-Gast	Durchschnittsplatzierung der Heim(Gast)mannschaft in den letzten X4 Saisons
89/90	Avg5PlatzHeim/-Gast	Durchschnittsplatzierung der Heim(Gast)mannschaft in den letzten X5 Saisons

Tabelle 2.5: Aufzählung und Erklärung der AvgPlatz-Gruppenattribute

Abstieg-Attribute

Bei Abstieg-Gruppenattributen handelt es sich um Attribute bei denen die Zugehörigkeit eines Teams zur 1.Bundesliga in den letzten X Jahren berechnet wird. Abstiegsattribute haben 2 Ausprägungen: AbstiegHeim (Anzahl der Saisons der Heimmannschaft in der Bundesliga in den letzten X Jahren) und AbstiegGast (Anzahl der Saisons der Gastmannschaft in der Bundesliga in den letzten X Jahren). Auch bei Abstieg gibt es eine Differenzausprägung „DiffAbstieg“, die bei den Differenzattributen gelistet wird.

Beispiel

X=10

Team A ist in den letzten 10 Jahren, 2x abgestiegen, daher wird das Attribut den Wert 8 zurückliefern, d.h. von den letzten 10 Saisons war Team A 8 Saisons in der Bundesliga. Dieser Wert wird nun mit dem Wert von Team B verglichen (bzw. für „DiffAbstieg“ eine Differenz gebildet).

Hauptidee hinter den Abstieg-Attributen

Ein Team, das selten abgestiegen ist, tendiert dazu bessere Ergebnisse zu erzielen, als die sogenannten Fahrstuhlmannschaften (Mannschaften, die oft auf- und absteigen).

Abstieg-Attribute		
gelten für Saisons: 1963-heute		
#	Name	Erklärung
134/135	Abstieg1Heim/-Gast	Anzahl der Bundesligasaisons des Heim(Gast)teams in den letzten X1-Saisons
137/138	Abstieg2Heim/-Gast	Anzahl der Bundesligasaisons des Heim(Gast)teams in den letzten X2-Saisons
140/141	Abstieg3Heim/-Gast	Anzahl der Bundesligasaisons des Heim(Gast)teams in den letzten X3-Saisons
143/144	Abstieg4Heim/-Gast	Anzahl der Bundesligasaisons des Heim(Gast)teams in den letzten X4-Saisons
146/147	Abstieg5Heim/-Gast	Anzahl der Bundesligasaisons des Heim(Gast)teams in den letzten X5-Saisons

Tabelle 2.6: Aufzählung und Erklärung der Abstieg-Gruppenattribute

2.2.3 Manuell hinzugefügte Attribute

In diesem Unterkapitel beschreibe ich Attribute, die mittels einer externen Datei und externen Quellen **[Kicker]** **[FDat]** **[UEFA]** manuell eingefügt wurden (siehe Kapitel 2.1.2).

Hauptidee hinter den manuell hinzugefügten Attributen

Durch manuell hinzugefügte Attribute kann man eine noch breiter gefächerte Attributenauswahl anbieten. Es können damit einige wichtige Attributbereiche berücksichtigt werden, die bisher von anderen Attributen nicht abgedeckt wurden (z.B. die Anzahl deutscher Nationalspieler im Kader).

<u>Manuell hinzugefügte (ADD)-Attribute</u>		
gelten für Saisons: 2002-2008		
#	Name	Erklärung
94/95	ADDNationalTop9Heim/-Gast	Anzahl der Nationalspieler aus Top9-Ländern beim Heim(Gast)team
96/97	ADDNationalLänderspieleTop9 Heim/-Gast	Anzahl der Länderspiele der Nationalspieler aus Top9-Ländern bei Heim(Gast)team
98/99	ADDNationalTop9PosHeim/-Gast	Anzahl der Nationalspieler aus Top9-Ländern des Heim(Gast)teams geordnet nach Position
100/101	ADDNationalLänderspielerTop9Pos Heim/-Gast	Anzahl der Länderspiele der Nationalspieler aus Top9-Ländern bei Heim(Gast)team geordnet nach Position
102/103	ADDMittelNatlspHeim/-Gast	Durchschnitt aus dem Positionsattribut #98 und #100 (bei Gast Durchschnitt zwischen #99 und #101)
104/105	ADDLsPktHeim/-Gast	Erzielte Punkte des Heim(Gast)teams in der letzten Saison
106/107	ADDGeschosseneToreHeim/-Gast	Geschossene Tore des Heim(Gast)teams in der letzten Saison
108/109	ADDDurchgelasseneToreHeim/-Gast	Durchgelassene Tore des Heim(Gast)teams in der letzten Saison
110/111	ADDPositionHeim/-Gast	Position des Heim(Gast)teams in der letzten Saison
112/113	ADDStadiongroesseHeim/-Gast	Stadiongröße des Heim(Gast)teams
114/115	ADD10AbstiegHeim/-Gast	Anzahl Abstiege des Heim(Gast)teams in der letzten 10 Saisons
116/117	ADDBrazilHeim/-Gast	Anzahl der Brasilianer (nicht unbedingt Nationalspieler) im Kader des Heim(Gast)teams
118/119	ADDBrazilNatHeim/-Gast	Anzahl brasilianischer Nationalspieler im Kader des Heim(Gast)teams
120/121	ADDGermanyNatHeim/-Gast	Anzahl deutscher Nationalspieler im Kader des Heim(Gast)teams
122/123	ADDAustriaHeim/-Gast	Anzahl der Österreicher im Kader des Heim(Gast)teams (nicht unbedingt Nationalspieler)
124/125	ADDAustriaNatHeim/-Gast	Anzahl der österreichischen Nationalspieler im Kader des Heim(Gast)teams
126/127	ADDEwigeTabelleHeim/-Gast	Position des Heim(Gast)teams in der „Ewigen Tabelle“ (stand vor Saison 2008)
128/129	ADDUefa1JahrHeim/-Gast	Position des Heim(Gast)teams in der Uefa-Tabelle des jeweils letzten Jahres [UEFA]
130/131	ADDUefa5JahreHeim/-Gast	Position des Heim(Gast)teams in der Uefa-Tabelle der jeweils letzten 5 Jahre [UEFA]
132/133	ADDPunkteHeimHeim/-GastGast	Punkte des Heim(Gast)teams in der Heim(Gast)tabelle der letzten Saison

Tabelle 2.7: Aufzählung und Erklärung der manuell hinzugefügten (ADD-)Attribute

Zur Verdeutlichung erkläre ich noch kurz aus welchen Ländern sich eine Top9 zusammensetzt. Top 9-Länder wurden von mir auf Basis meiner Erfahrungen und auf Basis der Ergebnisse der jeweiligen Nationalmannschaften zusammengestellt. Die Top 9 lautet folgendermaßen: 1. Deutschland, 2. Argentinien, 3. Brasilien, 4. England, 5. Spanien, 6. Italien, 7. Frankreich, 8. Portugal, 9. Niederlande.

2.2.4 Aus anderen Attributen kombinierte oder neu erstellte Attribute

In diesem Unterkapitel will ich Attribute beschreiben, die auf Basis schon vorhandener Attribute neu erstellt oder kombiniert wurden.

Hauptidee hinter den aus anderen Attributen kombinierten bzw. neu erstellten-Attributen:

Ziel war es Abhängigkeiten der schon vorhandenen Attribute festzustellen und für Weka zu „betonen“. Ferner war es ebenso ein Ziel mithilfe von einigen Attributen zu experimentieren und die Wirksamkeit einiger Lerntechniken zu überprüfen.

Noch eine Anmerkung vorab: mit Testattributen sind die Attribute (in der Tabelle 2.8) gemeint, bei denen schwächere Ergebnisse in Kauf genommen wurden um vor allem die Wirksamkeit der Lernmethoden bei diesem Attribut zu testen (Beispielsweise war es interessant zu prüfen ob ein Wrapper es schafft dieses Attribut herauszufiltern bzw. schlecht zu bewerten).

Kombinierte bzw. auf Basis von anderen Attributen neu erstellte-Attribute		
gelten für Saisons: 1963-heute		
#	Name	Erklärung
18	Jahreszeit	<p>Hier wurde das Attribut „Monat“ in 4 Jahreszeiten unterteilt:</p> <ol style="list-style-type: none"> 1. „kalte Monate“ (Nov., Dez., Jan., Febr.) 2. „Monate mit durchschnittlicher Temperatur“ (März, April, Okt.) 3. „warme Monate“ (Rest)
20	Tabtendenz	<pre>if (\$stabplatzdiff<=-6) {\$stabtendenz=1;} elseif (\$stabplatzdiff<=-1) {\$stabtendenz=2;} elseif (\$stabplatzdiff<=5) {\$stabtendenz=3;} else {\$stabtendenz=4;}</pre> <p>Das Attribut versucht also große Tabellenplatzunterschiede zu erkennen</p>
21	TabDiffForm	<p>Das Attribut versucht Tendenzen aus Plusminusattribut (#4 bzw. #11) zu erkennen. Dabei sind für dieses Attribut einfach nur die Fälle interessant, wenn ein Team sich am letzten Spieltag verschlechtert und das andere Team sich gleichzeitig verbessert hat.</p> <pre>d.h. also: if (\$tabdiffheim>=0 && \$tabdiffgast<0) {\$tabdiffform=1;} elseif (\$tabdiffheim<0 && \$tabdiffgast>=0) {\$tabdiffform=2;} , else {\$tabdiffform=0;}</pre> <p>Dabei ist \$tabdiffheim = „Plusminusheim“ (Attr. #4) und \$tabdiffgast = „Plusminusgast“ (Attr. #11)</p>
22	TabDiffFormVerg	<p>Ähnlich zu Tabdiffform, jedoch werden hier die Vereine direkt miteinander verglichen</p> <pre>if (\$tabdiffheim>\$tabdiffgast) {\$tabdiffformverg=1;} elseif (\$tabdiffheim==\$tabdiffgast) {\$tabdiffformverg=0;} elseif (\$tabdiffheim<\$tabdiffgast) {\$tabdiffformverg=2;}</pre>

28	TorDiffForm	<p>Ähnlich wie Tabdiffform (#21) nur mit Tordifferenz.</p> <p>Wenn Team A also positive Tordifferenz hat und gleichzeitig Team B negative Tordifferenz hat usw.</p> <pre>if (\$tordiffheim>=0 && \$tordiffgast<0) {\$tordiffform=1;} elseif (\$tordiffheim<0 && \$tordiffgast>=0) {\$tordiffform=2;} else {\$tordiffform=0;}</pre>
29/30	TFHeim/-Gast	Wenn die Heim(Gast)mannschaft Tabellenführer ist, dann TFHeim(Gast)=1, sonst 0
31/32	Heimstark/Gaststart	Traditionell heimstarke (auswärtsschwache) Teams (manuell ausgewählt) erhalten eine 1, andere Teams 0
33	SiegeHeimTest	<p><i>Testattribut</i>, sobald die Anzahl der Heimsiege eine bestimmte Zahl nicht uebersteigt, SiegeHeimTest=0, else SiegeHeimTest=1</p> <p>Sinn dieses Attributs ist es sehr heimschwache Teams zu kennzeichnen</p>
49	TabGast	Versucht einen "sichereren Heimsieg" herauszufinden, ab einem bestimmten Tabellenplatz des Gastteams ist Tabgast=1, sonst 0
50	Tabvergleich	Vergleicht den Tabellenplatz beider Mannschaften, wobei bei gleichem Tabellenplatz (soll es oft nach 1.Spieltag usw. geben) Heimteam den Vorzug hat
		<pre>if (\$tabelleheim<\$tabellegast) {\$tabvergleich=1;} elseif (\$tabelleheim>\$tabellegast) {\$tabvergleich=2;} else {\$tabvergleich=1;}</pre>
76	VieleTore	<p><i>Testattribut</i>, Hier habe ich die Differenz: "Tordifferenz-Punkte" berechnet. Wenn die Tordifferenz die Punkte deutlich übersteigt gewinnt dieses Team oft sehr hoch (gute Moral)</p>
92/93	HaeufigstePlazierungHeim/-Gast	Dieses Attribut zählt auf welchem Tabellenplatz sich das Heim(Gast)team am häufigsten in der letzten Saison aufgehalten hat. Bei Gleichheit der Häufigkeit zweier Tabellenplätze, wird der am spätesten belegte genommen

Tabelle 2.8: Aufzählung und Erklärung von kombinierten bzw. auf Basis von anderen neu erstellten Attributen

2.2.5 Differenzen

Zum Schluß kommen nun die Differenzen. Mit den Differenzen sind die Attribute gemeint, die sich aus der Differenz von Heim und Gast-Attributen des gleichen Typs ergeben (z.B. PunkteHeim-PunkteGast=DiffPunkte).

Hauptidee hinter den Differenz-Attributen

Wenn Weka ein Attribut bzw. eine Zahl statt 2 verschiedenen Attributen bzw. Zahlen als Maßstab vorliegt, kann das Programm damit leichter umgehen und Tendenzen erkennen. Außerdem ist durch diese Differenz eine leichte Tendenz des Attributs erkennbar (Bsp. Wenn FormHeim wert von 3 hat, FormGast aber den Wert von 9, dann sieht man durch die -6, dass der Gast eine bessere Form vorweist).

Differenz-Attribute		
gelten für Saisons: Unterschiedlich, alle ADD-Differenzen gelten für 2002-heute, die restlichen Differenzen gelten für 1963-heute		
#	Name	Erklärung
19	Tabdiff	Differenz der aktuellen Tabellenplätze des Heim- und Gastteams
23	Siegdiff	Differenz der aktuellen Anzahl der Siege des Heim- und Gastteams
24	Remisdiff	Differenz der aktuellen Anzahl der Remis des Heim- und Gastteams
25	Niederlagendiff	Differenz der aktuellen Anzahl der Niederlagen des Heim- und Gastteams
26	Punktediff	Differenz der aktuellen Anzahl der Punkte des Heim- und Gastteams
27	TDdiff	Differenz der aktuellen Tordifferenz des Heim- und Gastteams
36,39,42, 45,48	VersusX (X ∈ 1..5)	Differenz von VersusXHeim und VersusXGast (X ∈ 1..5)
55,60,65, 70,75	DiffFormX (X ∈ 1..5)	Differenz von FormXHeim und FormXGast (X ∈ 1..5)
79,82,85, 88,91	DiffAvgX (X ∈ 1..5)	Differenz von AvgXPlatzHeim und AvgXPlatzGast (X ∈ 1..5)
136,139, 142,145, 148	DiffAbstiegX (X ∈ 1..5)	Differenz von AbstiegXHeim und AbstiegXGast (X ∈ 1..5)
149	DiffPunkteHeimGast	Differenz der Attribute AddPunkteHeimHeim und AddPunkteGastGast
150	DiffPktLs	Differenz von AddLsPktHeim und AddLsPktGast
151	DiffGstLs	Differenz von ADDGeschosseneToreHeim und ADDGeschosseneToreGast
152	DiffKassTls	Differenz von ADDDurchgelasseneToreHeim und ADDDurchgelasseneToreGast
153	DiffPosLs	Differenz von ADDPositionHeim und ADDPositionGast
154	DiffEwigeTabelle	Differenz von ADDEwigeTabelleHeim und ADDEwigeTabelleGast
155	DiffNat	Differenz von ADDNationalTop9Heim und ADDNationalTop9Gast
156	DiffNat2	Differenz von ADDNationalländerSpieleTop9Heim und ADDNationalländerSpieleTop9Gast
157	DiffNatPos	Differenz von ADDNationalTop9PosHeim und ADDNationalTop9PosGast
158	DiffNat2Pos	Differenz von ADDNationalländerSpieleTop9PosHeim und ADDNationalländerSpieleTop9PosGast
159	DiffNatPosMit	Differenz von ADDMittelNatlspHeim und ADDMittelNatlspGast
160	DiffStadGr	Differenz von ADDStadiongroesseHeim und ADDStadiongroesseGast
161	Diff10Abstieg	Differenz von ADD10AbstiegHeim und ADD10AbstiegGast
162	DiffBrazil	Differenz von ADDBrazilHeim und ADDBrazilGast
163	DiffBrazilN	Differenz von ADDBrazilNatHeim und ADDBrazilNatGast
164	DiffGermanyN	Differenz von ADDGermanyNatHeim und ADDGermanyNatGast
165	DiffAustria	Differenz von ADDAustriaHeim und ADDAustriaGast
166	DiffAustriaN	Differenz von ADDAustriaNatHeim und ADDAustriaNatGast
167	DiffUefa1J	Differenz von ADDUefa1JahrHeim und ADDUefa1JahrGast
168	DiffUefa5J	Differenz von ADDUefa5JahreHeim und ADDUefa5JahreGast
169	DiffHaeufPlaz	Differenz von HaeufigsteplatzierungHeim und HaeufigstePlatzierungGast

Tabelle 2.9: Aufzählung und Erklärung der Differenz-Attribute

Somit stehe ich am Ende des 2.Kapitels und komme nun zum Kapitel 3. Dieses Kapitel beinhaltet eine knappe Erklärung der fachlichen Grundlagen.

Kapitel 3: Fachliche Grundlagen

In diesem Kapitel habe ich mir als Ziel gesetzt dem Leser die Funktionsweise aller in der Diplomarbeit verwendeter Algorithmen und Klassifizierer in einer kurzen Zusammenfassung zu vermitteln. Das Kapitel unterteilt sich in 4 Teile. Im ersten Teil beschreibe ich die Funktionsweise der verwendeten Klassifizierer, um danach im zweiten Teil die verwendeten Lerntechniken zu erklären. Im dritten Teil erkläre ich die verschiedenen angewendeten Evaluierungsmethoden, bevor ich mit einer kurzen Erläuterung des Overfittings abschließe.

3.1. Beschreibung der verwendeten Klassifizierer

In diesem Unterkapitel erkläre ich die verschiedenen in der Diplomarbeit angewendeten Klassifizierer. Dabei geht es nach einer kurzen Einleitung konkret um: Naive Bayes, JRip-Regellerner, J48, Support Vector Machines und Ensemble Methoden (genauer gesagt Bagging, Boosting und Random Forest).

3.1.1 Einleitung

(Die Informationen in diesem Unterkapitel basieren auf den Büchern: Tom M. Mitchell, *Machine Learning* [Mit97] und I.H. Witten, E.Frank. *Data Mining Practical Machine Learning Tools and Techniques, 2nd Edition* [WIFR05])

Klassifikation ist eine systematische Einordnung in Klassen (Brockhaus) [Brockh97]

Nimmt man die Brockhaus-Definition als Maßstab, kann man erkennen, dass es bei Klassifizierung um Zusammenfassung bestimmter Objekte nach einer bestimmten vorgegebenen Form geht. Diese bestimmte vorgegebene Form sind die Klassifizierer. Die Werkzeuge von diesen können da durchaus unterschiedlich sein. Sowohl mathematische Funktionen, oder Aufstellen von verschiedenen Regeln als auch Aufbau von Entscheidungsbäumen oder Benutzung von Ensemble-Methoden kommen als Werkzeuge in Frage.

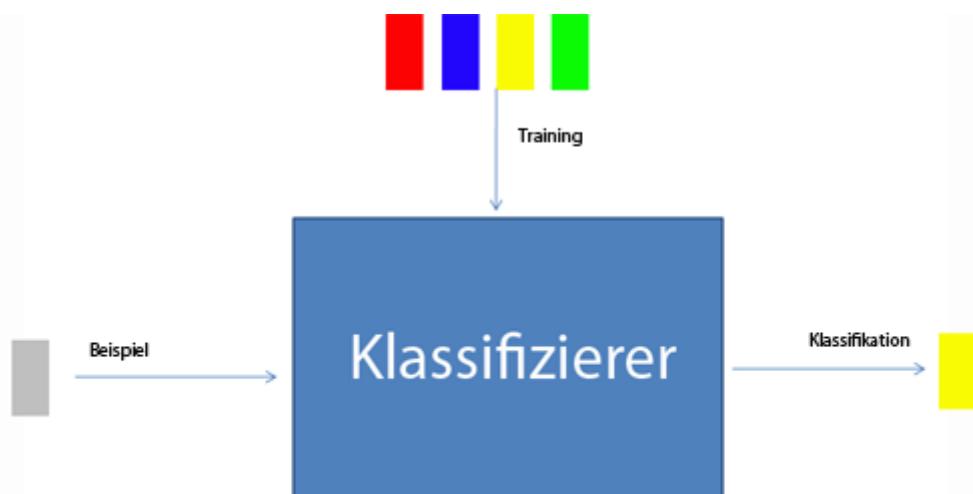


Bild 3.1: Skizze eines Klassifizierers

Dabei steht jedoch meistens eine zentrale Aufgabe im Vordergrund:

Lerne ein Modell, dass das Ergebnis einer abhängigen Variablen für eine gegebene Instanz vorhersagt. Dabei steht dem Lerner eine größere Menge an „Erfahrung“ zur Verfügung. Diese Erfahrung definiert sich in gegebenen Beispielen, wo die gesuchte Instanz durch Vorherige bereits stattgefundene Ereignisse bereits ein Ergebnis vorweist bzw. durch ein Training diese Beispiele bereits analysiert hat. (siehe Bild 3.1)

3.1.2 Naive Bayes

(Die Informationen zu Bayes-Klassifizierern und Naive Bayes basieren hauptsächlich auf dem Buch: J.Han, M.Kamber. *Data Mining Concepts and Techniques*, 2nd Edition [HanK06], an einigen Stellen wurden ebenso Informationen aus den Büchern: Tom M. Mitchell, *Machine Learning* [Mit97] und I.H. Witten, E.Frank. *Data Mining Practical Machine Learning Tools and Techniques*, 2nd Edition [WIFR05] verwendet)

Bayes-Klassifizierer:

Bayes-Klassifizierer gehören zu der Klasse der statistischen Klassifizierer. Sie können die Wahrscheinlichkeit einer Klassenzugehörigkeit vorhersagen. Genauer gesagt, für in meiner Diplomarbeit zutreffenden Fall, können die Bayes-Klassifizierer die Wahrscheinlichkeit, dass ein Tupel an Werten zu einer bestimmten Klasse gehört, vorhersagen.

Bayes Klassifizierer basieren auf dem Bayes Theorem, das weiter unten aufgeführt wird. Ein simpler Algorithmus aus der Klasse der Bayes-Klassifizierer, den ich bei meiner Diplomarbeit verwende ist Naive Bayes.

Bayes-Theorem:

Wie bereits erwähnt, basieren die Bayes-Klassifizierer (u.a. in der Diplomarbeit verwendeter Naive Bayes Klassifizierer) auf dem Bayesschen Theorem. Um dieses Theorem einzuführen, müssen noch einige Vordefinitionen getätigkt werden. Sei H eine Hypothese und D ein Datentupel, das zu einer bestimmter Klasse K gehört. Dann lautet das Bayes-Theorem:

$$P(H | D) = \frac{P(D | H)P(H)}{P(D)} \quad (3.1)$$

Zur Verdeutlichung eine Legende:

$P(H)$ = A priori Wahrscheinlichkeit der Hypothese H

$P(D)$ = A priori Wahrscheinlichkeit des Datentupels bzw. der Trainingsdaten D

$P(H|D)$ = Wahrscheinlichkeit von H *unter der Voraussetzung*, dass D gegeben ist (A posteriori Wahrscheinlichkeit)

$P(D|H)$ = Wahrscheinlichkeit von D *unter der Voraussetzung*, dass H gegeben ist

Beispiel

Nehmen wir nun folgende Datenbelegungen an:

-Die Hypothese ist: „Heimteam gewinnt“, also eine Tendenz von „1“

-D also unser Datentupel, der aus folgenden 3 Attributen besteht: „Name des Heimvereins“, „Position des Heimvereins“ und „Punkte aus dem letzten Spiel des Heimvereins“ und ist in meinem konkreten Beispiel mit „Eintr. Frankfurt“, „13“, „0“ belegt.

Dann erhält man nun basierend auf dem Bayes Theorem folgende Formel:

$$P(1 \mid \text{Eintr. Frankfurt}, 13, 0) = \frac{P(\text{Eintr. Frankfurt} \mid 1) P(13 \mid 1) P(0 \mid 1) P(1)}{P(\text{Eintr. Frankfurt}, 13, 0)} \quad (3.2)$$

In Worten gesagt bedeutet diese Formel: Die Wahrscheinlichkeit davon, dass im Spiel mit dem Heimverein Eintracht Frankfurt, der auf dem 13. Platz steht und 0 Punkte aus dem letzten Spiel hat ein Heimsieg rauspringt, berechnet sich aus der Einzelwahrscheinlichkeiten (jeweils einzeln für „Eintracht Frankfurt“, 13.Platz und 0 Punkte) für bisherige Heimsiege multipliziert mit der Gesamtwahrscheinlichkeit für einen Heimsieg und dividiert mit der Wahrscheinlichkeit für das Auftreten des kompletten Gesamtereignisses (Eintracht Frankfurt, 13, 0).

Naive Bayes

Nun komme ich zu Naive Bayes, einem Algorithmus, der auf dem Bayesschen Theorem basiert und der Klasse der Bayes-Klassifizierer zugeschrieben wird. Das Naive an diesem Algorithmus ist (woraus auch der Name abgeleitet wird), dass der Algorithmus davon ausgeht, dass alle Wahrscheinlichkeiten bzw. Ereignisse innerhalb einer Klasse voneinander unabhängig sind. Natürlich ist diese Annahme sehr gewagt (als Gegenbeispiel kann man beispielsweise die Attribute: „Position des Heimvereins“ und „Punkte des Heimvereins“ betrachten. Mit einer höheren Position erreicht der Heimverein mehr Punkte bzw. mit mehr Punkten darf man von einer höheren Position des Heimvereins ausgehen).

Naive Bayes ist für mittlere und große Trainingsmengen am besten geeignet und wird aufgrund der einigermaßen guten Ergebnisse und aufgrund der guten Geschwindigkeitseigenschaften oft vor allem bei der Klassifikation von Textdokumenten verwendet.

Umsetzung

Um die Umsetzung des Naive Bayes zu erklären müssen noch einige Definitionen getätigt werden:

1. Angenommen T ist ein Trainingsset von Datentupeln (mit Elementen $t_1, t_2 \dots t_n$) und ihren zugeordneten Werten. Somit wäre in unserem Fall T eine komplette Weka-Zeile mit sämtlichen Attributbelegungen (d.h. eintrachtfrankfurt, 1,10,10...,0).
2. Nehmen wir ebenso an, dass K die Klassen (mit Elementen $K_1, K_2 \dots K_i$) sind. In unserem Fall gibt es 3 Klassen und zwar die möglichen Tendenzen „0“, „1“ und „2“.
3. Wir wenden das oben eingeführte Bayes-Theorem ein:

$$P(K \mid T) = \frac{P(T \mid K)P(K)}{P(T)} \quad (3.3)$$

Ziel ist es die Wahrscheinlichkeit $P(K \mid T)$ aus dem Ergebnis zu maximieren. Dabei braucht man nur den Zähler zu maximieren, da der Nenner immer konstant bleibt und somit ignoriert werden kann.

Nun wird der Zähler also etwas genauer betrachtet:

3.1. Für jede einzelne Klasse K_i wird $P(T|K)$ berechnet. Dies geschieht mit der Aufeinander-Multiplikation der Einzelwahrscheinlichkeiten für jedes Element t_k aus dem Datentupel. Die Formel hierfür ist also:

$$P(T|K_i) = \prod_{k=1}^n p(t_k|K_i) \quad (3.4)$$

$$\text{d.h. also } P(T|K_i) = p(t_1|K_i) p(t_2|K_i) \dots p(t_n|K_i) \quad (3.5)$$

3.2 Das Ergebnis aus 3.1 wird mit der Gesamtwahrscheinlichkeit für die Klasse multipliziert, also $P(K)$.

4. Diese Prozedur wird nun für jede Klasse K_i durchgeführt und die Klasse mit der maximalen Wahrscheinlichkeit dem betrachteten Datentupel zugewiesen.

Beispiel

Gegeben sind Klassen: „0“, „1“ und „2“

Ein Datentupel besteht aus „Name des Heimvereins“ (NdHeim), „Position des Heimvereins (PositionH)“, „Punkte aus dem letzten Spiel des Heimvereins“ (PLSH),

Folgende Wahrscheinlichkeiten sind gegeben:

$$P(0) = 2/11$$

$$P(1) = 3/11$$

$$P(2) = 6/11$$

$$P(NdHeim=Eintracht Frankfurt/0) = 1/13$$

$$P(NdHeim=Eintracht Frankfurt/1) = 2/13$$

$$P(NdHeim=Eintracht Frankfurt/2) = 9/13$$

$$P(PositionH=13/0) = 4/13$$

$$P(PositionH=13/1) = 3/13$$

$$P(PositionH=13/2) = 6/13$$

$$P(PLSH=0/0) = 5/13$$

$$P(PLSH=0/1) = 7/13$$

$$P(PLSH=0/2) = 1/13$$

Nun betrachte und klassifiziere ich folgendes Beispiel mit Naive Bayes:
(Eintracht Frankfurt, 13, 0)

Folgende Berechnungen erfolgen nun:

$$P(0|Eintracht Frankfurt, 13, 0) = 1/13 * 4/13 * 5/13 * 2/11 = 22/24167$$

$$P(1|Eintracht Frankfurt, 13, 0) = 2/13 * 3/13 * 7/13 * 3/11 = 126/24167$$

$P(2|Eintracht Frankfurt, 13, 0) = 9/13 * 6/13 * 1/13 * 6/11 = 324/24167$ (höchster Wert von den 3 Klassen)

Somit ist die Vorhersage für das Heimteam Eintracht Frankfurt, auf 13.Position und mit 0 Punkten aus dem letzten Spiel eine 2 (Also eine Heimniederlage der Eintracht).

Problem von Naive Bayes

Das größte Problem von Naive Bayes wird offensichtlich wenn mindestens ein Datum aus den Datentupeln (d.h. ein t_k) fehlt oder gleich 0 ist. Denn in diesem Fall ergibt die gesamte Multiplikation ein Ergebnis von 0 und führt damit zu einer Nichtberücksichtigung von vielen anderen möglicherweise guten Datenwerten. Dieses Problem kann auf mehrere Weisen gelöst werden. Die einfachste Methode zur Lösung, die auch als Standardtechnik verwendet wird, ist eine Laplace-Schätzung. Diese addiert einfach jedem Zähler eine 1 auf und teilt dann (im Nenner) durch die Gesamtanzahl der in allen Attributen aufaddierten „Einser“ im Nenner.

3.1.3: JRip

(die Informationen zu Ripper und JRip basieren auf dem Buch: I.H. Witten, E.Frank. *Data Mining Practical Machine Learning Tools and Techniques, 2nd Edition [WIFR05]*)

Kommen wir nun zu den Regellernern. Regellerner stellen, wie der Name schon sagt, nach bestimmten Mustern Regeln auf, nach denen Sie dann versuchen zu klassifizieren. Dabei können sowohl einzelne Regeln als auch Regelmengen aufgestellt werden und verschiedene Pruning-Techniken angewendet werden um irrelevante Informationen „abzuschneiden“.

Ein Beispiel einer Regel bzw. einer Regelmenge wäre (entnommen aus dem Weka-Programm):

1. $(PUNKTEHEIM \leq 19) \text{ and } (ADDSTADIONGROESSEHEIM \geq 44308) \text{ and } (REMISHEIM \geq 4) \Rightarrow \text{tendenz}=0$
 $(62.0/28.0)$
2. $(DIFFAVG3 \geq 2) \text{ and } (DIFFUEFA1J \leq -4) \text{ and } (DIFFAVG4 \geq 11.7) \Rightarrow \text{tendenz}=2$ $(62.0/24.0)$
3. $\Rightarrow \text{tendenz}=1$ $(794.0/413.0)$

(die Bedeutung der einzelnen Attribute kann man im Kapitel 2.2 finden)

In diesem Unterkapitel betrachte ich den JRip-Algorithmus, der auf dem Regellernalgorithmus Ripper basiert. Ripper verwendet incremental reduced-error Pruning (I-REP). I-REP hat die Möglichkeit eine betrachtete Regel, im Fall der Überflüssigkeit oder Unbedeutenheit sofort nach dem Betrachten wegschneiden zu können (in Gegensatz zu dem Reduced Error Pruning-Algorithmus, wo hierfür erstmal die komplette Regelmenge untersucht werden muss).

Da es von „Ripper“ viele mögliche Ausprägungen und Implementierungen gibt, werde ich mich in meiner Diplomarbeit nur auf die Grundidee beschränken.

Die Vorgehensweise von Ripper

Der Ripper-Algorithmus kann in 4 grobe Phasen aufgeteilt werden, diese sind 1. Aufbau (BUILD) 2. Optimierung (OPTIMIZE) 3. Prüf-Phase (MOP UP) und 4. Lösch-Phase (CLEAN UP). Dabei werden die Klassen in aufsteigender Reihenfolge untersucht (d.h. also beginnend mit der kleinsten Klasse mit den wenigsten Instanzen und endend mit der größten Klasse). An einigen Stellen wird im Algorithmus die Description Length gebraucht. Description Length ist ein Verfahren, dass im Großen und Ganzen mit Hilfe einer komplexeren Formel, die Anzahl der Bits überprüft, die die Beispieldaten sowie die Regelmenge benötigt.

Kommen wir nun zu der Beschreibung der einzelnen Phasen:

1. Aufbau-Phase: Generiere für jede betrachtete Klasse eine Regelmenge mit Hilfe des I-Rep.
2. Optimierungs-Phase: Nachdem der Aufbau von Regelmengen für jede Klasse erfolgt ist, wird jede Regel noch einmal einzeln betrachtet. In dieser Phase werden jedoch alle Instanzen

entfernt die von anderen Regeln für diese Klasse abgedeckt werden. In anderen Worten: Als Pruningkriterium zählt nur der Erfolg auf den Instanzen, die die Regel selbst abdeckt. In dieser neuen Datenmenge können sowohl komplett neue Regeln entstehen, als auch ältere Regeln durch kleine Ergänzungen verändert werden. Beide Varianten werden einander gegenübergestellt und die optimalere Variante wird ausgewählt.

3. Prüf-Phase: Sollten nach dem 2. Schritt von Regeln unbedeckte Instanzen verbleiben, kehre zum 1.Schritt, also zur Aufbau-Phase zurück und baue auf diesen Instanzen neue Regeln auf.
4. Lösch-Phase: Hier wird die Description Length sowohl für das ganze Regelset, das in den ersten 3 Phasen konstruiert wurde, als auch für jede Regel aus diesem Regelset im Einzelnen berechnet. Sollte hier eine Regel die Description Length erhöhen wird diese gelöscht.

3.1.4 J48

(Das Unterkapitel zu J48, sowie zu ID3 und C4.5 wurde auf Basis der Informationen aus den Büchern: Tom M. Mitchell, *Machine Learning* [Mit97] und I.H. Witten, E.Frank. *Data Mining Practical Machine Learning Tools and Techniques, 2nd Edition* [WIFR05], sowie auf Basis des Artikels von Prof. G. Ingargiola [[Ing](#)] erstellt)

J48-Algorithmus basiert auf dem C4.5 Entscheidungsbaum-Lernverfahren, dass seinerseits eine Weiterentwicklung vom ID3-Algorithmus zur Generierung von Entscheidungsbäumen ist.

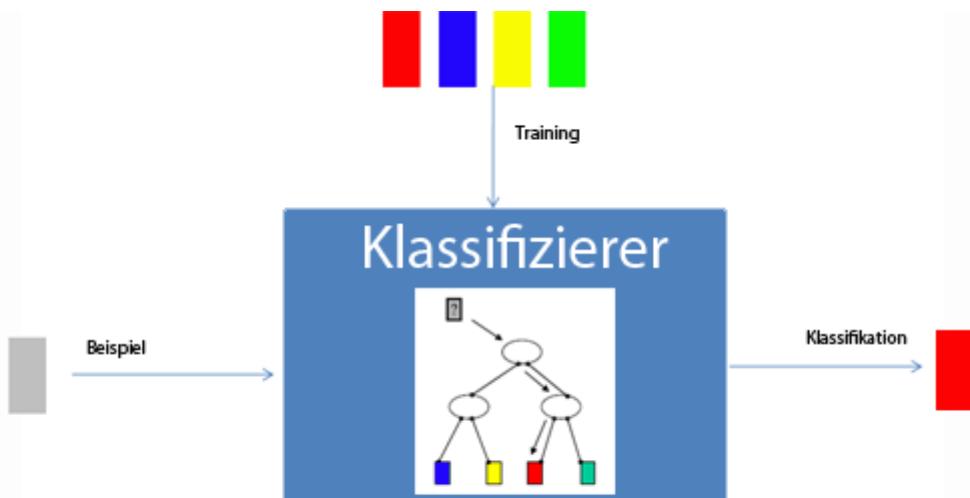


Bild 3.2: Klassifikation mit Entscheidungsbäumen

Entscheidungsbäume

Zunächst aber eine kurze Beschreibung der Entscheidungsbäume. Entscheidungsbäume ist eine Lösungsform für bestimmte Probleme, bei der die Knoten einen bestimmten Test für einen Attribut darstellen, die Kanten einerseits eine Verbindung zum nächsten Knoten liefern, sowie einem Ergebnis des vorangegangenen Tests entsprechen und die Blätter die Ergebnisse des Klassenattributs beinhalten. Somit muss man um ein Beispiel zu klassifizieren im Wurzelknoten starten, den Test darin ausführen und der Kante mit dem Ergebnis zum nächsten Knoten folgen. Diese Prozedur wird bis zum Erreichen des Blattknotens mit dem Ergebnis durchgeführt. Das Beispiel wird dann mit dem Inhalt des Blattknotens klassifiziert (siehe auch Bild 3.2).

ID3-Algorithmus

Da C4.5 auf ID3-Algorithmus basiert, beginne ich erst einmal mit der Beschreibung des ID3-Algorithmus um später einige der wichtigsten Erweiterungen des C4.5 gegenüber ID3 aufzuführen.

Angenommen T sei unser (nichtleeres) Trainingsset und die Beispiele im Trainingsset gehören zu verschiedenen Klassen $K_1, K_2 \dots K_i$ (anderenfalls hätten wir nur einen Knoten).

1. Wähle Knoten A nach einer bestimmten heuristischen Funktion (in ID3: GainRatio, siehe Kapitel 3.2.2)
2. A ist nun ein Testknoten, Generiere nun anhand derselben heuristischen Funktion Entscheidungsbäume für alle Ausgänge (Kanten) von A und verbinde die Wurzel dieser Entscheidungsbäume mit A mit Kanten.

Erweiterungen von C4.5 gegenüber ID3 [Ing]

C4.5 bietet viele Erweiterungen gegenüber ID3, einige wichtigsten davon sind:

- C4.5 kann neben diskreten auch mit kontinuierlichen Werten umgehen, somit kann C4.5 auch untere und obere Schranken erstellen um mit diesen Attributen zu arbeiten
- C4.5 kann mit fehlenden Attributen umgehen. Diese Attribute werden (vorausgegangen Sie sind entsprechend markiert) bei den Heuristikberechnungen (wie z.B. GainRatio) nicht berücksichtigt.
- C4.5 benutzt Post-Pruning. Somit können nachdem der Entscheidungsbaum erstellt wurde, noch einzelne überflüssige sowie irrelevante Knoten entfernt werden
- C4.5 benutzt als Heuristik GainRatio, während ID3 Information-Gain benutzt

3.1.5 Support Vektor Maschinen (SVM)

(Diese kurze Zusammenfassung über das umfangreiche Thema der Support Vektor Machines basiert auf folgender Literatur: Tom M. Mitchell, *Machine Learning* [Mit97] und J.Han, M.Kamber. *Data Mining Concepts and Techniques, 2nd Edition* [HanK06])

Das Themengebiet der Support Vektor Maschinen umfasst ein sehr breites und komplexes Feld, das den Rahmen der Diplomarbeit sprengen würde. Daher will ich hier nur einen kurzen Einblick in dieses Thema geben.

Support Vektor Maschinen (weiter in der Diplomarbeit nur noch als SVM bezeichnet) verwenden nichtlineares Mapping um die Original-Trainingsdaten in eine höhere Dimension zu transformieren, mit der Behauptung, dass in einer höheren Dimension sich Daten zweier Klassen viel besser separieren lassen. Diese Separation verschiedener Klassen geschieht durch eine Hyperebene. SVM findet diese Ebene mit Hilfe der Support Vektoren und der sogenannten „Margins“, die durch die Support Vektoren aufgespannt werden. Es wird also versucht die Beispielmenge durch eine geeignete Hyperebene so zu separieren, dass die Trainingsbeispiele durch diese Linie in 2 möglichst verschiedene Klassen definiert werden. Dabei werden die am nächsten zur Hyperebene befindende Beispiele „Support Vektoren“ genannt und der Abstand von diesen nächsten Beispielen zur Hyperebene als \vec{w} , sowie der kleinste Abstand von 2 Beispielen verschiedener Klassen als (auf verschiedener Seite der Hyperebene) „Margin“ bezeichnet.

Das Ziel von SVM ist die Hyperebene also so zu legen, dass \vec{w} minimal und der Margin maximal ist (siehe auch Bild 3.3).

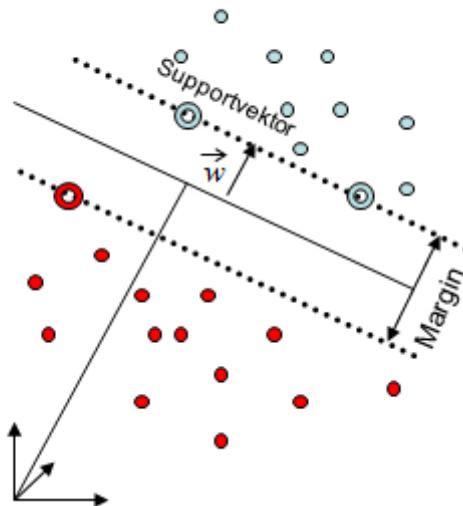


Bild 3.3: Verdeutlichung des Margins und Support Vektoren bei linearer Trennbarkeit

Bei linear trennbaren SVM ist die Hyperebene durch eine einfache Linie dargestellt.

Bei Nicht-linear trennbaren SVM wird versucht eine Funktion zu finden um die Klassen zu trennen. Ebenso werden hier imaginäre Variablen (d.h. die sogenannten Schlupfvariablen) benutzt. Die Funktion der Hyperebene bei nichtlinearer Trennbarkeit heißt Kernel (siehe Bild 3.4).

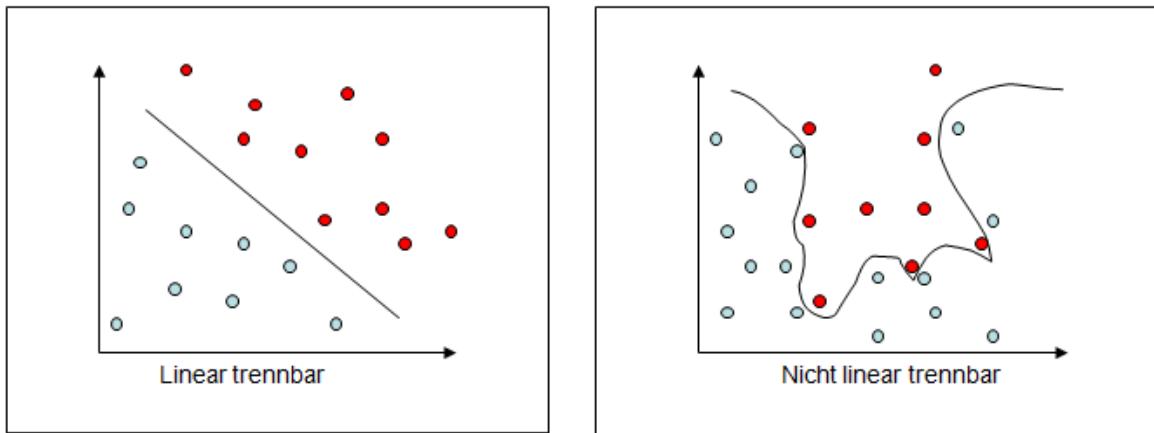


Bild 3.4: Hyperebene der Support Vektor Machines bei linearer und nichtlinearer Trennbarkeit

Bis heute zählen die Support Vektor Maschinen zu einem der genauesten Verfahren zur Klassifikation von verschiedenen Klassen. Eins der größten Nachteile von SVM ist jedoch der enorm hohe Bedarf an Zeit und Ressourcen, was später auch bei den Analysen in meiner Diplomarbeit zu einem größeren Problem wurde.

3.1.6 Ensembles

In diesem Unterkapitel komme ich zu den Ensemble Methoden. Diese zeichnen sich dadurch aus, dass sie mit Kombinationen von verschiedenen Klassifizierern arbeiten. Ensembles setzen sich als Ziel ein optimales Modell mit Hilfe einer Kombination von vorhandenen Klassifizierern zu schaffen. In diesem Unterkapitel betrachte ich 3 verschiedene Ensemble Methoden: 1. Bagging 2. Boosting und 3. Random Forest.

(*Die Informationen zu den nachfolgenden Ensemble-Verfahren „Bagging“ und „Boosting“ basieren auf den Informationen aus den Büchern: J.Han, M.Kamber. Data Mining Concepts and Techniques, 2nd Edition [HanK06] und I.H. Witten, E.Frank. Data Mining Practical Machine Learning Tools and Techniques, 2nd Edition [WIFR05]*)

Bagging:

Bagging steht für „bootstrap aggregation“ und wählt aus vielen verschiedenen Klassifizierer-Ergebnissen das am häufigsten Vorkommende. Dabei verwendet Bagging die Bootstrap-Schätzmethode. Bootstrap basiert darauf, Stichproben aus einer Datenmenge zu entnehmen um die Trainingsdaten zu formen und gezogene Instanzen im Original zu ersetzen [WIFRD01]

Bagging Algorithmus:

1. Für jede Iteration: Ziehe ein Bootstrap-Sample aus den Daten und lerne einen Klassifizierer K_i darauf. Man erhält also nach diesem Durchlauf nach i Iterationen i Klassifikatoren
2. Für jedes Beispiel sollte nun jeder der im 1.Schritt erhaltenen Klassifikatoren ausprobiert werden.
3. Die Klasse, die die meisten Stimmen von den Klassifizierern im Schritt 2 erhalten hat gewinnt.

Ein Klassifizierer, der mit Bagging bearbeitet wurde, hat oft eine größere Genauigkeit, als ein einzelner Klassifizierer, da das zusammengesetzte Modell oft die Varianz zwischen den verschiedenen Klassifizierern senkt. Außerdem ist ein Attribut mit Bagging stabiler gegenüber verrauschten Daten.

Boosting:

Im Gegensatz zum Bagging-Verfahren benutzt das Boosting-Verfahren Gewichtungen für die Klassifizierer. Die Gewichte sind an jedes Datentupel gebunden und werden mit jedem Durchgang neu aktualisiert. Die Idee ist, dass die Gewichte nach einer Klassifizierung mit dem Klassifizierer K_i so verändert werden, dass der nächst kommende Klassifizierer K_{i+1} automatisch die Gewichte genauer betrachtet die K_i falsch klassifiziert hat. Der finale Klassifizierer kombiniert die Stimmen von jedem Klassifizierer. Ein sehr bekannter Boosting-Algorithmus ist AdaBoost.

Boosting-Algorithmus (am Beispiel von AdaBoost):

1. Initialisiere alle Gewichte, in dem am Anfang alle Klassifizierer gleich gewichtet werden
2. Benutze den ersten Klassifizierer, prüfe dort jedes Beispiel und mache eine Fehlerabschätzung auf dem Ergebnis der Klassifikation. Senke die Gewichtung aller korrekt klassifizierten Beispiele, steigere die Gewichtung aller inkorrekt klassifizierten Beispiele
3. Benutze den (die) nächsten Klassifizierer weiter wie in Schritt 2.
4. Sage die Klasse mit der höchsten Gewichtung am Ende voraus.

Boosting kann eine noch bessere Genauigkeit erzielen als Bagging und somit eine bessere Genauigkeit als ein einzelnes Modell erzielen. Jedoch ist ein Boosting-Modell, durch seine zu große Berücksichtigung von missklassifizierten Daten mehr Overfitting (siehe Kapitel 3.2.1)-gefährdet.

Random Forest

(Das Unterkapitel zu Random Forest basiert auf den Informationen aus dem Buch: T.Hastie, R.Tibshirani, J.Friedman. *The Elements of Statistical Learning: Data Mining, Inference and Prediction, 2nd Edition [HATF08]*)

Random Forest (RF) arbeitet, wie der Name schon sagt, mit Bäumen und ist eine substanzielle Modifikation des Bagging-Verfahrens. Als Ensemble-Methode verwendet RF im Schlusschritt das bereits erwähnte Bagging Verfahren. Die Variablen für einen Split in dem Knoten werden zufällig ausgewählt.

Random Forest-Algorithmus

1. Ziehe ein Bootstrap-Sample Z der Größe N aus den Trainingsdaten
2. Erstelle auf Basis von Z einen Random Forest Baum T_b mit Hilfe folgender Schritte:
 - 2.1 Wähle zufällig m von p möglichen Variablen.
 - 2.2 Suche aus m die bestmögliche Variable für den bestmöglichen Split aus
 - 2.3 Splitte den Vaterknoten anhand der bestmöglichen Variablen in 2 Äste (linker Ast mit Werten kleiner als die Variable und rechter Ast mit Werten größer als die Variable).
 - 2.4 Solange das Stopkriterium nicht erreicht ist, wende die Schritte 2.1-2.3 auf den neu entstandenen linken und rechten Ast des Baums an
3. Wiederhole Schritte 1-2 mit neuen Z , solange möglich bzw. notwendig
4. Aggregiere die B so erhaltenen Bäume mit Bagging zu einer Lösung und gebe die Lösung aus

Somit stellt Random Forest eine einfache und sehr effiziente Ensemble-Methode dar, die gute Ergebnisse liefert und einfach zu trainieren und einzustellen ist. **[HATF08]** Jedoch verwendet die Random Forest-Methode bei der Erstellung vom zufälligen Bäumen kein Pruning.

Hiermit schließe ich das Klassifizierer-Unterkapitel ab und komme im nächsten Kapitel zu den Verfahren zur Attribut- bzw. Featureauswahl, die während dieser Diplomarbeit verwendet wurden.

3.2 Erklärung der Verfahren zur Featureauswahl

In diesem Unterkapitel komme ich zur Erklärung von einem wichtigen Punkt im Bereich des Data Minings, der „Feature Subset Selection“. Ich werde hierfür zuerst einleitend definieren um danach auf die verwendeten Lerntechniken genauer einzugehen und diese zu erklären.

3.2.1 Einleitung

Im vorangegangenen Kapitel wurden dem Leser die verwendeten Attribute gezeigt. Nun ist es so, dass bei 170 Attributen bestimmt einige Attribute dabei sind, die weniger helfen die Tendenz zu erraten und eher sowohl zu störenden, als auch zu übertrieben genauen Daten (Overfitting) führen können. Ferner habe ich auch, wie bereits im vorangegangenen Kapitel gesehen wurde, Gruppenattribute dabei, die viele redundante Informationen beinhalten. Um diese Attribute zu finden und zu entfernen, ist eine Feature Subset Selection notwendig.

Eine Feature Subset Selection beschäftigt sich mit dem Auffinden und Entfernen von irrelevanten, redundanten sowie für die Klassifikation uninteressanten Attributen. Diese Attribute zu entfernen kann sowohl die Genauigkeit verbessern, als auch eine Verhinderung von Overfitting erreichen. Was mit Overfitting genau gemeint gilt es nun zu erklären.

Overfitting

Mit Overfitting wird eine zu starke Annäherung an die Daten bezeichnet. Man kann mit einem Lerner einen Fall ganz genau Klassifizieren und erreicht in ebendiesem Fall einen hohen Klassifikationserfolg, jedoch schließt es durch seine zu genaue Betrachtung eventuell andere, deutlich mehr aussagende, allgemeinere Fälle aus. Sehen wir uns nun Overfitting in einem Beispiel an. Dieses Beispiel wurde absichtlich übertrieben um das Overfitting sehr deutlich zu machen:

Der Verein Eintracht Frankfurt hat in der Saison 2007/2008 ein Heimspiel gegen Bayern München im Monat April verloren, vor dem Spiel hatte Eintracht 42 Punkte und eine Tordifferenz von -2.

Jetzt erstellt ein Regellerner folgende Regel:

(Heim=Eintracht Frankfurt) and (Gast=Bayern München) and (Punkteheim=42) and (Monat=04) and (TDHeim=-2) => tendenz=2

Ist diese Regel gut?

Nun für die Saison 2007/08 ist sie zumindest richtig, auch wenn sie sehr wenig abdeckt. Für die anderen Saisons wird sie nur in größten Ausnahmefällen zutreffen. Eine Regel *(Gast=Bayern München) => tendenz=2* ist viel allgemeiner und nicht so speziell auf den Fall zugeschnitten, wird aber mit Sicherheit über die Jahre mehr Erfolg haben.

Ausgewählte Lerntechniken in der Diplomarbeit

Nun zurück zur Feature Subset Selection. Für die Diplomarbeit wurden folgende Lernverfahren ausgewählt:

-InfoGain, als eine sehr schnelle und gute Möglichkeit die Attribute zu sortieren und zu bewerten.
-Wrapper mit verschiedenen Hilfsmethoden, als eine sehr zuverlässige Methode die Attribute zu bewerten. Ein Wrapper wird in Kombination zu einer Suchmethode benutzt. Folgende Suchmethoden wurden von mir in der Diplomarbeit in der Zusammenarbeit mit einem Wrapper verwendet:

- BestFirst (BackwardElimination)
- RankSearch (Ranking der Attribute mit GainRatio-Evaluator)
- LinearForwardSelection (Extension von einer Best-First Search Forward Elimination)

Nun komme ich zu ausführlicheren Erläuterungen von Feature Ranking Verfahren „InfoGain“ und „GainRatio“.

3.2.2 Feature-Ranking

In diesem Unterkapitel werden Ranking-Verfahren beschrieben. Dabei werde ich erst einmal auf den Information-Gain eingehen und danach wird GainRatio besprochen.

Information-Gain (InfoGain)

(*Informationen zu Information-Gain wurden auf Basis der Literaturquellen: Tom M. Mitchell, Machine Learning [Mit97] und I.H. Witten, E.Frank. Data Mining Practical Machine Learning Tools and Techniques, 2nd Edition [WIFR05] erstellt*)

Information-Gain (InfoGain) benutzt das Konzept der Entropie. Entropie misst die Unreinheit eines Datensatzes. Sollte also der Datensatz bei 2 Klassen aus je Hälfte der Beispiele aus jeder der beiden Klasse bestehen, gilt es als maximal „verunreinigt“ und die Entropie ist in diesem Fall Maximal (=1). Sollte ein Datensatz fast nur aus Beispielen einer Klasse bestehen gilt es als „sauber“ und die Entropie ist minimal (also nahe 0, siehe auch Bild 3.5). Entropie kann aber auch als: „Die Menge der Informationen (Bits), die gebraucht wird um die Klasse eines Beispiels zu erfahren“ bezeichnet werden. Bei einer 50/50 Verteilung zwischen 2 Klassen ist die Menge an gebrauchten Informationen maximal, bei Verteilungen wie 99/1-sehr gering. Da die Entropie-Information in Bits kodiert ist, wird bei der Entropie-Formel die \log_2 Funktion benutzt.

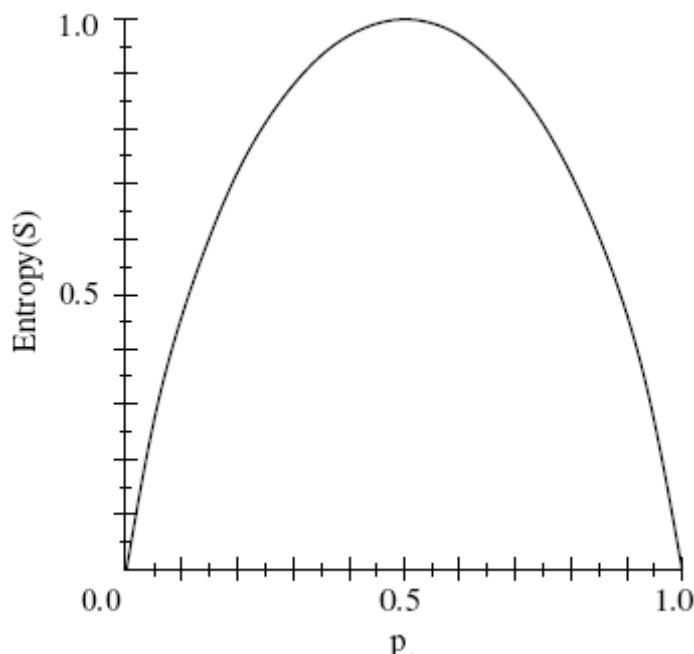


Bild 3.5: Funktion der Entropie

(Diese Graphik wurde aus dem Buch: Tom M. Mitchell, *Machine Learning*, Seite 57 [Mit97] entnommen)

Entropie definiert sich durch die folgende Formel:

$$\text{Entropy}(S) \equiv -p_+ \log_2 p_+ - p_- \log_2 p_- \quad (3.6)$$

Die Formel für Information-Gain ($Gain(S, A)$ mit der oben benutzten Definition der Entropie) ist dann:

$$Gain(S, A) \equiv Entropy(S) - \sum_{v \in Werte(A)} \frac{|S_v|}{|S|} Entropy(S_v) \quad (3.7)$$

Dabei sind:

S : Menge von Trainingsbeispielen

A : ein Test

S_v : Teilmenge der Klasse S , die als Ausgang den Wert v annimmt (z.B. Teilmenge der Klasse Tendenz, die als Ausgang den Wert „1“ annimmt)

p_+ : Beispiele der Klasse +

p_- : Beispiele der Klasse –

Damit ist also Information-Gain definiert als die Differenz der Gesamtentropie aller Klassen und der Summe der jeweils einzelnen Entropien einzelner Klassen (jeweils mit dem Klassenanteil multipliziert).

Da InfoGain sehr oft als die Splittingheuristik bei Entscheidungsbäumen benutzt wird, kann man auch die Formel auch folgendermaßen erklären: $Entropy(S)$ beschreibt den Zustand vor der Verzweigung

und $\sum_{v \in Werte(A)} \frac{|S_v|}{|S|} Entropy(S_v)$ den Zustand nach der Verzweigung eines Knotens. Dies bedeutet je

höher der InfoGain ist, desto eher wird verzweigt.

In Weka wird bei den Lerntechniken ein InfoGain Evaluator verwendet, der Attribute sehr gut und schnell bewerten kann. Dabei wird der Verdienst (Merit) eines Attributs basierend auf dem InfoGain berechnet. Je höher der Verdienst, desto wertvoller ist das Attribut für den Datenset. Was InfoGain allerdings nicht kann, ist die Abhängigkeit der Attribute voneinander zu erkennen. Für diesen Zweck wurden in der Diplomarbeit die unterschiedlichen Wrapper-Methoden verwendet. Ein weiteres Problem von InfoGain ist der Umgang mit Attributen mit vielen verschiedenen Ausgängen. Hierzu wird im nächsten Kapitel eine Verbesserung des InfoGain vorgestellt, das „GainRatio“.

GainRatio

(Informationen in diesem Unterkapitel wurden auf Basis der Literaturquellen: J.Han, M.Kamber. *Data Mining Concepts and Techniques*, 2nd Edition [HanK06] erstellt)

Wie bereits erwähnt hat Information-Gain Probleme mit Attributen, die sehr viele Ausgänge haben. Um dies zu verdeutlichen sehe ich mir ein kurzes Beispiel an:

Zu meiner Attributmenge addiere ich ein imaginäres Attribut „Spielnummer“ hinzu. Das heißt, dass jede Zeile in Wekas ARFF-Datei je eine unterschiedliche Spielnummer erhält (mit 1 anfangend). Dies bedeutet, dass ich sehr viele Ausgänge mit je einem Datentupel hätte, dass zu einer bestimmten Klasse gehört. Dies bedeutet aber auch, dass der Ausdruck $\sum_{v \in Werte(A)} \frac{|S_v|}{|S|} Entropy(S_v)$ in der

Information-Gain Formel gleich 0 sein wird und der Information-Gain bei diesem eigentlich nichts aussagenden Attribut einen unverhältnismäßig hohen Wert anzeigt.

Als Lösung für dieses Problem wurde das Maß GainRatio erfunden.

Die Formel für GainRatio lautet:

$$GainRatio(S, A) = \frac{Gain(S, A)}{SplitInformation(S, A)} \quad (3.8)$$

GainRatio berücksichtigt in seiner Formel eine Splitinformation, die folgendermaßen definiert ist:

$$SplitInformation(S, A) \equiv -\sum_{i=1}^c \frac{|S_i|}{|S|} \log_2 \frac{|S_i|}{|S|} \quad (3.9)$$

Wie man unschwer erkennen kann, führt Splitinformation eine Art Normalisierung durch, in dem sie Information-Gain zu der Gesamtanzahl der Datentupel in ein Verhältnis setzt.

Auch GainRatio wird bei Entscheidungsbäumen sehr gerne verwendet (z.B. bei C4.5). Ebenso (wie bei Info-Gain und ID3) wird das Attribut mit dem maximalen GainRatio zum Splitten ausgewählt.

Ein Problem von GainRatio ist dabei wenn die SplitInformation sich 0 nähert. In diesem Fall wird die Formel zunehmend instabil. Aus diesem Grund wurde auch eine Nebenbedingung eingefügt, dass der Information-Gain in GainRatio hinreichend groß sein muss (mindestens so groß wie der DurchschnittsGain über alle Tests).

Damit beende ich das Unterkapitel mit den Feature-Ranking Verfahren und komme zu einer anderer Methode der Feature Subset Selection-dem Wrapper.

3.2.3 Wrapper

(Informationen in diesem Unterkapitel wurden auf Basis der Literaturquellen: J.Han, M.Kamber. *Data Mining Concepts and Techniques*, 2nd Edition [**HanK06**] sowie I.H. Witten, E.Frank. *Data Mining Practical Machine Learning Tools and Techniques*, 2nd Edition [**WIFR05**] erstellt)

In diesem Unterkapitel beschreibe ich die Wrapper. Dabei erläutere ich zunächst die Funktionsweise und erkläre danach kurz die Wrapper-Suchmethode-Kombinationen, die in meiner Diplomarbeit verwendet wurden.

Funktionsweise des Wrappers

Ein Wrapper sucht unter allen Attributen. Er fängt mit der kleinsten (oder größten, je nach Verfahren) Elementenmenge an, untersucht jedes mögliche Feature Subset und wählt am Ende das Feature Subset aus, das zusammen mit der bereits ausgewählten Elementenmenge die beste Genauigkeit ab liefert. Dabei verwendet ein Wrapper die Genauigkeiten der Klassifikationen der ihm zugrunde liegender ausgewählten Algorithmen und arbeitet in Kombination mit bestimmten ausgewählten Suchmethoden. Bei den Suchmethoden gibt es einige Variationen die ich nun in folgenden Unterkapiteln erläutere.

Wrapper mit Best-First Search

Diese Methode durchsucht den Attributraum mit einem einfachen und für Wrapper sehr typischen Hillclimbing. Davon gibt es 2 Ausprägungen, eine Forward Selection und eine Backward Elimination Methode. Dabei ist es so, dass die Forward Selection Methode deutlich schneller ist als die Backward Elimination und meistens ein kleineres Featureset als Ergebnis zurückliefert. Die Ergebnisse der Backward Elimination jedoch sind meistens genauer, als die der Forward Selection. Ferner gibt es mit „Linear Forward Selection“, eine Erweiterung für Forward Selection auf die ich hier auch kurz eingehen werde.

Best-First Search Forward Selection (FS)

Diese Methode fängt mit der leeren Featuremenge an, sucht das Feature, das die Genauigkeit am Meisten erhöht und fügt es in die Featuremenge hinzu. Im zweiten Schritt fügt FS das Feature hinzu, das die Genauigkeit in der Kombination mit bereits hinzugefügtem Feature am Meisten erhöht. Diese Prozedur wird solange fortgesetzt, bis die Genauigkeit durch ein Hinzunehmen des neuen Features nicht mehr erhöht werden kann. Als Ergebnis hat man dann ein Featureset, das die Genauigkeit maximal erhöht.

Best-First Search Backward Elimination (BE)

Backward Elimination geht den genau umgekehrten Weg von Forward Selection. Es startet mit einem vollem Feature Set und löscht die Attribute aus ebendiesem sofern die Genauigkeit erhöht wird. Dabei ist es auch hier so, dass das Attribut mit dem größten Genauigkeitsgewinn durch seine Löschung, als erstes eliminiert wird und es in BE solange gelöscht wird bis ein Genauigkeitsverlust durch ein Löschen vorhanden ist.

Linear Forward Selection (LFS)

In Weka gibt es mit Linear Forward Selection ein ausgeweitetes Forward Selection Verfahren. Dieses Verfahren erlaubt dem Benutzer, unter anderem, vor jeder Subsettauswahl die Anzahl der Attribute zu beschränken, sowie weitere Restriktionen auf dem Attributset vorzunehmen. Genauer gesagt wird bei dieser Suchmethode bei jedem Schritt statt aller Attribute, nur eine (nach bestimmten Kriterien) ausgewählte Untermenge von Attributen betrachtet. Diese Kriterien können vom Benutzer ausgewählt werden, was auch bedeutet, dass die Attributmenge nach diesem Kriterium vor der LFS-Durchführung sortiert wird. Diese Art von Suchmethode, verschnellert die Suche deutlich, ohne dabei an der Genauigkeit der Auswahl gegenüber einer einfachen FS signifikant zu verlieren. [GFHK09] Dies ist auch der Grund, wieso ich eine LFS in Kombination mit einem Wrapper in meiner Diplomarbeit einer „schlichten“ Forward Selection vorgezogen habe.

Wrapper mit Ranksearch

Ferner wurde in der Diplomarbeit auch ein Wrapper mit Ranksearch verwendet. Dieser sortiert das Ergebnis nach einer bestimmten Heuristik. In unserem Fall wurde als Heuristik GainRatio verwendet (siehe Kapitel 3.2.2). Zuerst erfolgt hier eine Sortierung der Attribute nach dem GainRatio Maß. Danach wird wie bei Forward Selection vorgegangen d.h. das Attribut mit dem besten GainRatio Maß wird als Erstes genommen, dass mit dem zweitbesten GainRatio Maß als Zweites usw.

Nun bin ich am Ende des Unterkapitels zu den in der Diplomarbeit verwendeten Lerntechniken angelangt. Im letzten Unterkapitel von diesem Kapitel erläutere ich die in der Diplomarbeit verwendeten Evaluierungstechniken.

3.3. Erklärung der verschiedenen Evaluierungsarten

In diesem Unterkapitel werden verschiedene verwendete Evaluierungsarten erklärt. Hierbei gehe ich zunächst auf die in der Diplomarbeit sehr oft verwendete Cross-Validation ein. Abschließend erwähne ich noch die ebenso verwendete Möglichkeit der Evaluierung mit Supplied Test Set.

3.3.1 Cross-Validation

Cross-Validation (CV) wird oft bei begrenzten Mengen an Trainings- und Testdaten benutzt. Die Grundidee davon ist, dass die Daten in X gleich große Partitionen geteilt werden. Dabei werden X-1 Partitionen für Training und eine Partition für das Testen der trainierten Daten benutzt. Die Anzahl der Durchläufe (Folds) der CV entspricht dabei der Anzahl der Partitionen, da jede Partition genau einmal als Testset benutzt wird. Am Besten wird dies auf dem Beispielbild einer 4-Fold Cross-Validation verdeutlicht (siehe Bild 3.6).

Der Standardwert für X liegt bei einer CV bei 10. Auch in meiner Diplomarbeit wurde immer eine 10-Fold Cross-Validation verwendet. Der Leser mag sich fragen: „Wieso gerade 10?“ Nun, umfangreiche Tests haben bewiesen, dass 10 eine optimale Anzahl an Durchläufen ist um die beste Fehlerabschätzung zu erhalten. [WIFRD01]

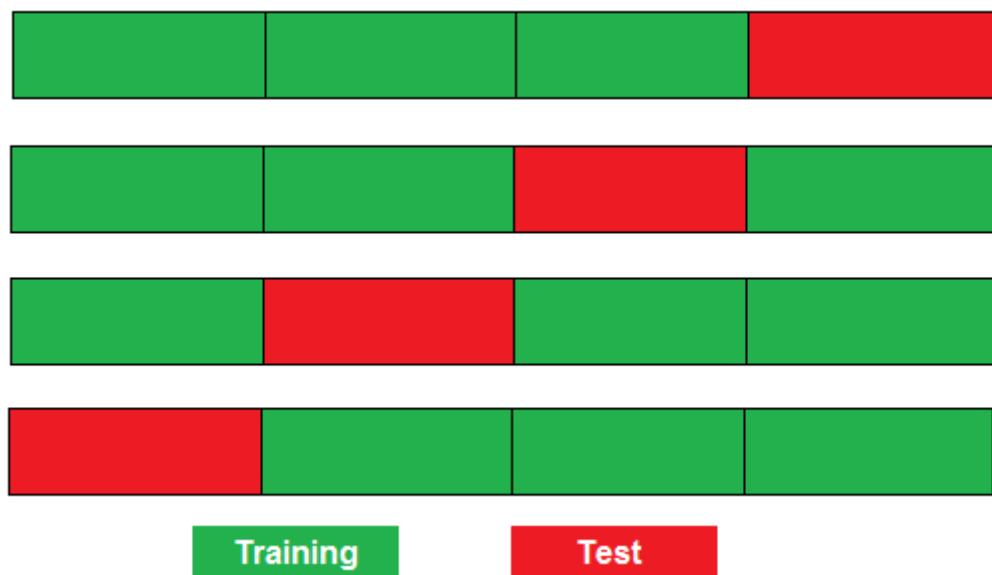


Bild 3.6: Beispiel einer 4-Fold Cross-Validation

3.3.2 Supplied Test Set

Mit einem Supplied Test Set (STS) ist ein von Trainingsset getrenntes Testset gemeint. Dies bedeutet, dass für den Test ein spezielles externes Datenset verwendet wird.

STS wurde von mir in der Diplomarbeit verwendet um beispielsweise auf Daten der Bundesliga-Saisons 2005-2007 zu lernen und das Gelernte auf der Saison 2008 zu testen, bzw. die Spiele der Saison 2008 versuchen auf Basis des Gelernten aus den vorherigen 3 Saisons vorherzusagen. Natürlich muss das Supplied Test Set, genau dieselben Attribute beinhalten wie das verwendete Trainingsset.

Hiermit schließe ich das dritte Kapitel ab. Im nächsten Kapitel fange ich mit der Analyse von den vorhandenen Daten mittels verschiedener Lerntechniken an.

Kapitel 4: Analyse der Attribute

4.1. Einleitung und zu beantwortende Fragen

In diesem Kapitel beginne ich mit der Analyse der vorhandenen Attribute. Das Ziel dieses Kapitels ist, die Besten Attribute mit Hilfe von Trainingssets verschiedener Zusammensetzung und Größe für unterschiedliche Klassifizierer zu bestimmen. Dies wird einerseits mit Tests zur optimalen Belegung bei den Variablen von Gruppenattributen und andererseits mit Wrappern, InfoGain-Tests und Ergebnisvergleichen unter verschiedenst möglichen Einstellungen versucht zu erreichen.

Bevor ich aber mit dem Kapitel anfange noch eine kleine Begriffsklärung. Eine Saison in der Fußball-Bundesliga umfasst 2 Jahre. Beispielsweise fängt die Saison 2009/2010 im August 2009 an und endet in Mai 2010. Um die Bezeichnung zu vereinfachen, wird in beiden folgenden Kapiteln nur das Startjahr als Bezeichnung gebraucht (z.B. Saison 2007/2008->2007; oder Saison 2002/2003->2002). Dies vereinfacht die Darstellung um einiges. Nur in Ausnahmefällen wird auf die volle Bezeichnung zurückgegriffen.

Nun aber zurück zur Analyse, unter anderem betrachte ich in diesem Kapitel folgende Fragen:

- Was sind die optimalen Belegungen für die Gruppenattribute?
- Welche Attribute tragen am Meisten dazu bei, die richtige Tendenz zu erraten?
- Welche Klassifizierer kommen mit welchen Attributen am Besten klar?
- Was ist die optimale Auswahl an Attributen für einen Datensatz?

Alle Analysen in diesem Kapitel wurden auf 10 Trainingssets verschiedener Größe durchgeführt. Diese sind:

- 4 Trainingssets mit der Länge von 3 Saisons (2002-2004, 2003-2005, 2004-2006, 2005-2007)
- 3 Trainingssets mit der Länge von 4 Saisons (2002-2005, 2003-2006, 2004-2007)
- 2 Trainingssets mit der Länge von 5 Saisons (2002-2006, 2003-2007)
- 1 Trainingsset mit der Länge von 6 Saisons (2002-2007)

Die jeweiligen Testbedingungen sowie Einzelheiten werden in den einzelnen zugehörigen Unterkapiteln genauer beschrieben. Die meisten Ergebnisse werden in übersichtlichen Graphiken zusammengefasst. Die kompletten Ergebnisse können im Anhang C gefunden werden.

4.2. Analyse der Gruppenattribute

In diesem Unterkapitel werden die 4 vorhandenen Gruppenattribute (Versus, Form, AvgPlatz und Abstieg) auf die optimalen Belegungen analysiert. Generell besteht die Analyse aus folgenden Schritten:

1. Bestimmung des Testfensters (erfahrungsisierte Größe 10, diese Werte werden durch manuelle Tests mit InfoGain an den Randbereichen bestimmt. Aus Platzgründen werden diese Tests nur mit den Ergebnissen erwähnt, auf Details wird im Allgemeinen hier nicht eingegangen).
2. Bestimmung der optimalen Attribut-Ausprägung für das im vorherigen Schritt bestimmte Testfenster (Hier werden zu den bereits schon vorhandenen Ausprägungen auch die jeweiligen entsprechenden Differenzausprägungen mitbetrachtet).
3. Bestimmung der optimalen Belegung und Zuweisung zu den entsprechenden Variablen, für die im vorherigen Schritt festgelegte Ausprägung.

Bei Formattributen ist ein zusätzlicher Zwischenschritt notwendig, da Formatattribute eine zusätzliche Option haben (saisonübergreifend/nicht saisonübergreifend). Vor dem ersten Schritt wird in diesem Fall zusätzlich getestet ob eine saisonübergreifende Statistik einen Sinn hat oder man sich nur auf eine saisoninterne Statistik beschränken sollte. Alle Untersuchungen basieren auf dem Merit des Infogains (siehe Kapitel 3.2.2) auf den bereits in der Einleitung erwähnten 10 Trainingssets.

4.2.1 Versus

Definition: Beim Versus-Attribut werden die letzten X Aufeinandertreffen der Teams aus der betrachteten Paarung unter sich betrachtet (weitere Informationen siehe Kapitel 2.2.2). Das optimale X gilt es hier zu bestimmen.

Mögliche Ausprägungen des Versus-Attributs: VersusXHeim, VersusXGast, VersusX (Differenzausprägung aus VersusXHeim und VersusXGast)

Optimales Testfenster: 10-19 (siehe hierzu den Punkt 1, in der Einleitung des Kapitels 4.2)

Bestimmung der optimalen Attributsausprägung

Aus Platz- und Übersichtlichkeitsgründen folgen Graphiken aus 4 Trainingssets verschiedener Größen. Die anderen 6 Trainingssets entsprachen überwiegend den Trends, die in diesen Graphiken sichtbar gemacht wurden.

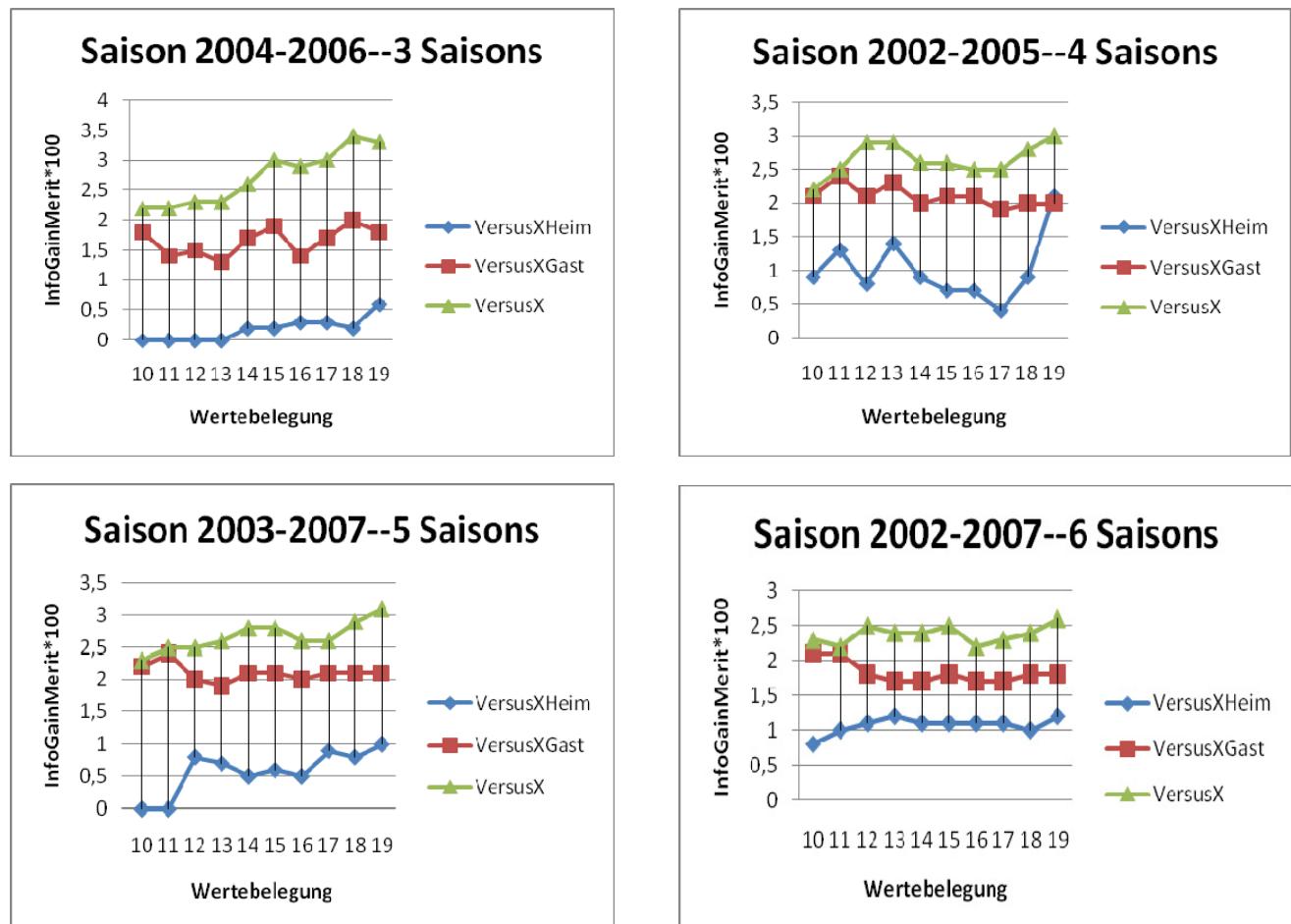


Bild 4.1 Übersicht der Versusausprägungen mit verschiedener Trainingssetlänge

Erklärung der Bilder

Auf der X-Achse sind die Wertebelegungen aus dem optimalen Testfenster zu finden (da das optimale Testfenster sich hier von 10-19 erstreckt, befinden sich auch hier f鰎glich die Werte von 10-19 auf der X-Achse). Die Y-Achse wird von den entsprechenden Werten des InfoGain-Merits belegt (Ubersichtshalber mit 100 multipliziert). Diese Erklrung gilt auch fr Bilder anderer Gruppenattribute (somit auch fr Bilder 4.2-4.5).

Wie aus dem Bild 4.1 ersichtlich ist, liegt der hchste Infogain bei Versus-Attributen bei der VersusX ($X=1,\dots,5$)-Ausprgung (grne Linie auf dem Bild). Dabei ist der Abstand von VersusX-Linie zur VersusXGast-Linie zwar nicht gro und dennoch immer vorhanden und eindeutig. VersusXHeim-Linie ist dagegen sehr weit unten auf der Koordinatenachse zu finden. Oft ist diese sogar sehr nahe oder gar gleich 0. Man sieht, dass die maximalen Werte der Ausprgungen bei 18 und 19 liegen, was aber im nchsten Punkt deutlicher und genauer beschrieben wird.

Bestimmung der optimalen Belegung und Zuweisung

Aus der untenstehenden Tabelle 4.1 kann man nun die Ergebnisse der Analysen entnehmen. Hierbei wurde der beste Wert fett und die Pltze 2-5 kursiv dargestellt. Auch wurden die ersten 5 Positionen der Belegung grau unterlegt. Ebenso wurde auch hier der Ubersicht halber der Meritwert des InfoGains mit 100 multipliziert um die fhrenden Nullen zu entfernen. Als Endergebnis wird der Beste Wert als Belegung der Variable Versus1 zugewiesen, der zweitbeste Versus2 etc bis Versus5.

Platz	VersusX-Wert	IG-Average (in IG*100)	Variable zugewiesen
1	19	3	Versus1
2	18	2,88	<i>Versus2</i>
3	15	2,75	<i>Versus3</i>
4	12	2,7	<i>Versus4</i>
5	14	2,68	<i>Versus5</i>
6	17	2,64	---
7	16	2,57	---
8	13	2,53	---
9	11	2,38	---
10	10	2,26	---

Tabelle 4.1: Optimalitt der Versus-Belegungen

Die Tabelle 4.1 ist somit auf folgende Art zu lesen und zu verstehen: Versus-Wert mit Belegung 19 hat ein InfoGain-Merit von 3 erzielt, belegte damit den 1. Platz unter allen Belegungen und wurde entsprechend der Variable Versus1 aus dem Enddatensatz zugewiesen. Der Verdacht aus dem graphischen Auszug bei der Auswahl der Attributsausprgung hat sich somit bestigt, die besten Werte liegen nahe dem Ende des Testfensters.

Wie man aus der Tabelle 4.1 schlielich entnehmen kann, wurden im Datensatz fr die Versusbelegungen die Werte 19,18,15,12 und 14 verwendet.

4.2.2 Form

Definition: Form-Attribut berechnet die erzielten Punkte aus den letzten X Spielen eines Teams. (weitere Informationen siehe Kapitel 2.2.2). Das optimale X soll hier zu bestmtt werden. Wie bereits erwhnt, ist bei Formattribut neben den verschiedenen Ausprgungen auch die Mglichkeit gegeben, die Formberechnung saisonbergreifend oder nicht saisonbergreifend zu gestalten. Diese Mglichkeit musste zustzlich vor dem Belegungstest mit InfoGain-Vergleich berprft werden.

Mögliche Ausprägungen der Form-Attribute: FormXHeimHeim, FormXGastAusw, FormXHeim, FormXGast, DiffFormX (Differenzausprägung von FormXHeim und FormXGast)

Optimales Testfenster: 10-19

Überprüfung Saisonübergreifend/nicht Saisonübergreifend

Hierfür wurden dieselben Trainingssets einmal saisonübergreifend und einmal nicht saisonübergreifend betrachtet und die daraus folgenden Infogain-Werte verglichen. Das Ergebnis ist eindeutig, daher beschränke ich mich auf 2 Diagramme der DiffForm-Ausprägungen als Repräsentanten der gesamten Ergebnismenge.

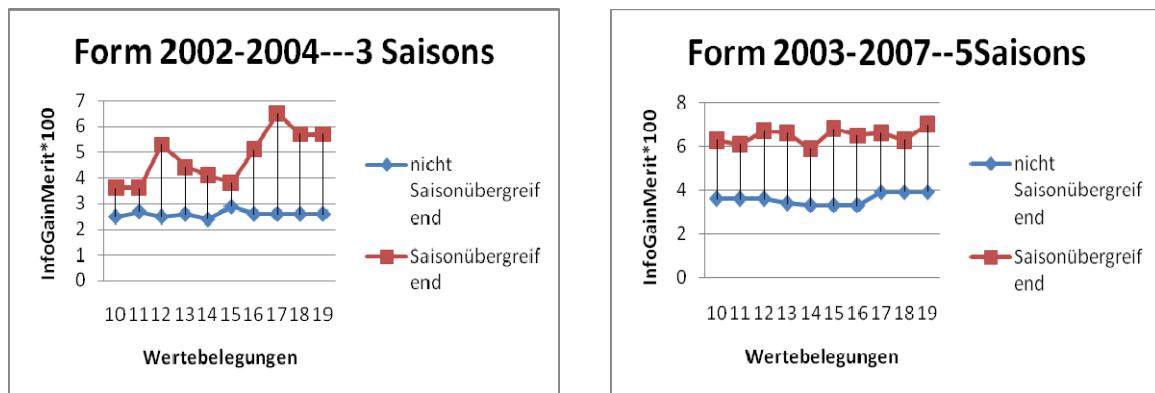


Bild 4.2: Form-Attribut Prüfung saisonübergreifend/saisonintern

Auf Bild 4.2 lässt sich der relativ deutliche Unterschied erkennen. Es gab insgesamt keinen einzigen Punkt, wo die nicht saisonübergreifende Formberechnung besser war als die saisonübergreifende. Somit wird, aufgrund dieser Erkenntnisse, eine saisonübergreifende Formberechnung verwendet.

Bestimmung der optimalen Attributsausprägung

Es folgen Graphiken aus 4 Datensets verschiedener Größen, die anderen 6 Datensätze entsprachen überwiegend den Trends, die in diesen Graphiken deutlich gemacht wurden.

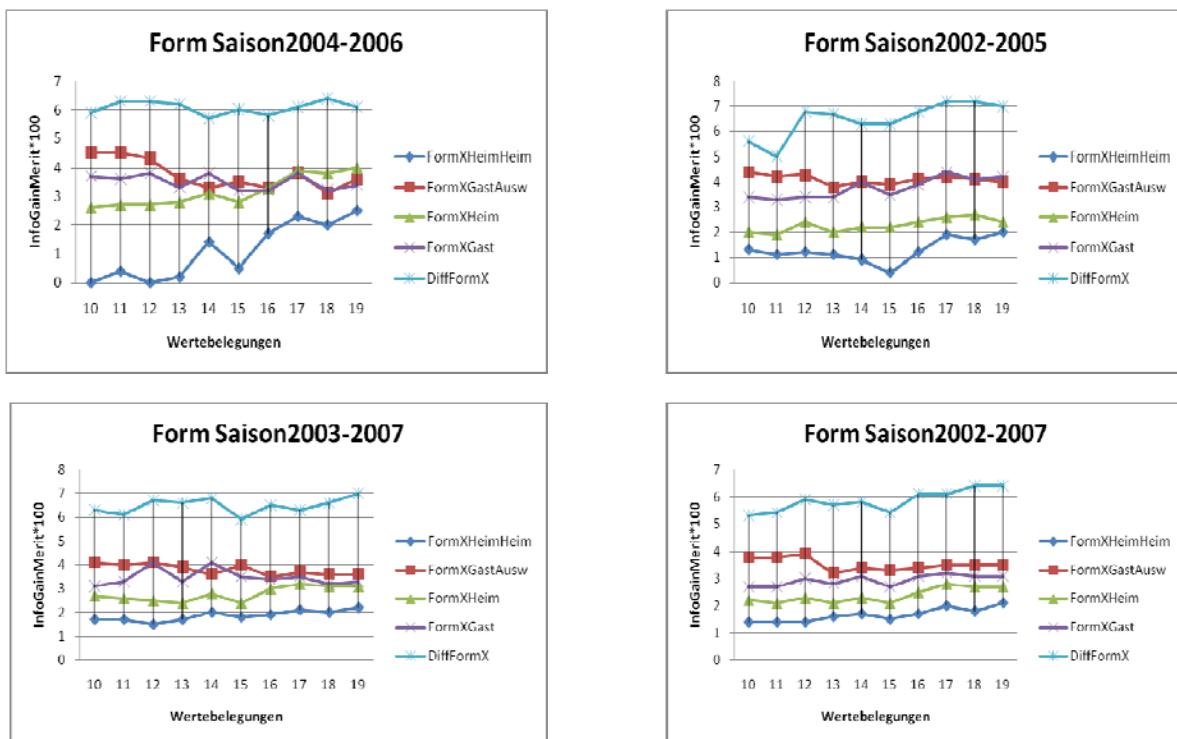


Bild 4.3: Übersicht der Formausprägungen mit verschiedener Trainingssetzlänge

Wie aus Bild 4.3 ersichtlich ist, liegt hier der höchste Infogain bei den DiffFormX ($X=1,\dots,5$)-Ausprägungen (hellblaue Linie auf dem Bild). Wie man ebenso auf dieser Graphik sehen kann, erzielt die Formdifferenz-Ausprägung DiffFormX mit Abstand die besten Ergebnisse. Der Verlauf der DiffFormX-Linie ist relativ gleichmäßig, es lässt sich aber erkennen, einmal bei den Werten 12-13 und dann noch bei den Werten 17-19 ein etwas höherer Anstieg erfolgt.

Bestimmung der optimalen Belegung und Zuweisung

Aus der untenstehenden Tabelle 4.2 kann man nun die Ergebnisse der Analysen entnehmen. Hier gelten die gleichen Optionen wie bei vorangehenden Versustests. Als Endergebnis wird der beste Wert als Belegung der Variable Form1 zugewiesen, der zweitbeste Form2 etc bis Form5.

Platz	FormX-Wert	IG-Average (in $IG \cdot 100$)	Variable zugewiesen
1	17	6,68	Form1
2	19	6,51	Form2
3	13	6,42	Form3
4	12	6,39	Form4
5	18	6,29	Form5
6	16	6,29	---
7	15	6,11	---
8	14	5,89	---
9	11	5,69	---
10	10	5,59	---

Tabelle 4.2: Optimalität der Form-Belegungen

Wie man der Tabelle 4.2 entnehmen kann, sind haben DiffForm-Belegungen von 16 und 18 gleichen Durchschnittsinfogainwert auf den 10 benutzten Datensätzen. Aufgrund des Hinzuziehens von zusätzlichen Tests (u.a. andere Datensatzlänge) wurde die Attributbelegung 18 als Form5-Belegung übernommen. Ansonsten sieht man, dass der Abstand von der besten Belegung zu den anderen Belegungen relativ deutlich ist. Dabei hat sich, auch hier der Verdacht aus der graphischen Bestimmung der Ausprägungen vollständig bestätigt, denn sowohl Werte 17-19, als auch 12-13 bilden die Top5.

Somit wurden in den Gesamtdatensatz für die Formbelegungen die Werte 17,19,13,12 und 18 übernommen. Die Formberechnung erfolgt „Saisonübergreifend“.

4.2.3 AvgPlatz

Definition: AvgPlatz-Attribute berechnen die Durchschnittsplatzierung eines Teams in den letzten X Saisons. (weitere Informationen siehe Kapitel 2.2.2). Das optimale X sollte hier bestimmt werden.

Mögliche Ausprägungen des AvgPlatz-Attributs: AvgPlatzXHeim, AvgPlatzXGast, DiffAvgX (Differenz ausprägung aus AvgPlatzXHeim und AvgPlatzXGast)

Optimales Testfenster: 1-10

Bestimmung der optimalen Attributsausprägung

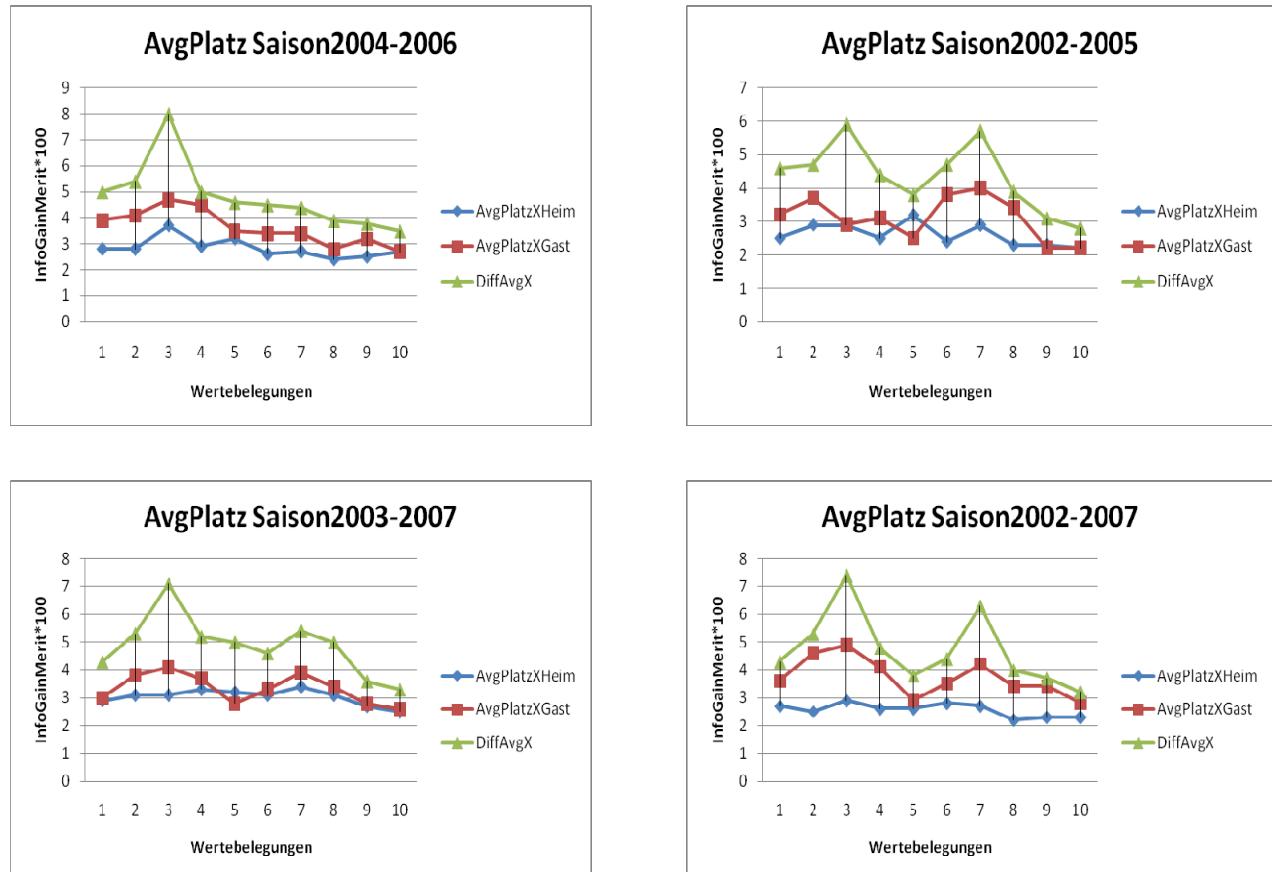


Bild 4.4: Übersicht der AvgPlatz-Ausprägungen mit verschiedener Trainingssetlänge

Wie aus Bild 4.4 zu ersehen, liegt der höchste Infogain bei bei den DiffAvgX ($X=1,\dots,5$)-Ausprägungen (grüne Linie auf dem Bild). Dieser Unterschied ist relativ deutlich erkennbar, da die grüne Linie des DiffAvgX immer höher ist als die AvgPlatzHeim und AvgPlatzGast-Linien. Dabei sind besonders bei den grünen und roten Linien 2 lokale Extrema sehr deutlich, einmal ein größerer Anstieg bei 3 und dann (in einigen Fällen) ein etwas kleinerer Anstieg bei 7.

Bestimmung der optimalen Belegung und Zuweisung

Aus der untenstehenden Tabelle 4.3 lassen sich nun die Ergebnisse der Analysen entnehmen. Als Endergebnis wird der Beste Wert als Belegung der Variable AvgPlatz1 zugewiesen, der zweitbeste AvgPlatz2 etc bis AvgPlatz5.

Bei AvgPlatz ist das Ergebnis so deutlich wie bei keinem anderen Gruppenattribut. Bei Wert 3 ist der DiffAvgPlatzX-Wert deutlich besser als alle anderen Werte (was das Extremum bei dem graphischen Ausprägungsbestimmungs-Teil bestätigt). Auch der andere kleinere Anstieg bei den Graphiken wird mit der Übernahme von 7 als zweiter Belegungswert bestätigt.

Platz	DiffAvgX-Wert	IG-Average (in IG*100)	Variable zugewiesen
1	3	6,69	AvgPlatz1
2	7	5,32	AvgPlatz2
3	2	5,21	AvgPlatz3
4	4	4,89	AvgPlatz4
5	6	4,54	AvgPlatz5
6	1	4,50	---
7	5	4,32	---
8	8	4,27	---
9	9	3,48	---
10	10	3,20	---

Tabelle 4.3: Optimalität der AvgPlatz-Belegungen

Wie man also der Tabelle 4.3 entnehmen kann werden für den Gesamtdatensatz für AvgPlatz-Belegungen die Werte 3,7,2,4 und 6 übernommen.

4.2.4 Abstieg

Definition: Bei Abstieg-Gruppenattributen handelt es sich um Attribute bei denen die Zugehörigkeit eines Teams zur 1.Bundesliga in den letzten X Jahren berechnet wird. (weitere Informationen siehe Kapitel 2.2.2). Das optimale X sollte nun in folgendem Abschnitt bestimmt werden.

Mögliche Ausprägungen des Abstieg-Attributs: AbstiegXHeim, AbstiegXGast, DiffAbstiegX (Differenzausprägung aus AbstiegXHeim und AbstiegXGast)

Optimales Testfenster: 6-15

Bestimmung der optimalen Attributausprägung

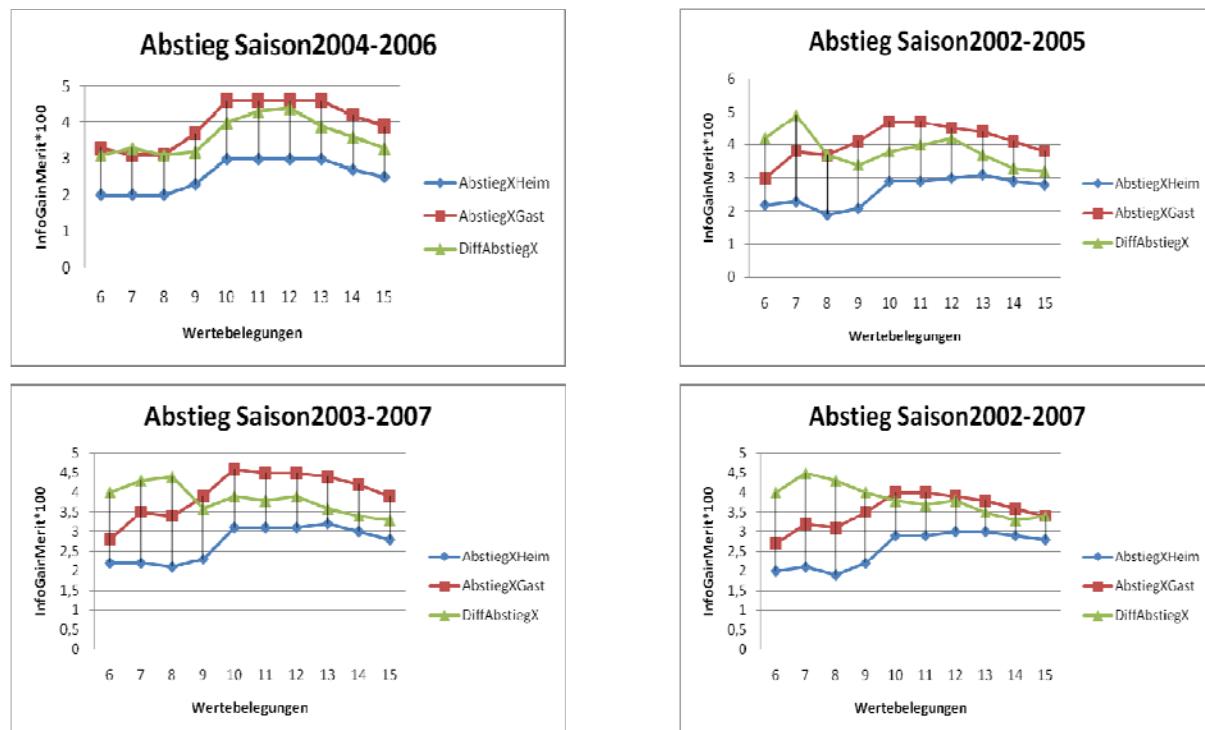


Bild 4.5: Übersicht der Abstieg-Ausprägungen mit verschiedener Trainingssetlänge

Wie auf Bild 4.5 zu erkennen, liegen bei Abstieg-Gruppenattribut alle 3 Ausprägungen relativ nah beieinander. Dabei stechen besonders 2 Ausprägungen ins Auge: „AbstiegXGast“ (rote Linie) und „DiffAbstiegX“ (grüne Linie). Graphisch erkennt man, dass DiffAbstiegX eher am Anfang des Testfensters seinen maximalen Anstieg hat, AbstiegXGast hingegen eher zur Mitte des Testfensters. Bei Bestimmung der InfoGain-Durchschnittswerte hat sich AbstiegXGast als etwas besser als DiffAbstiegX rausgestellt. Auch der Anstieg von DiffAbstiegX hat sich als nicht so stark, wie der bei AbstiegXGast herausgestellt. Damit ist Abstieg das einzige Gruppenattribut, wo die Differenzausprägung des Attributs etwas schlechter ist als eine alternative Ausprägung.

Bestimmung der optimalen Belegung und Zuweisung

Platz	DiffAvgX-Wert	IG-Average (in IG*100)	Variable zugewiesen
1	10	4,56	Abstieg1
2	11	4,53	Abstieg2
3	12	4,46	Abstieg3
4	13	4,36	Abstieg4
5	14	4,08	Abstieg5
6	9	3,86	---
7	15	3,82	---
8	7	3,44	---
9	8	3,35	---
10	6	2,92	---

Tabelle 4.4: Optimalität der Abstieg-Belegungen

Auch die Ergebnisse der Bestimmung von optimaler Werteberechnung sind bei Abstieg-Attribut nicht so eindeutig wie bei anderen Gruppenattributen. Hier ist der Vorsprung vom ersten zum zweiten Platz relativ knapp.

Wie man also der Tabelle 4.4 entnehmen kann werden für den Gesamtdatensatz für die Abstieg-Belegungen die Werte 3,7,2,4 und 6 übernommen.

Somit bin ich am Ende der Analyse von den Gruppenattributsbelegungen angekommen. Ich habe nun bei jedem der Gruppenattribute eine aus 5 Werten bestehende (aus der Sicht von InfoGain) optimale Belegung vorliegen. Bevor ich aber zur großen Analyse unter Zuhilfenahme von Wrappern schreite, sollte kurz eine Momentaufnahme der aktuellen Ergebnisse erstellt werden. Diese wird dann am Ende des Kapitels mit den Ergebnissen nach der Analyse verglichen werden.

4.3 Ergebnisüberblick vor der großen Analyse

In diesem Unterkapitel werden die Ergebnisse der verwendeten Klassifizierer sowohl graphisch, als auch tabellarisch kurz betrachtet, um diese nach der Analyse noch einmal zu vergleichen. Zur Erinnerung noch einmal die in Rahmen meiner Diplomarbeit verwendeten Klassifizierer: 1. Naive Bayes, 2.JRip, 3. J48, 4.SVM und 5.Random Forest.

Nun sehen wir uns die Ergebnisse erst einmal tabellarisch Jahr für Jahr an.

Vorhersage der Saison 2005

Hier kann ich Trainingssets mit der Länge von 3 Saisons testen. Die Genauigkeiten der Vorhersagen sind:

Klassifizierer Trainingsset – Testset	Naive Bayes	JRip	J48	SVM	Random Forest	ZeroR (Tendenz=1)
2002/2004---2005	44,77	46,73	38,89	48,69	41,83	42,81

Tabelle 4.5: Ergebnisüberblick Saison2005 (alle Werte in %, gerundet auf 2 Nachkommastellen)

Wie man erkennt, ist hier mit SVM das beste Ergebnis erreicht worden. Dies ist bemerkenswert, denn bei allen anderen Trainingsets fällt die Leistung der Support Vector Machines bei dem Komplett-Datensatz zum Teil deutlich (um bis zu 8%) ab. Ebenso zu bemerken ist, dass 3 von 5 Klassifizierern über den ZeroR-Wert (Regel mit Tendenz immer=1) liegen.

Vorhersage der Saison 2006

Hier können Datensätze der Länge von 3 und 4 Saisons getestet werden. Die Genauigkeiten der Vorhersagen sind:

Klassifizierer Trainingsset – Testset	Naive Bayes	JRip	J48	SVM	Random Forest	ZeroR (Tendenz=1)
2003/2005---2006	44,44	47,06	39,87	40,52	40,85	
2002/2005---2006	43,46	42,81	39,87	42,81	41,50	43,79

Tabelle 4.6: Ergebnisüberblick Saison2006 (alle Werte in %, gerundet auf 2 Nachkommastellen)

Zu erkennen ist, dass Naive Bayes und JRip bei Hinzuziehung von einer weiteren Saison unter das ZeroR fallen. SVM steigt leicht, insgesamt bleibt das Ergebnis für die Saison 2006 jedoch ziemlich schwach.

Vorhersage der Saison 2007

Hier können Trainingssets der Länge 3,4 und 5 Saisons getestet werden:

Klassifizierer Trainingsset – Testset	Naive Bayes	JRip	J48	SVM	Random Forest	ZeroR (Tendenz=1)
2004/2006---2007	46,73	46,08	38,89	44,44	38,89	
2003/2006---2007	49,35	50,00	43,79	44,12	45,10	
2002/2006---2007	48,69	46,73	39,22	46,41	45,10	46,73

Tabelle 4.7: Ergebnisüberblick Saison2007 (alle Werte in %, gerundet auf 2 Nachkommastellen)

Für 2007 scheint die Trainingssettänge von 4 Saisons die optimale zu sein. Allerdings ist auch bei der Trainingssettänge von 5 Saisons anzumerken, dass bei den Genauigkeiten gegenüber 3 Saisons ein starker Anstieg erfolgt ist.

Vorhersage der Saison 2008

Hier können Datensätze der Länge 3,4,5 und 6 Saisons verwendet werden:

Klassifizierer Trainings. —Tests.	Naive Bayes	JRip	J48	SVM	Random Forest	ZeroR (Tendenz=1)
2005/2007---2008	47,14	48,48	36,36	40,40	44,78	48,15
2004/2007---2008	47,47	48,82	38,38	41,08	42,42	
2003/2007---2008	47,47	47,14	39,73	46,13	43,10	
2002/2007---2008	47,81	50,84	35,02	48,15	43,10	

Tabelle 4.8: Ergebnisüberblick Saison2008 (alle Werte in %, gerundet auf 2 Nachkommastellen)

Man erkennt hier, dass bei Naive Bayes, JRip und SVM der Genauigkeitswert mit der zunehmenden Trainingsseterhöhung gewachsen ist. Bei JRip überstieg dieser sogar die 50%-Marke.

Komplettübersicht

Klassifizierer Trainings. —Testset	Naive Bayes	JRip	J48	SVM	Random Forest
2002/2004---2005	44,77	46,73	38,89	48,69	41,83
2003/2005---2006	44,44	47,06	39,87	40,52	40,85
2004/2006---2007	46,73	46,08	38,89	44,44	38,89
2005/2007---2008	47,14	48,48	36,36	40,40	44,78
2002/2005---2006	43,46	42,81	39,87	42,81	41,50
2003/2006---2007	49,35	50,00	43,79	44,12	45,10
2004/2007---2008	47,47	48,82	38,38	41,08	42,42
2002/2006---2007	48,69	46,73	39,22	46,41	45,10
2003/2007---2008	47,47	47,14	39,73	46,13	43,10
2002/2007---2008	47,81	50,84	35,02	48,15	43,10
Länge 3-Durchschn.	45,77	47,09	38,50	43,51	41,59
Länge 4-Durchschn.	46,76	47,21	40,68	42,67	43,01
Länge 5-Durchschn.	48,08	46,94	39,48	46,27	44,10
Gesamtdurchschn.	46,73	47,47	39,00	44,28	42,67

Tabelle 4.9: Gesamtergebnisüberblick über alle Saisons (alle Werte in %, gerundet auf 2 Nachkommastellen)

Zum Abschluss des Unterkapitels noch ein graphischer Verlauf der verwendeten Algorithmen. (siehe Bild 4.6, zur besseren Übersicht wurden hier die Punkte verbunden). Die einzelnen Datensatzlängen sind durch eine senkrechte orangene Linie getrennt. Auf der X-Achse sind dabei die einzelnen Trainingssets-Datensätze zu finden (204 bedeutet Trainingsset Saison2002-2004 usw.), auf der Y-Achse ist die Genauigkeit in Prozent dargestellt.

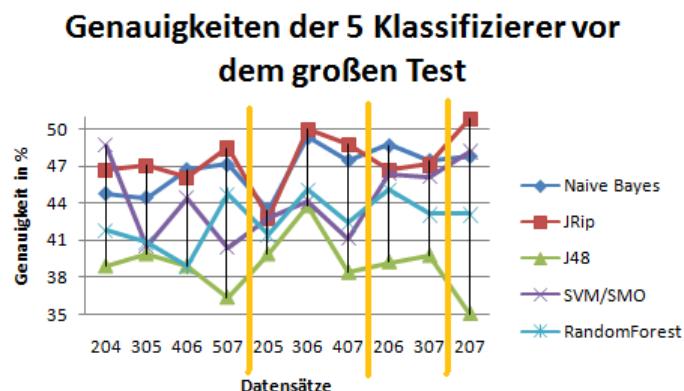


Bild 4.6: Graphischer Verlauf der verwendeten Algorithmen über alle Trainingssets

4.4 Hauptanalyse der kompletten Attributmenge

Bevor ich mit diesem Unterkapitel beginne, möchte ich mich bei Prof. J. Fürnkranz sowie bei Sang-Hyeun Park dafür bedanken, dass ich die Möglichkeit hatte, den Clusterserver sowie die „parallele Cross-Validation“ des Fachgebiets „Knowledge Engineering“ an der TU Darmstadt zur Ergebnisberechnung in diesem Unterkapitel zu verwenden.

In diesem Kapitel führe ich eine größere Analyse durch, die die vorhandenen 169 Attribute (d.h. exklusive dem Klassenattribut „Tendenz“) genaustens unter die Lupe nimmt.

Hierfür wurden mittels Wrapper und InfoGain werden 4 verschiedene Tests auf Basis der 5 Klassifizierer und 10 Trainingssets durchgeführt.

Die verwendeten Tests sind:

- Wrapper mit entsprechendem Klassifizierer mit Suchmethode: RankSearch (mittels GainRatio), später als *WrapperRSG* bezeichnet
- Wrapper mit entsprechendem Klassifizierer mit Suchmethode: BestFirst (Backward Elimination), später als *WrapperBFBE* bezeichnet
- Wrapper mit entsprechendem Klassifizierer mit Suchmethode: LinearForwardSelection, später als *WrapperLFS* bezeichnet
- InfoGain mit Ranker (Klassifiziererunabhängig, Kapitel 4.4.6)

Die verwendeten Klassifizierer sind (zur Erinnerung):

- Naive Bayes
- JRip
- J48
- SVM
- Random Forest

Die verwendeten Trainingssets sind (zur Erinnerung):

- 2002-2004, 2003-2005, 2004-2006, 2005-2007 (Länge 3 Saisons)
- 2002-2005, 2003-2006, 2004-2007 (Länge 4 Saisons)
- 2002-2006, 2003-2007 (Länge 5 Saisons)
- 2002-2007 (Länge 6 Saisons)

Die verwendete Evaluationstechnik und Erklärung der Bewertung

Die verwendete Evaluierungstechnik eine 10-fold Cross-Validation (siehe Kapitel 3.3.1).

Die Ergebnisse werden dann auch in Folds (Durchläufe) ausgedrückt. Diese zeigen an, in wievielen Folds das Attribut vom Wrapper mit der entsprechenden Suchmethode und Klassifizierer in die ausgewählte Attributmenge übernommen wurde. Somit ist bei einer 10-Fold CrossValidation ein Wert 10 (aus maximal 10 Folds) optimal und ein Wert 0 sehr schlecht. Bei meiner Bewertung gebe ich je einem Fold einen Punktewert von 1. Wenn ein Attribut also bei 8/10 Folds in die Attributmenge übernommen wurde, entspricht dies einer 8, bei dem entsprechenden Datensatz und Klassifizierer. Von den gesamten Werten des Wrappers wird am Ende für jeden Klassifizierer ein Durchschnittswert gebildet. Die Sortierung für jeden Klassifizierer erfolgt nach diesem Durchschnittswert.

Beispiel

Bei Klassifizierer Naive Bayes erhält das Attribut A folgende Werte:

- Wrapper 1---in 10 von 10 Folds übernommen, also eine 10
- Wrapper2---in 6 von 10 Folds übernommen, also eine 6
- Schließlich Wrapper 3---wird das Attribut in 8 von 10 Folds übernommen, also eine 8

Somit ist der Durchschnittswert des Attribut für Naive Bayes $(10+8+6)/3=8$. Somit wurde dieses Attribut durchschnittlich in 8 von 10 Folds in die Attributmenge übernommen. Nach diesem Wert wird für den Klassifizierer ein Ranking erstellt.

Da ich aus Platzgründen nicht jedes Mal eine Tabelle bestehend aus 169 Attributen in meiner Diplomarbeit präsentieren kann, muss ich mich für jeden Klassifizierer auf eine kurze tabellarische Beschreibung der durchschnittesten und -schlechtesten 10 Attributen für den Klassifizierer beschränken. Dabei gehe ich jedoch in der anliegenden Diskussion auf die besonderen Ausreißer auch außerhalb dieser besten/schlechtesten 10 ein und erwähne auch die Auffälligkeiten bei den einzelnen Wrapper-Ergebnissen, sowie untersuche einzelne Attributgruppen, die im 2.Kapitel zu finden sind. Die kompletten Ranking-Ergebnisse finden Sie im Anhang D.

4.4.1 Naive Bayes

Im Folgenden werden die Ergebnisse der TOP10 für den Klassifizierer Naive Bayes dargestellt.

Top10 Naive Bayes

Platz	Attributname	D-Wert Folds aus 10
1.	DiffAVG1	5,57 (55,7%)
2.	DiffGermanyN	5,33 (53,3%)
-.	DiffUEFA1J	5,33 (53,3%)
4.	DiffUEFA5J	5,10 (51,0%)
5.	Form4GastAusw	5,07 (50,7%)
6.	DiffNatPos	5,03 (50,3%)
7.	DiffAVG5	4,97 (49,7%)
8.	DiffAVG2	4,83 (48,3%)
9.	DiffForm3	4,80 (48,0%)
10.	ADDGermanyNatGast	4,73 (47,3%)

Tabelle 4.10: Top10-Durchschnittswerte für die Analyse mit Naive Bayes

Die Tabelle ist folgendermaßen zu verstehen, in 5,57 von maximal 10 Folds einer Cross-Validation wurde das Attribut DiffAVG1 von allen 3 Wrappern im Durchschnitt bei Naive Bayes ausgewählt. Damit wurde also das Attribut „DiffAvg1“ von allen anderen Attributen durchschnittlich am häufigsten ausgewählt.

Wie wir sehen haben die Differenzattribute bei Naive Bayes mit Abstand die besten Ergebnisse erzielt. Die ersten 4 Plätze und Plätze 6-9 sind allesamt von Differenzattributen belegt. Ferner ist es auffällig, dass die Differenzausprägungen der Gruppenattribute ein gutes Ergebnis schaffen. Auf den Plätzen 1, 7 und 8 befinden sich jeweils Differenzen der Gruppenattribute, auf Platz 5 ist eine weitere Ausprägung eines Gruppenattributs zu finden. Den abschließenden 10.Platz belegt ein Attribut aus der Gruppe der manuell hinzugefügten Attribute.

Um das Ranking etwas genauer zu analysieren, gehe ich auf die jeweils 3 besten Plätze von im Kapitel 2 vorgestellten Untergruppen ein. Dabei werde ich jedoch auf die Gruppe mit Differenzen nicht eingehen, da die Platzierungen der 3 besten Differenzen (8 Attribute aus der Top10) bereits an der Top10 zu erkennen ist.

Untersuchung der im Kapitel 2 vorgestellten Gruppen bei Naive Bayes

Top 3 aus der Gruppe der „Fußball Studio“-Attribute bei Naive Bayes (Kapitel 2.2.1)

Platz	Attributname	D-F10
59.	TDGast	3,53 (35,3%)
62.	PosGast	3,47 (34,7%)
91.	Gast	3,13 (31,3%)

Tabelle 4.11: Top3-Durchschnittswerte der „Fußball Studio“-Attribute bei Naive Bayes

Man erkennt, dass die „Fußball Studio“-Attribute ziemlich schlecht abgeschnitten haben. Das beste „Fußball Studio“-Attribut befindet sich auf Platz 59. Ferner kann festgestellt werden, dass die Attribute mit den Namen der Heim- und Gastvereine (die sogenannten Vereinsnamen-Attribute) nicht sehr gut bei Naive Bayes abgeschnitten haben. Attribut mit dem Namen der Gastvereine befindet sich auf Platz 91 und der Heimvereine gar auf Platz 133.

Top 3 aus den Gruppenattributen bei Naive Bayes (Kapitel 2.2.2)

Platz	Attributname	D-F10
5.	Form4GastAusw	5,07 (50,7%)
15.	AVG1PlatzGast	4,50 (45,0%)
17.	Form2GastAusw	4,40 (44,0%)

Tabelle 4.12: Top3-Durchschnittswerte der Gruppenattribute bei Naive Bayes

Hier kann erkannt werden, dass die Gruppenattribute sich deutlich besser platziert haben als die z.B. vorher erwähnten „Fußball Studio“-Attribute. Auffällig ist hier wie auch überall bei Naive Bayes, dass die Gastausprägungen der Attribute oft besser platziert sind als die Heimausprägungen. Ebenso ist es auffällig, dass die „Form“ und „AvgPlatz“-Gruppenattribute (sowie deren Differenzausprägungen) von allen Attributen am Besten abschneiden. Gruppenattribut „Abstieg“ (bester Wert Platz 19, Abstieg4Gast, 4,33 Folds) schneidet etwas schwächer und Gruppenattribut „Versus“ (besten Wert Platz 68, Versus2Gast, 3,37 Folds) deutlich schlechter ab.

Top 3 der manuell hinzugefügten Attribute bei Naive Bayes (Kapitel 2.2.3)

Platz	Attributname	D-F10
10.	ADDGermanyNatGast	4,73 (47,3%)
22.	ADDEUFA1JahrGast	4,27 (42,7%)
35.	ADDEwigetabelleGast	4,07 (40,7%)

Tabelle 4.13: Top3-Durchschnittswerte der manuell hinzugefügten Attribute bei Naive Bayes

Die Ergebnisse der manuell hinzugefügten Attribute sind ebenso noch gut. Auch hier ist zu erkennen, dass Gastattribute besser platziert sind, als die dazugehörigen Heimatattribute.

Top 3 der aus anderen kombinierten oder neu erstellten Attribute bei Naive Bayes (Kapitel 2.2.4)

Platz	Attributname	D-F10
71.	Tabvergleich	3,33 (33,3%)
85.	Heimstark	3,20 (32,0%)
96.	Tabtendenz	3,10 (31,0%)

Tabelle 4.14: Top3-Durchschnittswerte der kombinierten bzw. neuerstellten Attribute bei Naive Bayes

Die „kombinierten“ Attribute schneiden von allen Gruppen am schlechtesten ab. Was mit Sicherheit daran liegt, dass viele Attribute aus dieser Gruppe Experimentierattribute waren.

Untersuchung der einzelnen Wrapper und Kommentar zu deren Vorgehensweise bei Naive Bayes:

Wrapper mit Ranksearch (GainRatio) bei Naive Bayes

Top3:

Platz (Ges.)	Attributname	D-F10 bei Ranksearch	D-F10 Gesamt
4.	DIFFUEFA5J	5,00 (50,0%)	5,10 (51,0%)
1.	DIFFAVG1	4,60 (46,0%)	5,57 (55,7%)
13.	DIFFForm5	4,40 (44,0%)	4,57 (45,7%)

Tabelle 4.15: Top3-Durchschnittswerte von WrapperRSG bei Naive Bayes

Generell erfolgte die Attributauswahl bei WrapperRSG mittelhäufig. Im Durchschnitt wurden bei Ranksearch knapp 63 Attribute jedes Mal ausgewählt, also etwas mehr als ein Drittel aller Attribute. Die Foldanzahl variierte stark von 1 bis 10, die meisten Attribute erhielten jedoch 1-4 Folds. Der Durchschnittswert aller Folds bei Ranksearch lag bei 1,23, war also damit recht niedrig.

Wrapper mit LinearForwardSelection bei Naive Bayes

Top3:

Platz (Ges.)	Attributname	D-F10 bei LFS	D-F10 Gesamt
2.	DIFFUEFA1J	4,90 (49,0%)	5,33 (53,3%)
2.	DiffGermanyN	3,70 (37,0%)	5,33 (53,3%)
1.	DiffAvg1	3,50 (35,0%)	4,57 (45,7%)

Tabelle 4.16: Top3-Durchschnittswerte von WrapperLFS bei Naive Bayes

Bei Wrapper mit LinearForwardSelection erfolgte die Auswahl der Attribute sehr selten, im Durchschnitt wurden knapp 17 Attribute bei einem Fold ausgewählt, was in der Gesamtheit etwa ein Zehntel aller Attribute ausmachte. Die Schwankungen um den Mittelwert waren dabei recht klein. Ein Attribut wurde in den meisten Fällen ein- bis zweimal (von 10 möglichen) ausgewählt. Beachtenswert ist ebenso, dass die Top3 der LinearForwardSelection im Großen und Ganzen mit der Gesamt-Top3 bei Naive Bayes übereinstimmt. Der Durchschnittswert von allen Folds ist mit 0,23 mit Abstand der niedrigste von allen 3 Wrappern.

Wrapper mit BestFirst Backward Elimination

Top5 (da hier 5 Attribute mit dem Idealwert vorlagen)

Platz (Ges.)	Attributname	D-F10 bei BestFirstBE	D-F10 Gesamt
27.	Form5Gast	10,00 (100,0%)	4,17 (41,7%)
28.	Form3GastAusw	10,00 (100,0%)	4,13 (41,3%)
30.	Form2Gast	10,00 (100,0%)	4,10 (41,0%)
46.	Avg5Platzgast	10,00 (100,0%)	3,87 (38,7%)
60.	Addnationallaenderspielertop9gast	10,00 (100,0%)	3,53 (35,3%)

Tabelle 4.17: Top3-Durchschnittswerte von WrapperBFBE bei Naive Bayes

Beim Wrapper mit BestFirst Backward Elimination wurden jede Runde alle oder fast alle Attribute ausgewählt, somit wurden jede Runde mindestens 167 und maximal alle 169 Attribute ausgewählt. Die Foldwerte bei Attributen waren meist im hohen (8-9) bis sehr hohem Bereich (10), was man auch

an den Durchschnittswerten der besten Attribute (jeweils 10) erkennen kann. Der Durchschnittswert aller Folds ist mit 8,43 ebenso sehr hoch. Ebenso bemerkenswert ist, dass von den 5 immer ausgewählten Attributen 4 Gruppenattribute sind und 3 davon aus dem Formgruppenattribut stammen. Auffällig ist ebenso, dass der Abstand zwischen Attributen relativ gering ist. Platz 1-70 trennt beispielsweise ein Fold. Dennoch wurden schwächere Attribute bei WrapperBFBE bei Naive Bayes auch schwächer bewertet (Beispiel „Violetore“ hat ein Folddurchschnittswert von 0,20 und „TFHeim“ von 1,20).

Schlechtesten 10, Plätze 160-169

Platz	Attributname	D-Wert Folds aus 10
160.	DiffAustriaN	1,83 (18,3%)
161.	Plusminusgast	1,80 (18,0%)
162.	Remisheim	1,73 (17,3%)
163.	AddAustriaNatGast	1,60 (16,0%)
164.	Remisgast	1,53 (15,3%)
-.	Jahreszeit	1,53 (15,3%)
-.	AddAustriaNatHeim	1,53 (15,3%)
167.	Plusminusheim	1,50 (15,0%)
168.	TFHeim	1,27 (12,7%)
169.	Violetore	0,20 (2,0%)

Tabelle 4.18: Die 10 schlechtesten Attribute bei Naive Bayes

Wie man erkennen kann ist das Testattribut „Violetore“ mit Abstand das schleteste Attribut. Ferner sieht man, dass die Attribute mit österreichischen Nationalspielern (sowohl die Heim-, als auch Gast- und auch die Differenzausprägung) ebenso unter den schlechtesten 10 zu finden sind. Weiterhin findet man in dieser Gruppe die selbstgebauten Plusminusattribute, Remisattribut und das zeitliche Attribut „Jahreszeit“ (das andere zeitliche Attribut „Monat“ findet man auf Platz 152, also auch nicht weit vom Ende entfernt).

Wenn man generell die letzten Tabellenplätze bei Naive Bayes betrachtet kann man ebenso erkennen, dass viele Heim-Attribute sich auf den eher hinteren Plätzen befinden.

Zusammenfassung

- Differenzen erzielen mit Abstand die besten Werte bei Naive Bayes-Analyse. Dazu zählen vor allem die Differenzen von Gruppenattributen und manuell hinzugefügten Attributen
- „Fußball Studio“-Attribute sind eher in der Mitte des Analysefeldes platziert, dabei überraschen Heim und Gastattribute (Attribute mit den Namen des Heim und Gastvereins) mit relativ schlechten Platzierungen in der hinteren Analysefeldhälfte (Gast) oder gar Drittelf (Heim)
- Gruppenattribute und manuell hinzugefügte Attribute schneiden relativ gut ab
- Gastausprägungen der Attribute sind im Allgemeinen deutlich besser platziert als die Heimausprägungen (bei Formattributen die Gastausw-Ausprägungen)
- Die selbsterstellten und neuhinzugefügten Attribute sind bei Naive Bayes überwiegend in der hinteren Hälfte des Analysefeldes platziert

Bei Einzelbetrachtungen von Wrappern bei Naive Bayes hat sich folgendes rausgestellt:

- Bei WrapperLFS ist die Anzahl der ausgewählten Attribute bei jedem Fold sehr gering, das Ergebnis von LinearForwardSelection entsprach im Wesentlichen dem Gesamtergebnis
- Bei WrapperRSG ist die Anzahl der ausgewählten Attribute bei jedem Fold gering bis mittelhoch. Hier wurden deutlich mehr Attribute als bei LinearForwardSelection ausgewählt und auch die

Schwankungen der ausgewählten Attributmenge bei jedem Fold waren deutlich größer. Auch hier jedoch entsprach die Auswahl im Wesentlichen dem Gesamtergebnis

-Beim Wrapper mit BestFirst Backward Elimination ist die Anzahl der ausgewählten Attribute bei jedem Fold hoch. Das Ergebnis spiegelte hier eher nicht das Gesamtergebnis wieder. Allerdings ist der Abstand zwischen dem ersten und den nachfolgenden Plätzen so gering, dass die Ergebnisse ziemlich nah beieinander liegen

-Das Feld der 10 schlechtesten Attribute bilden zum größten Teil viele Testattribute bzw. selbst zusammengestellte und kombinierte Attribute. Mit Abstand das schlechteste Attribut ist „VieleTore“.

4.4.2 JRip

Top10 JRip

Platz	Attributname	D-Wert Folds aus 10
1.	DiffForm3	7,13 (71,3%)
2.	DiffAvg1	6,77 (67,7%)
3.	DiffForm4	6,73 (67,3%)
4.	DiffUEFA5J	6,40 (64,0%)
5.	DiffNatPos	6,37 (63,7%)
6.	DiffForm2	6,30 (63,0%)
7.	DiffNat	6,23 (62,3%)
8.	DiffForm1	6,13 (61,3%)
9.	DiffForm5	6,07 (60,7%)
-.	DiffAvg2	6,07 (60,7%)

Tabelle 4.19: Top10-Durchschnittswerte für die Analyse mit JRip

„DiffForm3“ ist das mit Abstand stärkste Attribut bei JRip. Auffällig ist ebenso, dass die Top10 komplett aus den Differenzattributen gebildet wurde, darunter die gesamte „DiffForm“ und 2 „DiffAvg“-Ausprägungen. Das erste Nichtdifferenzattribut befindet sich erst auf Platz 15 (Abstieg1Gast bzw. Abstieg2Gast). Ebenso ist die hohe Anzahl an Folds auffällig, was mit Sicherheit an den seltsamen Ergebnissen des Backward Elimination-Wrappers liegt (dazu aber später mehr).

Kommen wir jedoch nun zu der Betrachtung der einzelnen Gruppen.

Untersuchung der im Kapitel2 vorgestellten Gruppen bei JRip

Top 3 aus der Gruppe der „Fußball Studio“Attribute bei JRip (Kapitel 2.2.1)

Platz	Attributname	D-F10
74.	TDHeim	4,40 (44,0%)
-.	TDGast	4,40 (44,0%)
85.	SiegeGast	4,23 (42,3%)

Tabelle 4.20: Top3-Durchschnittswerte der „Fußball Studio“-Attribute bei JRip

Diese Gruppe von Attributen ist bei JRip eher mittelmäßig bis schlecht platziert. Man erkennt, dass das erste Attribut hier erst auf Platz 74 zu finden ist. Ebenso schlecht schneiden die Vereinsnamenattribute ab. Attribut „Gast“ ist erst auf Platz 94 zu finden und Attribut „Heim“ ist mit Platz 131 noch weiter unten.

Top 3 aus den Gruppenattributen bei JRip (Kapitel 2.2.2)

Platz	Attributname	D-F10
15.	Abstieg1Gast	5,87 (58,7%)
-.	Abstieg2Gast	5,87 (58,7%)
23.	Form4GastAusw	5,70 (57,0%)

Tabelle 4.21: Top3-Durchschnittswerte der Gruppenattribute bei JRip

Diese Gruppe schneidet ziemlich gut ab und ist, hinter den Differenzen, die beste Gruppe bei JRip. Die anderen 2 Gruppenattribute befinden sich jeweils auf Platz 27 („Avg1PlatzGast“) und auf Platz 95 („Versus1Heim/-Gast“, „Versus2Gast“). Ebenso auffällig ist, dass hier die Gastausprägung deutlich stärker abschneidet, als die Heimausprägung.

Top 3 der manuell hinzugefügten Attribute bei JRip (Kapitel 2.2.3)

Platz	Attributname	D-F10
18.	ADDEwigeTabelleGast	5,80 (58,0%)
31.	ADDUEFA1JahrGast	5,30 (53,0%)
37.	ADDNationalländerstiletop9Heim	4,07 (40,7%)

Tabelle 4.22: Top3-Durchschnittswerte der manuell hinzugefügten Attribute bei JRip

Die Ergebnisse der manuell hinzugefügten Attribute sind noch gut. Man erkennt, dass ein Attribut in der Top20 und weitere 2 im vorderen Viertel des Attributfeldes platziert sind.

Top 3 der aus anderen kombinierten oder neu erstellten Attribute bei JRip (Kapitel 2.2.4)

Platz	Attributname	D-F10
33.	TFHeim	5,27 (52,7%)
78.	Heimstark	4,33 (43,3%)
115.	Haeufigsteplatzierungheim	3,87 (38,7%)

Tabelle 4.23: Top3-Durchschnittswerte der kombinierten bzw. neuerstellten Attribute bei JRip

Wie zu erkennen ist, schneidet nur das Attribut „TFHeim“ aus dieser Gruppe gut ab (ziemlich überraschend, da TFHeim bei Naive Bayes das zweitschlechteste Attribut wurde). Die anderen Attribute schneiden auch hier schlecht ab und befinden sich in der zweiten Hälfte des Attributfeldes.

Untersuchung der einzelnen Wrapper und Kommentar zu deren Vorgehensweise bei JRip

Wrapper mit Ranksearch (GainRatio) bei JRip

Top3:

Platz (Ges.)	Attributname	D-F10 bei Ranksearch	D-F10 Gesamt
1	DiffForm3	8,70 (87,0%)	7,13 (71,3%)
2.	DIFFAVG1	8,50 (85,0%)	6,77 (67,7%)
3.	DiffForm4	8,50 (85,0%)	6,73 (67,3%)

Tabelle 4.24: Top3-Durchschnittswerte von WrapperRSG bei JRip

Wie man bemerken hier kann, spiegelt die Top3 von WrapperRSG die Top3 der Gesamtwertung komplett wieder. Es ist von daher auch nicht verwunderlich, dass WrapperRSG bei JRip die zentrale Rolle für die Gesamtwertung spielte, da WrapperLFS insgesamt eher wenige und WrapperBFBE bei JRip schlichtweg alle Attribute aussuchte. Generell erfolgte die Attributauswahl bei WrapperRSG häufig. Im Durchschnitt wurden bei Ranksearch 132 Attribute pro CV-Durchlauf ausgewählt, also ungefähr drei Viertel aller Attribute (wobei die Abweichungen vom Mittelwert bei einzelnen Tests

recht groß sind). Der Durchschnittswert aller Folds bei Ranksearch lag bei 4,17 und war damit deutlich über den Wert von Naive Bayes.

Wrapper mit LinearForwardSelection bei JRip

Top3

Platz (Ges.)	Attributname	D-F10 bei LFS	D-F10 Gesamt
1.	DiffForm3	2,70 (26,0%)	7,13 (71,3%)
2.	DiffAvg1	2,20 (22,0%)	6,77 (67,7%)
3.	DiffForm4	2,00 (20,0%)	6,73 (67,3%)

Tabelle 4.25: Top3-Durchschnittswerte von WrapperLFS bei JRip

Hier stimmt die Platzierung der Top3 ebenso mit der Gesamtwertung komplett überein, allerdings ist hier die Anzahl der Folds deutlich niedriger.

Bei Wrapper mit LinearForwardSelection erfolgte die Auswahl der Attribute, wie auch bei Naive Bayes, eher selten. Im Durchschnitt wurden knapp 32 Attribute bei einem Fold ausgewählt, was in der Gesamtheit etwa ein Fünftel aller Attribute ausmachte. Durchschnittlich wurde ein Attribut mit 1,54 Folds pro CV ausgewählt.

Wrapper mit BestFirst Backward Elimination bei JRip

Bei JRip wurden leider fast alle Attribute in 10/10 Folds, nur wenige in 9/10 und nur einzelne in 8/10 ausgewählt. Alles in allem macht dies eine TOP3 hier mehr als überflüssig, da den maximalen Durchschnittswert von 10 über alle Trainingssets allein 35 Attribute für sich beanspruchen können (danach folgen noch 30 Attribute mit dem Wert von 9,90). Der Grund für diese Ergebnisse lässt sich nur raten, da die Ursache dafür nicht sehr leicht ausfindig gemacht werden kann. Die Fehlerquellen hierfür können also sehr viele Ursachen haben, von algorithmenspezifischen Besonderheiten bis programminternen Fehlern bei Weka bei der Berechnung.

Im Großen und Ganzen ändert jedoch BestFirst Backward Elimination jedoch nur sehr wenig an den Gesamtergebnissen. Lediglich werden hier die im gesamten Durchschnitt schwächeren (und natürlich auch stärkeren) Attribute deutlich mehr Folds in der Gesamttabelle beinhalten, als z.B. schwächeren (stärkeren) Attribute bei Naive Bayes oder später bei J48, wo BestFirst Backward Elimination adäquatere Ergebnisse liefert hat. Aufgrund der wenig aussagenden Ergebnisse bei JRip BestFirst Backward Elimination, werde ich hier nicht detaillierter auf die Ergebnisse eingehen (detaillierte Ergebnisliste finden Sie im Anhang D).

Schlechtesten Attribute bei JRip Plätze 158-169 (da Plätze 158-161 gleich)

Platz	Attributname	D-Wert Folds aus 10
158.	Gaststark	3,43 (34,3%)
-.	SiegeHeimTest	3,43 (34,3%)
-.	ADD10AbstiegHeim	3,43 (34,3%)
-.	ADDAustriaHeim	3,43 (34,3%)
162.	TabDiffForm	3,40 (34,0%)
-.	TabDiffFormVerg	3,40 (34,0%)
164.	RemisDiff	3,37 (33,7%)
-.	DiffAustrian	3,37 (33,7%)
166.	Versus3Heim	3,33 (33,3%)
-.	DiffAustrian	3,33 (33,3%)
168.	Versus4Heim	3,30 (33,0%)
-.	AddAustriaGast	3,30 (33,0%)

Tabelle 4.26: Die schlechtesten Attribute bei JRip

Unter den schlechtesten Attributen bei JRip sind keine große Überraschungen vorhanden, lediglich die Abwesenheit von Attribut „VieleTore“ und den zeitlichen Attributen überrascht ein wenig (allerdings sind diese Attribute nicht weit von den schlechtesten 10 entfernt). Auffällig ist die hohe Anzahl an Folds bei den schlechtesten Attributen, denn die 2 schwächsten Attribute wurden immerhin durchschnittlich in 1/3 aller Folds ausgewählt. Dieses hohe Fold-Ergebnis lässt sich aber mit den Ergebnissen von WrapperBFBE sehr leicht erklären.

Zusammenfassung

- „DiffForm3“ ist mit Abstand von ca 0,4 Folds das beste Attribut in der JRip-Analyse. Generell wurden hier die ersten 15 Plätze von Differenzen belegt, davon mehr als die Hälfte durch Differenzen von Gruppenattributen

- Bei der Gruppenbetrachtung schneiden die Gruppenattribute als zweitbeste Gruppe (hinter den bereits erwähnten Differenzen) ab, manuell erstellte Attribute sind ebenso noch gut. Enttäuscht haben die „Fußball Studio“-Attribute (wo auch „Heim“ und „Gast“ Vereinsnamenattribute in der zweiten Hälfte des Attributfeldes landeten) und auch hier die „kombinierten“-Attribute, wo nur ein Attribut in unter der Top40 zu finden ist (dieses jedoch sehr überraschend „TFHeim“ ist, das bei Naive Bayes das zweitschlechteste Attribut wurde)

- Bei den Wrappern hatten sowohl WrapperRSG als auch WrapperLFS mit der Gesamtwertung übereinstimmende Ergebnisse, lediglich in der Anzahl der Durchschnittsfolds pro CV unterscheiden sich die beiden deutlich. Auch wurden bei WrapperRSG deutlich mehr Attribute ausgewählt als bei ebendiesem Wrapper bei Naive Bayes. WrapperBFBE wählt auch hier eine zu große Attributmenge in jedem Fold der CV aus, so dass dieser Wrapper hier nur wenig Aussagekraft besitzt. (35 Attribute mit einem Durchschnittsfoldwert von 10, weitere 30-9,90)

- Bei den schlechtesten Attributen fehlen die bisher bekannten Attribute „VieleTore“, sowie „Monat“ und „Jahreszeit“. Ansonsten gibt es hier jedoch keine großen Überraschungen.

4.4.3 J48

Auch hier zuerst die Ergebnisse der Besten 10

Platz	Attributname	D-Wert Folds aus 10
1.	DiffGermanyN	5,00 (50,0%)
2.	DiffForm3	3,53 (35,3%)
3.	Heim	3,20 (32,0%)
-.	Gast	3,20 (32,0%)
-.	DiffForm1	3,20 (32,0%)
6.	DiffNatPos	3,17 (31,7%)
7.	DiffNat	3,00 (30,0%)
8.	DifferenzPunkteHeimGast	2,93 (29,3%)
9.	ADDUefa1JahrHeim	2,90 (29,0%)
10.	DiffForm2	2,87 (28,7%)
-.	ADDAustriaNatHeim	2,87 (28,7%)
-.	DiffUefa1J	2,87 (28,7%)

Tabelle 4.27: Top10-Durchschnittswerte für die Analyse mit J48

Auffällig ist, dass „DiffGermanyN“ bei J48 mit ungefähr 1,5 Folds Abstand die Tabelle anführt. Ebenso auffällig sind die hohen Platzierungen von Vereinsnamenattributen „Heim“ und „Gast“, sowie die Zugehörigkeit zu der Top10 des bisher als sehr schwach betrachteten Attributs der österreichischen Nationalspieler „ADDAustriaNatHeim“. Ebenso ist zunächst zu bemerken, dass die Heim- und Gastausprägungen der Attribute keine gravierenden Unterschiede vorzeigen. Ob dieser Verdacht von den Einzelergebnissen bestätigt wird, wird später zu prüfen sein. Ferner ist hier auch bei J48 die Dominanz der Differenzattribute zu erkennen, sowie das gute Abschneiden der Differenz-Gruppenattribute zu bemerken. Schließlich fällt ebenso auf, dass die Anzahl der Folds in einem J48-CV-Durchgang in der Top10 von J48 sehr niedrig ist. Mit nur 2,87 Folds wurde bereits Platz 10 erreicht, es ist somit von sehr kleinen Abständen zwischen den Attributen auszugehen.

Untersuchung der im Kapitel 2 vorgestellten Gruppen bei J48

Top 3 aus der Gruppe der „Fußball Studio“-Attribute bei J48 (Kapitel 2.2.1)

Platz	Attributname	D-F10
3.	Heim	3,20 (32,0%)
-.	Gast	3,20 (32,0%)
29.	RemisHeim	2,63 (26,3%)

Tabelle 4.28: Top3-Durchschnittswerte der „Fußball Studio“-Attribute bei J48

Die Vereinsnamen-Attribute „Heim“ und „Gast“ sind, wie bereits erwähnt, ziemlich weit vorne. Auch das nächste Attribut „RemisHeim“ befindet sich noch in der Top30. Somit schneiden die „Fußball Studio“-Attribute auch insgesamt gesehen hier deutlich besser ab als bei bisherigen Klassifizierern.

Top 3 aus den Gruppenattributen bei J48 (Kapitel 2.2.2)

Platz	Attributname	D-F10
22.	Form3GastAusw	2,70 (27,0%)
38.	Versus2Heim	2,57 (25,7%)
45.	Avg1PlatzGast	2,50 (25,0%)
-.	Abstieg1Gast	2,50 (25,0%)

Tabelle 4.29: Top3-Durchschnittswerte der Gruppenattribute bei J48

Bei J48, sind die Gruppenattribute zwar gut aber dennoch deutlich schlechter platziert als bei bisherigen Klassifizierern. Es ist auch kein Gruppenattribut unter J48 in der Top20 zu finden.

Top 3 der manuell hinzugefügten Attribute bei J48 (Kapitel 2.2.3)

Platz	Attributname	D-F10
9.	ADDUefa1JahrHeim	2,90 (29,0%)
11.	AddAustriaNatHeim	2,87 (28,7%)
15.	AddAustriaNatGast	2,80 (28,0%)

Tabelle 4.30: Top3-Durchschnittswerte der manuell hinzugefügten Attribute bei J48

Wie bereits erwähnt, überraschen bei J48 die bisher als sehr schwach betrachteten „AustriaNat“-Attribute mit einer relativ guten Platzierung. Ebenso fällt hier auf, dass die Heimausprägungen sogar ein etwas besseres Ergebnis zeigen als die Gastausprägungen. Insgesamt sind die Attribute sowohl im vorderen Feld, als auch unter den schlechtesten 10 zu finden und sind somit im Attributfeld sehr stark verteilt.

Top 3 der aus anderen kombinierten oder neu erstellten Attribute bei J48 (Kapitel 2.2.4)

Platz	Attributname	D-F10
13.	Tabgast	2,83 (28,3%)
26.	TFHeim	2,67 (26,7%)
35.	Gaststark	2,60 (26,0%)

Tabelle 4.31: Top3-Durchschnittswerte der kombinierten bzw. neuerstellten Attribute bei J48

Bei J48 schneiden diese Attribute eher gut und somit auch besser als bei vorangegangenen Klassifizierern ab. Es gibt zwar auch hier Ausnahmen (das Attribut „VieleTore“ ist mit 0 Folds Tabellenletzter), viele Attribute aus dieser Gruppe befinden sich jedoch bei J48 in der ersten Hälfte des Attributfeldes.

Untersuchung der einzelnen Wrapper und Kommentar zu deren Vorgehensweise bei J48:

Wrapper mit Ranksearch (GainRatio) bei J48:

Top3:

Platz (Ges.)	Attributname	D-F10 bei Ranksearch	D-F10 Gesamt
3.	DiffForm1	2,30 (23,0%)	3,20 (32,0%)
2.	DiffForm3	2,10 (21,0%)	3,53 (35,3%)
17.	DiffUefa5J	2,10 (21,0%)	2,80 (28,0%)

Tabelle 4.32: Top3-Durchschnittswerte von WrapperRSG bei J48

Man sieht, dass bei Wrapper mit RankSearch bei J48 die „DiffForm“-Ausprägungen vorne liegen. Ebenfalls zu bemerken ist die niedrige Anzahl der Folds bei hier führenden Attributen.

Insgesamt ist die Vergabe der Folds bei J48-WrapperRSG sehr niedrig (2,36 Folds im Durchschnitt). Auch wurden bei einer Cross-Validation insgesamt in 10 Folds nur wenige Attribute ausgewählt (Durchschnittlich 11,7 Attribute pro CV).

Wrapper mit LinearForwardSelection bei J48:

Top Attribute:

Platz (Ges.)	Attributname	D-F10 bei LFS	D-F10 Gesamt
1.	DiffGermanyN	6,70 (67,0%)	5,00 (50,0%)
6.	DiffNatPos	1,50 (15,0%)	3,17 (31,7%)
4 weitere Attribute mit je 0,70 Folds			

Tabelle 4.33: Top3-Durchschnittswerte von WrapperLFS bei J48

Auch hier sind die Folds bis auf die Ausnahme von „DiffGermanyN“ sehr niedrig. Die vorher erwähnte Ausnahme „DiffGermanyN“ hat dagegen über 5 Folds Vorsprung und legt hier auch den Baustein für den Gesamtsieg bei J48.

Die Anzahl der verteilten Folds ist auch hier unterdurchschnittlich, im Durchschnitt wurden 1,74 Folds pro Attribut verteilt. Die durchschnittliche Anzahl der Attribute bei einer Cross-Validation ist zwar niedrig, liegt aber noch über dem Wert bei WrapperRSG (14,3 Attribute pro CV)

Wrapper mit BestFirst Backward Elimination bei J48

Top 3 Attribute

Platz (Ges.)	Attributname	D-F10 bei Ranksearch	D-F10 Gesamt
3.	Heim	9,60 (96,0%)	3,20 (32,0%)
3.	Gast	9,60 (96,0%)	3,20 (32,0%)
124.	AddAustriaNatTeam	8,60 (86,0%)	2,87 (28,7%)

Tabelle 4.34: Top3-Durchschnittswerte von WrapperBFBE bei J48

Auffällig ist, dass die Vereinsnamenattribute beide die besten Werte bei WrapperBFBE bei J48 erreichen. Ebenso zu bemerken ist wieder die starke Platzierung von „AddAustriaNatTeam“ in dieser Kategorie. Die Differenzen schneiden hier dagegen vergleichsweise relativ schwach ab (nur 2 Differenzen in der Top10).

Insgesamt wurde bei WrapperBFBE ein Attribut durchschnittlich in 6,61 Folds ausgewählt. Fast in jedem Durchgang wurden hier in Cross-Validation 168 Attribute ausgewählt (einmal 167, mit 0 bei Avg2PlatzHeim) ausgewählt. Das einzige Attribut, das bei keinem Durchlauf ausgewählt wurde ist „VieleTore“)

Schlechtesten 10, Plätze 159-169 (da Plätze 159-161 gleich)

Platz	Attributname	D-Wert Folds aus 10
159.	Form5Gast, DiffStadGr, AddStadiongroesseGast	1,67 (16,7%)
162.	AddNationalspieleTop9Gast	1,60 (16,0%)
163.	Versus3	1,53 (15,3%)
-.	Avg1PlatzHeim	1,53 (15,3%)
165.	ADDMittelNatSpGast	1,47 (14,7%)
166.	Avg2PlatzGast	1,27 (12,7%)
167.	DiffAvg2	1,20 (12,0%)
168.	Avg2PlatzHeim	0,93 (9,30%)
169.	Violetore	0,00 (0,00%)

Tabelle 4.35: Die 10 schlechtesten Attribute bei J48

Man sieht, dass bei J48 viele Attribute, die bei anderen Klassifizierern vorwiegend in der Mitte oder gar in der oberen Hälfte des Feldes zu finden sind, sich ganz unten befinden. Überraschend ist auch, dass die gesamte „Avg2“-Ausprägung sich unter den schlechtesten 10, auf den unteren 4 Plätzen, zu finden ist. Das Attribut „VieleTore“ schließt auch hier (wie bei Naive Bayes) das Feld ab.

Zusammenfassung J48

- „DiffGermanyN“ ist bei J48 mit Abstand von ca 1,5 Folds das beste Attribut. Weiter in der Top10 zu finden sind sowohl Differenzen von „Form“-Gruppenattribut, als auch die „Heim“ und „Gast“-Vereinsnamenattribute

- Generell ist die Foldvergabe in einem CV-Durchlauf bei J48 sehr klein. Der Sieger wurde in 50% aller Folds durchschnittlich ausgewählt, der zweitplatzierte gar in 35% aller Folds. Somit sind die Abstände zwischen den Plätzen bei J48 ebenso gering

- Ein überraschend starkes Abschneiden schaffen die Attribute rund um die österreichischen Nationalspieler (3 Attributausprägungen auf den Plätzen 11-20)

- Bei den einzelnen Attributgruppen schneiden die „Fußball Studio“-Attribute und kombinierte Attribute etwas besser, Gruppenattribute etwas schlechter als bei bisherigen Klassifizierern ab. Manuelle Attribute sind im Attributfeld stark verteilt und damit sowohl weit vorne, als auch ganz hinten zu finden

- Bei Untersuchungen der einzelnen Wrapper, fiel bei WrapperRSG der niedrige Wert bei Attributauswahl pro CV-Durchlauf auf, der sogar unter dem Wert der Attributauswahl von WrapperLFS lag. Bei WrapperRSG führten die DiffForm-Ausprägungen das Feld knapp an, bei WrapperLFS gewann das gesamtbeste Attribut „DiffGermanyN“ mit einem großen Abstand von über 5 Folds. Bei WrapperBFBE waren die Vereinsnamenattribute „Heim“ und „Gast“ vorne.

- Unter den schlechtesten 10 sind bei J48 die Attribute zu finden, die normalerweise bei anderen Klassifizierern im Mittelfeld des Attributfeldes zu finden sind. Ebenso überraschend war eine gesamte „Avg2“-Ausprägung unter den schlechtesten 4. Das Attribut „VieleTore“ schließt das Feld mit 0 Folds ab.

4.4.4 SVM

Zunächst die Top10 von SVM:

Platz	Attributname	D-Wert Folds aus 10
1.	DiffAvg1	6,60 (66,0%)
2.	DiffNatPos	6,20 (62,0%)
3.	DiffForm4	5,50 (55,0%)
4.	DiffPunkteHeimGast	5,37 (53,7%)
5.	DiffForm2	5,20 (52,0%)
-.	DiffForm3	5,20 (52,0%)
7.	DiffUefa5J	5,03 (50,3%)
8.	DiffForm1	4,93 (49,3%)
9.	DiffAvg3	4,90 (49,0%)
-.	DiffGermanyN	4,90 (49,0%)

Tabelle 4.36: Top10-Durchschnittswerte für die Analyse mit SVM

Man erkennt, dass bei SVM alle 10 Plätze von Differenzen belegt werden. Ferner sind 7 von diesen 10 Differenzen, Differenzen von Gruppenattributen. Ebenso zu den Top10 gehört das bisher beste Attribut „DiffGermanyN“ und die Attribute „DiffNatPos“ und „DiffPunkteHeimGast“. Das beste Attribut mit einem Vorsprung von 0,4 Folds ist „DiffAvg1“. Auch das zweitplatzierte Attribut hat

einen relativ kleinen Vorsprung von 0,7 Fold gegenüber dem Dritten und auch dahinter sind die Vorsprünge relativ gering. Dennoch ist zu bemerken, dass die durchschnittliche Anzahl der Folds, deutlich höher als beispielsweise bei J48 liegt.

Untersuchung der im Kapitel 2 vorgestellten Gruppen bei SVM

Top 3 aus der Gruppe der „Fußball Studio“-Attribute bei SVM (Kapitel 2.2.1)

Platz	Attributname	D-F10
57.	TDGast	3,47 (34,7%)
75.	PunkteGast	3,33 (33,3%)
86.	NiederlagenHeim	3,30 (33,0%)

Tabelle 4.37: Top3-Durchschnittswerte der „Fußball Studio“-Attribute bei SVM

Wie man an der Top3 erkennen kann, sind nur 2 „Fußball Studio“-Attribute in der ersten Hälfte des Attributfeldes zu finden. Somit schneiden die „Fußball Studio“-Attribute eher schwach ab. Die Vereinnamenbezogenen Attribute „Gast“ und „Heim“ befinden sich unter den schlechtesten 5 („Heim“ ist sogar das zweitschlechteste Attribut).

Top 3 aus den Gruppenattributen bei SVM (Kapitel 2.2.2)

Platz	Attributname	D-F10
12.	Abstieg1Gast	4,73 (47,3%)
13.	Abstieg2Gast	4,60 (46,0%)
19.	Form4GastAusw	4,37 (43,7%)

Tabelle 4.38: Top3-Durchschnittswerte der Gruppenattribute bei SVM

Die Gruppenattribute sind bei SVM gut platziert. Man sieht, dass die Gastausprägungen auch hier weiter vorne sind. Die übrigen besten Ausprägungen der Gruppenattribute sind folgendermaßen platziert - „Avg1PlatzGast“ (Platz 22, 4,23 Folds) und „Versus2Heim“ (Platz 70, 3,37 Folds)

Top 3 der manuell hinzugefügten Attribute bei SVM (Kapitel 2.2.3)

Platz	Attributname	D-F10
25.	ADDNationalaenderspieltop9Heim	4,17 (41,7%)
-.	ADD10AbstiegGast	4,17 (41,7%)
28.	ADDEwigeTabelleHeim	4,13 (41,3%)

Tabelle 4.39: Top3-Durchschnittswerte der manuell hinzugefügten Attribute bei SVM

Viele der Attribute befinden sich in der ersten Hälfte des Attributfeldes oder gar im ersten Drittel. Zu bemerken ist, dass die Gastausprägungen auch hier nicht bevorzugt werden.

Top 3 der aus anderen kombinierten oder neu erstellten Attribute bei SVM (Kapitel 2.2.4)

Platz	Attributname	D-F10
29.	TFHeim	4,07 (40,7%)
70.	Tabtendenz	3,37 (33,7%)
101.	HäufigsteplatzierungHeim	3,23 (32,3%)

Tabelle 4.40: Top3-Durchschnittswerte der kombinierten bzw. neu erstellten Attribute bei SVM

Nur „TFHeim“ überrascht hier mit einem guten Ergebnis. Es hat immer mehr den Anschein, dass das Ergebnis bei Naive Bayes ein Ausrutscher für „TFHeim“ war. Die restlichen Attribute aus dieser Gruppe befinden sich eher weiter hinten.

Untersuchung der einzelnen Wrapper und Kommentar zu deren Vorgehensweise bei SVM

Wrapper mit Ranksearch (GainRatio) bei SVM

Top3

Platz (Ges.)	Attributname	D-F10 bei Ranksearch	D-F10 Gesamt
5.	DiffForm3	6,60 (66,0%)	5,20 (52,0%)
3.	DiffForm4	6,40 (64,0%)	5,50 (55,0%)
1.	DiffAvg1	6,30 (63,0%)	6,60 (66,0%)

Tabelle 4.41: Top3-Durchschnittswerte von WrapperRSG bei SVM

Differenzen von Gruppenattributen belegen hier die ersten Ränge. Im Allgemeinen befinden sich auch hier auf den vorderen 6 Plätzen jeweils Differenzattribute. Durchschnittlich wurden in WrapperRSG bei SVM aus jedem Datensatz 46,4 Attribute ausgewählt. Ein Attribut wurde bei WrapperRSG bei SVM durchschnittlich in 3,4 Folds ausgewählt.

Wrapper mit LinearForwardSelection bei SVM

Top3

Platz (Ges.)	Attributname	D-F10 bei Ranksearch	D-F10 Gesamt
2.	DiffNatPos	5,20 (52,0%)	6,20 (62,0%)
1.	DiffAvg1	4,40 (44,0%)	6,60 (66,0%)
9.	DiffGermanyN	3,00 (30,0%)	4,90 (49,0%)

Tabelle 4.42: Top3-Durchschnittswerte von WrapperRSG bei SVM

Generell sind hier die Folds etwas niedriger, als bei anderen Wrappern bei SVM. Ansonsten ist hier zu bemerken, dass auch hier die Differenzen die vordersten Plätze belegen.

Durchschnittlich wurden bei WrapperLFS pro CV bzw. aus jedem Datensatz 36,3 Attribute ausgewählt, was zwar den niedrigsten Wert von allen Wrappern bei SVM aufweist, dennoch im Vergleich mit anderen Klassifizierern einen ziemlich hoher Wert darstellt. Im Durchschnitt wurde ein Attribut bei WrapperLFS bei SVM in 1,9 Folds ausgewählt.

Wrapper mit BestFirst Backward Elimination bei SVM

Top3

Platz (Ges.)	Attributname	D-F10 bei LFS	D-F10 Gesamt
1.	ADDUefa5JahreHeim	9,90 (99,0%)	3,43 (34,3%)
-.	Abstieg2Heim	9,90 (99,0%)	3,33 (33,3%)
4 weitere Attribute mit je 9,80 Folds			

Tabelle 4.43: Top3-Durchschnittswerte von WrapperBFBE bei SVM

Überraschenderweise gibt es bei dem WrapperBFBE trotz der enorm hohen Foldwerte, kein einziges Attribut, das einen perfekten Wert von 10 aufweisen kann. Dennoch waren hier die Foldwerte im Allgemeinen höher. Es wurde hier bei jedem Durchgang der Cross-Validation 168 von 169 Attributen ausgewählt. Lediglich das Attribut „VieleTore“ wurde mit 0 Folds bei jedem Durchgang auch bei SVM zurückgelassen. Generell lag hier die durchschnittliche Foldanzahl für ein Attribut bei 9,14 Folds, was einen relativ hohen Wert darstellt.

Schlechtesten 10 Plätze, Plätze 158-169 (da Plätze 158-160 gleich)

Platz	Attributname	D-Wert Folds aus 10
158.	TabDiffForm	2,90 (29,0%)
-.	TabDiffFormVerg	2,90 (29,0%)
-.	Abstieg4Heim	2,90 (29,0%)
161.	Versus5	2,83 (28,3%)
-.	Tabgast	2,83 (28,3%)
163.	Tabvergleich	2,77 (27,7%)
164.	Jahreszeit	2,67 (26,7%)
165.	Gast	2,63 (26,3%)
166.	Monat	2,53 (25,3%)
-.	SiegeHeimTest	2,53 (25,3%)
168.	Heim	2,37 (23,7%)
169.	Violetore	0,00 (0,00%)

Tabelle 4.44: Die 10 schlechtesten Plätze bei SVM

Sehr schwach schneiden die Vereinsnamenattribute bei SVM ab. Ferner findet man aber auf den hinteren Plätzen nur noch wenige Überraschungen. Sowohl zeitliche Attribute als auch viele selbstgemachte Attribute sind im Attributfeld auch hier hinten zu finden. Auffällig ist die hohe Foldanzahl auf den hinteren Plätzen. Mit 2,9 Folds wäre man beispielsweise bei J48 noch in der Top10, bei SVM reicht das jedoch nur für die schlechtesten 10.

Zusammenfassung

- „DiffAvg1“ ist das beste Attribut bei SVM. Generell werden die 10 vorderen Plätze allesamt von Differenzen belegt, darunter sind 7 Differenzen von Gruppenattributen. Durchschnittliche Anzahl der Folds bei einer CV liegt deutlich über der von J48.

- „Fußball Studio“-Attribute sowie kombinierte Attribute haben eher mittelmäßig bis schlecht abgeschnitten, dabei belegen Vereinsnamen-Attribute „Heim“ und „Gast“ die unteren Plätze („Gast“ fünfletzten, „Heim“ gar den zweitletzten Platz).

-Gruppenattribute und manuell hinzugefügte Attribute sind eher gut platziert, viele Attribute der beiden Gruppen befinden sich in der ersten Hälfte des Attributfeldes

-Bei WrapperRSG bei SVM belegen die Differenzen von Gruppenattributen allesamt die ersten Ränge

-Auch bei WrapperLFS belegen die Differenzen (nicht nur von Gruppenattributen) die vorderen Plätze. Die Anzahl der ausgewählten Attribute pro Crossvalidation ist für LFS im Vergleich zu vorangegangenen Wrappern mit 36,3 relativ groß

-WrapperBFBE fällt wieder mit hohen Foldwerten auf. Trotz dieser hoher Werte, schafft es kein Attribut 10/10 Folds bei WrapperBFBE bei SVM zu erzielen. Die besten Attribute von WrapperBFBE bei SVM liegen im Gesamtattributfeld eher in der Mitte

- Die größte Überraschung bei den schlechtesten 10 sind, wie bereits erwähnt, die Vereinsnamenattribute. Ansonsten finden sich unter den schlechtesten 10 viele „alte Bekannte“ von anderen Klassifizierern wieder. Auffällig ist auch hier die relativ hohe Foldanzahl im Vergleich zu den bisherigen Klassifizierern

4.4.5 Random Forest

Auch hier erstmal die TOP10:

Platz	Attributname	D-Wert Folds aus 10
1.	DiffGermanyN	7,77 (77,7%)
2.	DiffForm3	5,60 (56,0%)
-.	DiffPunkteHeimGast	5,60 (56,0%)
4.	DiffForm4	5,57 (55,7%)
-.	DiffAvg1	5,57 (55,7%)
-.	DiffAvg3	5,57 (55,7%)
7.	DiffForm2	5,53 (55,3%)
-.	Abstieg1Gast	5,53 (55,3%)
-.	Abstieg2Gast	5,53 (55,3%)
-.	DiffNatPos	5,53 (55,3%)

Tabelle 4.45: Top10-Durchschnittswerte für die Analyse mit Random Forest

Bemerkenswert ist sicherlich auch hier der Sieg von „DiffGermanyN“ mit einem größeren Abstand. Bei Random Forest ist der Abstand sogar mit über 2 Folds etwas deutlicher als bei J48. Mit Sicherheit ist dies auch der deutlichste Abstand aller führenden Attribute in allen Top10 bisher. Die Begründung dieses Abstands wird etwas später bei den Ergebnissen der „LinearForwardSelection“ deutlich. Ebenfalls sehr auffällig ist der sehr kleine Abstand von Platz 2 zu Platz 10 (0,07 Folds), was mit Sicherheit unter anderem auf die Resultate bei „BestFirst Backward Elimination“ und auf die minimalistische Auswahlart bei „LinearForwardSelection“ zurückzuführen ist (dazu später mehr). Ferner fällt wieder auf, dass die Differenzattribute am stärksten abgeschnitten haben und dass mehr als die Hälfte davon aus Differenzen von Gruppenattributen bestehen. Neu ist hingegen, dass 2 Gastausprägungen von Abstieg-Gruppenattributen sich unter den besten 10 befinden.

Untersuchung der im Kapitel 2 vorgestellten Gruppen bei Random Forest

Top 3 aus der Gruppe der „Fußball Studio“-Attribute bei Random Forest (Kapitel 2.2.1)

Platz	Attributname	D-F10
81.	TDHeim	4,53 (45,3%)
90.	SiegeGast	4,47 (44,7%)
102.	PosGast	4,40 (44,0%)

Tabelle 4.46: Top3-Durchschnittswerte der „Fußball Studio“-Attribute bei Random Forest

Auch hier ist das Ergebnis der „Fußball Studio“-Attribute eher mittelmäßig. Das erste Attribut befindet sich in der Mitte, die anderen beiden gar in der zweiten Hälfte des Attributfeldes. Ebenso sind hier die kleineren Abstände auffällig (0,13 Folds bei 21 Plätzen Differenz). Die Vereinsnamen-Attribute „Heim“ und „Gast“ liegen relativ weit unten im Feld. „Gast“ befindet sich auf Platz 125, „Heim“ gar auf Platz 138.

Top 3 aus den Gruppenattributen bei Random Forest (Kapitel 2.2.2)

Platz	Attributname	D-F10
8.	Abstieg1Gast	5,53 (55,3%)
8.	Abstieg2Gast	5,53 (55,3%)
21.	Abstieg3Gast	5,30 (53,0%)

Tabelle 4.47: Top3-Durchschnittswerte der Gruppenattribute bei Random Forest

Hier sieht man, die Gastausprägungen des Abstiegsattributs bei Random Forest am besten abschneiden. Auch bei den anderen Gruppenattributen ist diese Tendenz der „überlegenen

Gastausprägungen“ zu bemerken. Die erste Versus-Ausprägung findet man dabei auf Platz 108 (Versus1Gast), die erste Formattribut-Ausprägung auf Platz 32 und die erste AvgPlatzausprägung auf Platz 28 (AvgPlatz1Gast) der Random Forest-Analyse.

Top 3 der manuell hinzugefügten Attribute bei Random Forest (Kapitel 2.2.3)

Platz	Attributname	D-F10
24.	ADDEwigeTabelleGast	5,27
29.	ADDUefa1JahrGast	5,17
37.	ADDNationalTOP9Gast	5,03

Tabelle 4.48: Top3-Durchschnittswerte der manuell hinzugefügten Attribute bei Random Forest

Diese Attribute sind nach wie vor gut in der ersten Hälfte des Attributfeldes platziert. Auch hier ist das bessere Abschneiden der Gastausprägung bemerkbar.

Top 3 der aus anderen kombinierten oder neu erstellten Attribute bei Random Forest (Kapitel 2.2.4)

Platz	Attributname	D-F10
11.	TFHeim	5,50
31.	Heimstark	5,13
104.	Tabvergleich	4,40

Tabelle 4.49: Top3-Durchschnittswerte der kombinierten bzw. neuerstellten Attribute bei Random Forest

Die Attribute „TFHeim“ und „Heimstark“, beide die großen Überraschungen der Random Forest-Analyse. Nach diesen beiden positiven Überraschungen kommt aber lange nichts mehr aus dieser Gruppe. Die anderen kombinierten oder neu erstellten Attribute sind relativ weit unten im Attributfeld platziert.

Untersuchung der einzelnen Wrapper und Kommentar zu deren Vorgehensweise bei Random Forest

Wrapper mit Ranksearch (GainRatio) bei Random Forest

Top3:

Platz (Ges.)	Attributname	D-F10 bei Ranksearch	D-F10 Gesamt
2.	DiffForm3	7,20 (72,0%)	5,60 (56,0%)
4.	DiffForm4	6,90 (69,0%)	5,57 (55,7%)
4.	DiffAVG1	6,90 (69,0%)	5,57 (55,7%)

Tabelle 4.50: Top3-Durchschnittswerte von WrapperRSG bei Random Forest

Die Ergebnisse des Wrappers mit Ranksearch (GainRatio) als Suchmethode waren bei dem Klassifizierer Random Forest die Aussagestärksten. Ferner wurden hier deutlich mehr Attribute ausgewählt, als beispielsweise bei RankSearch mit Naive Bayes als Klassifizierer.

Auffällig ist hier ebenso, dass die Differenzen der Gruppenattribute die ersten 4 Plätze belegen und insgesamt die Differenzen sich auf den vorderen 14 Plätzen befinden. Durchschnittlich wurden bei Random Forest und Ranksearch 4 Folds pro Attribut vergeben, was, (wenn man nach den Erfahrungen von anderen Klassifizierern urteilt) einen relativ hohen Wert für RankSearch darstellt. Durchschnittlich wurden 152 pro CV ausgewählt.

Wrapper mit LinearForwardSelection bei Random Forest

Top3

Platz (Ges.)	Attributname	D-F10 bei LFS	D-F10 Gesamt
1.	DiffGermanyN	7,40 (74,0%)	7,77 (77,7%)
31.	Heimstark	2,10 (21,0%)	5,13 (51,3%)
11.	TFHeim	1,10 (11,0%)	5,50 (55,0%)

Tabelle 4.51: Top3-Durchschnittswerte von WrapperLFS bei Random Forest

Die Ergebnisse zeigen die große Favorisierung des Attributs „DiffGermanyN“ von der „LinearForwardSelection“-Methode bei Random Forest. Danach folgen mit Abstand von über 5 Folds 2 bisher als Außenseiter abgestempelte Attribute, deren Werte jedoch durch die gute „LinearForwardSelection“-Bewertung in der Gesamtwertung an einigem hinzugewonnen haben. Generell zeigte „LinearForwardSelection“ bei Random Forest eine sehr minimale Auswahl der Attribute pro 1 CrossValidation-Durchgang. (Durchschnittlich nur 5,3 Attribute pro Fold ausgewählt). Durchschnittlich wurden hier 3 Folds pro Attribut verteilt.

Wrapper mit BestFirst Backward Elimination

Bei Random Forest gab es leider wieder die bereits bei JRip (und teilweise bei SVM) aufgetretenen Probleme von zu großen ausgewählten Attributmengen bei einem Fold. Hier erzielten allein 21 Attribute den optimalen Durchschnittswert von 10 (danach folgen noch 35 Attribute mit dem Wert von 9,90).

Da dieses Problem bereits bei JRip ausführlich erläutert wurde, werde ich hier nicht näher darauf eingehen. Detaillierte Ergebnisliste finden Sie im Anhang D.

Schlechtesten 10, Plätze 160-169

Platz	Attributname	D-Wert Folds aus 10
160.	ADDAustriaNatHeim	3,53 (35,3%)
161.	Gaststark	3,50 (35,0%)
-.	ADDAustriaHeim	3,50 (35,0%)
163.	ADDAustriaGast	3,47 (34,7%)
-.	DiffAustria	3,47 (34,7%)
-.	DiffAustriaN	3,47 (34,7%)
166.	Jahreszeit	3,43 (34,3%)
167.	TabDiffFormVerg	3,40 (34,0%)
-.	SiegeHeimTest	3,40 (34,0%)
169.	TabDiffForm	3,37 (33,7%)

Tabelle 4.52: Die 10 schlechtesten Attribute bei Random Forest

Man sieht schon, dass die Foldanzahl bei den schlechtesten Attributen etwa 2 Mal so hoch ist, wie z.B. bei Naive Bayes. Dies geschah vor allem durch die überdurchschnittlichen Foldanzahl-Resultate bei BestFirst Backward Elimination. Ansonsten, bestätigen hier die Ergebnisse der Attribute im Großen und Ganzen den Gesamteindruck von den vorhergehenden Klassifizierern. Es finden sich sowohl die Attribute mit österreichischen Nationalspielern, als auch die zeitlichen und auch selbst erstellten Attribute. Lediglich die Abstände der Folds sind deutlich kleiner als bei den meisten anderen Klassifizierern.

Zusammenfassung

- Bestes Attribut ist, mit einem sehr großen Abstand von über 2 Folds, „DiffGermanyN“. Der große Abstand liegt an der Überlegenheit des Attributs bei der „LinearForwardSelection“. Die Abstände von weiteren Attributen sind relativ klein, es fällt jedoch wieder auf, dass die Differenzattribute (vor allem wieder Differenzen von Gruppenattributen) vorne liegen. Ferner sind auch 2 Ausprägungen des „Abstieg“-Gruppenattributs sind in der TOP10
- Die Ergebnisse der „Fußball Studio“-Attribute sowie die der selbst erstellten Attribute sind wieder (wie bereits bei den vorangegangenen Klassifizierern) mittelmäßig bis schlecht. Die Platzierungen von Gruppenattributen und manuell hinzugefügten Attributen sind eher gut, was bei Gruppenattributen zum Teil an dem guten Ergebnis der Gastausprägungen des Abstiegsattributs liegt. Differenzattribute belegen wie bereits bei den vorangegangenen Klassifizierern die vorderen Plätze.
- Die genauere Betrachtung der einzelnen Wrapper kristallisiert Ranksearch mit GainRatio-Maß als die mit Abstand aussagestärkste Suchmethode bei Random Forest heraus. LinearForwardSelection wählt bei Random Forest nur eine minimale Anzahl von Attributen pro CrossValidation (im Durchschnitt 5,3 Attribute pro CV) aus. Dafür hat WrapperLFS aber mit dem Attribut „DiffGermanyN“ einen ganz klaren Favoriten, der mit über 5 Folds Vorsprung gewinnt und sich damit auch die beste Platzierung in der Gesamtwertung sichert. BestFirst mit Backward Elimination wählt auch hier wie bei JRip eine zu große Attributmenge in jedem Fold der CV aus, so dass das Ergebnis von Wrapper BFBE bei Random Forest nur wenig Aussagekraft besitzt
- Bei den schwächsten Attributen, ist die Foldanzahl (durch die hohen BestFirst BE-Foldwerte) etwa um das Zweifache höher als zum Beispiel bei Naive Bayes. Die Attribute in der Liste der Schwächsten selbst, sind aber im Großen und Ganzen zu anderen Klassifizierern ähnlich. Die Abstände in den Folds zwischen den Attributen sind relativ klein

4.4.6 InfoGain

Zuerst einmal die Top10 von InfoGain:

Platz	Attributname	D-Wert Folds aus 10
1.	Gast	8,49
2.	Heim	7,32
3.	DiffAVG1	6,69
4.	DiffForm1	6,68
5.	DiffGermanyN	6,67
6.	DiffNat	6,55
7.	DiffForm2	6,51
8.	DiffForm4	6,44
9.	DiffNatPos	6,44
10.	DiffForm3	6,41

Tabelle 4.53: Top10-Durchschnittswerte für die Analyse mit InfoGain

Somit kann man erkennen, dass die Vereinsnamen-beinhaltende Attribute „Gast“ und „Heim“ bei InfoGain die größten Werte erzielen könnten. Dabei ist das Attribut mit den Gastvereinen bei InfoGain noch etwas besser bewertet als die Heimvereine. Platz 3-10 (und wenn man einen Blick außerhalb der Top10 in der Durchschnittstabelle wagt, sogar Platz 3-23) bilden ohne Ausnahme die Differenzattribute. Ebenso beachtenswert ist, dass sich unter den besten 10 4 von den 5 Ausprägungen des Attributs „DiffForm“ befinden. Schließlich findet man unter den besten 10 auch die Differenzen von deutschen Nationalspielern und Nationalspielern aus der TOP9-Ländern.

Untenstehend folgt die Analyse von InfoGain. Dabei gehe ich zunächst wieder auf die einzelnen Gruppen aus dem 2.Kapitel ein, nicht jedoch logischerweise auf die einzelnen Wrapper, da InfoGain ein komplett alternatives und eigenständiges Verfahren zum Wrapper ist.

Ebenso werde ich bei der Analyse nicht auf die Gruppe der Differenzen und auf die Gruppenattribute eingehen, da einerseits die Differenzen aus der TOP10 ersichtlich sind und dazu ergänzend die InfoGain-Werte der Gruppenattribute bereits bei der Analyse der Gruppenattribute (Kapitel 4.2) detailliert aufgezeigt wurden.

Untersuchung der im Kapitel2 vorgestellten Gruppen bei InfoGain

Top 3 aus der Gruppe der „Fußball Studio“-Attribute bei InfoGain (Kapitel 2.2.1)

Platz (Ges.)	Attributname	D-Wert Folds aus 10
1.	Gast	8,49
2.	Heim	7,32
108.	TDGast	2,40

Tabelle 4.54: Top3-Durchschnittswerte der „Fußball Studio“-Attribute bei InfoGain

Man erkennt, dass diese Gruppe bei InfoGain deutlich besser abschneidet, als bei den anderen Klassifizierern, was aber daran liegt, dass die Attribute „Gast“ und „Heim“ sehr stark bewertet wurden und die ersten 2 Plätze belegen. Die anderen Attribute fallen sehr deutlich ab, was man auch daran sieht, dass der InfoGain des drittplatzierten Attributs „TDGast“ einen mehr als um das dreifache niedrigeren InfoGain hat, als die beiden Erstplatzierten. Fast alle Attribute aus dieser Gruppe (außer „Heim“ und „Gast“) finden sich im unteren Drittel des Infogain-Rankings wieder.

Top 3 der manuell hinzugefügten Attribute bei InfoGain (Kapitel 2.2.3)

Platz	Attributname	D-F10
38.	ADDUEFA1JahrGast	3,88
49.	ADDGermanyNATGast	3,68
50.	ADDNationalTOP9Heim	3,64

Tabelle 4.55: Top3-Durchschnittswerte der manuell hinzugefügten Attribute bei InfoGain

Die manuell hinzugefügten Attribute haben bei InfoGain etwas schlechter abgeschnitten als bei den Wrappern. Dabei haben hier die Heimausprägungen im Großen und Ganzen aufgeholt, jedoch sind die Gastausprägungen immer noch etwas besser.

Top 3 der aus anderen kombinierten oder neu erstellten Attribute bei InfoGain (Kapitel 2.2.4)

Platz	Attributname	D-F10
36.	Tabtendenz	3,92
64.	Tordiffform	3,27
82.	Heimstark	2,95

Tabelle 4.56: Top3-Durchschnittswerte der kombinierten bzw. neuerstellten Attribute bei InfoGain

Die Attribute aus dieser Gruppe schneiden bei InfoGain etwas besser ab, als bei den anderen Klassifizierern. Dabei überrascht lediglich das Attribut Tabtendenz mit einem relativ hohen InfoGain-Wert positiv, viele andere Attribute aus dieser Gruppe finden sich im unteren Drittel des Attributfeldes wieder.

Die schlechtesten Attribute

Da mehr als 10 (um genauer zu sagen 16) Attribute den InfoGainWert 0 haben, zähle ich diese Attribute hier nichttabellarisch auf.

Plusminusheim, Plusminusgast, VieleTore, TabdiffForm, TabdiffFormVerg, Remisheim, Remisgast, Remisdif, Monat, Jahreszeit, ADDAustriaHeim, ADDAustriaGast, ADDAustriaNatHeim, ADDAustriaNatGast, DiffAustria, DiffAustrian

Wie man unschwer erkennt, befinden sich von den 16 Attributen, alle 6 Attribute die sich mit österreichischen Nationalspielern auseinandersetzen. Ferner tauchen auch hier die zeitlichen Attribute „Monat“ und „Jahreszeit“ auf und auch die Testattribute sind allesamt dabei (unter anderem das schlechteste Attribut „VieleTore“ sowie „Tabdiffform“ -Ausprägungen). Schließlich findet man hier alle Remis-Attribute, die, realistisch gesehen, generell ebenso keine gute Aussagekraft haben, so dass alles in allem das Ergebnis von InfoGain sehr nachvollziehbar erscheint.

Zusammenfassung

- Vereinsnamen beinhaltende Attribute „Gast“ und „Heim“ schneiden bei InfoGain mit Abstand am Besten ab. (im Gegensatz zu anderen Klassifizierer-Tests, wo diese nur im Mittelfeld bzw. unterem Mittelfeld gelandet sind)
- Differenzattribute erzielen auch bei InfoGain sehr gute Ergebnisse und belegen gar in der Gesamtwertung die Plätze 3-23. Differenzen der Gruppenattribute schneiden ebenso gut ab (besonders DiffForm und DiffAVG), die Gruppenattribute selbst haben einen eher mittelmäßigen InfoGain
- Bei der Gruppe der „Fußball Studio“-Attribute schneiden „Gast“ und „Heim“, wie bereits erwähnt, sehr gut ab, die anderen Attribute sind jedoch sehr weit unten im Attributfeld zu finden (Mittelfeld oder gar unteres Drittel)
- die manuell hinzugefügten Attribute schneiden hier eher etwas schlechter ab, als bei den anderen Klassifizierern
- die selbst kombinierten und neu erstellten Attribute schneiden dagegen etwas besser ab. Davon ausgenommen sind die Testattribute, die sich mit einem InfoGain von 0 ganz unten in der Tabelle befinden
- unter den schlechtesten Attributen mit einem InfoGain von 0 befinden sich die selbsterstellten Testattribute, zeitlichen Attribute, sowie die wenig aussagende Attribute mit Remis und rund um die österreichischen Nationalspieler. Somit ist das Ergebnis der schlechtesten Attribute von InfoGain sehr nachvollziehbar.

Damit bin ich mit allen Tests fertig und fasse im nachfolgenden Unterkapitel die erzielten Ergebnisse zusammen.

4.4.7 Fazit und Gesamtergebnisüberblick

Um die Berücksichtigung aller Klassifizierer-Ergebnisse im gleichen Maße zu gewährleisten, habe ich mich entschlossen die Ergebnisse auf folgende Art und Weise primär im wichtigsten Ranking zu bewerten:

1. Von jedem Klassifizierer sowie InfoGain werden die Platzierungen der Attribute übernommen
2. Von diesen Platzierungen wird eine mittlere Platzierung errechnet (alle Platzierungen werden addiert und durch 6 geteilt)
3. Nach dieser mittleren Platzierung wird ein Ranking aufgebaut, in dem nach dieser mittleren Platzierung aufsteigend sortiert wird. Bei Gleichheit an der Spitze zählt die Mehrheit an ersten Plätzen bei den Klassifizierern

Diese Sortierung, die als Basis für viele für Kapitel 5 übernommene Datensätze dienen soll wird in der Zukunft unter dem Namen „Ranking1“ geführt. Es gibt aber auch die „Ranking2“-Methode. Diese besteht aus den gemittelten Wrapper-Ergebnissen. Es berücksichtigt also die Durchschnittsergebnisse der einzelnen Folds.

Abschließend kommt ein Gesamtergebnisüberblick mit Statistiken zu den einzelnen Wrappern und Attributgruppen, bevor die wichtigsten Erkenntnisse aus den vorherigen Unterkapiteln zusammengefasst werden.

Ranking1-Ranking nach Plätzen

Die 10 gesamtbesten Attribute von Ranking1 sind:

Pl.	Attributname	NB	JRip	J48	SV	RF	IG	DPlatz
1.	DiffAvg1	1	2	18	1	4	3	4,83
2.	DiffForm3	9	1	2	5	2	10	4,83
3.	DiffNatPos	6	5	6	2	7	8	5,67
4.	DiffGermanyN	2	22	1	9	1	5	6,67
5.	DiffForm2	11	6	10	5	7	7	7,67
6.	DiffForm4	11	3	18	3	4	8	7,83
7.	DiffUefa5J	4	4	15	7	11	12	8,83
8.	DiffPunkteHeimGast	14	13	8	4	2	17	9,67
9.	DiffForm1	33	8	3	8	11	4	11,17
10.	DiffNat	16	7	7	16	18	6	11,67

Tabelle 4.57: Top10-Ranking1

Wie bereits aus den Einzeltests der Klassifizierer sowie InfoGain zu vermuten war, schneiden die Differenzattribute am besten ab. Das beste Attribut ist „DiffAvg1“, dicht gefolgt von „DiffForm3“. Beide erreichen einen beachtlichen Platzdurchschnitt von 4,83, „DiffAvg1“ hat jedoch eine Erstplatzierung mehr vorzuweisen. Auffällig ist, dass ebenso viele Gruppenattributdifferenzen (insbesondere „DiffForm“) sich in der Top10 befinden.

Ferner werden weiter unten einige weitere interessante Statistiken zu der Gesamtwertung geliefert:

Die meisten Erstplatzierungen

2- DiffAvg1 (Platz 1) und DiffGermanyN (Platz 4)

„DiffGermanyN“ hat ebenso wie der Tabellenführer „DiffAvg1“ 2 erste Plätze vorzuweisen, jedoch macht eine schlechte Platzierung bei JRip eine bessere Gesamtplatzierung als den 4.Platz für „DiffGermanyN“ nicht möglich.

Die Platzierungen der Vereinsnamen-Attribute:

Pl.	Attributname	NB	JRip	J48	SV	RF	IG	DPlatz
72.	Gast	91	94	3	165	125	1	79,83
96.	Heim	133	131	3	168	138	2	95,83

Tabelle 4.58: Platzierung der Vereinsnamen-Attribute bei Ranking1

Man sieht, dass beide Platzierungen nicht gerade gut sind. Es hat den Anschein, dass die Vereinsnamen für die Vorhersage gar nicht so wichtig sind. Auch dies werde ich im 5.Kapitel prüfen, in dem ich die entsprechenden Datensätze gegeneinander antreten lasse. Ferner prüfe ich bereits am Ende des 4.Kapitels im „Vorher-Nachher-Test“ 2 identische Datensätze, einmal mit „Heim“ „Gast“ und einmal ohne diese beiden Attribute.

Top3-Nichtdifferenzattribute

Pl.	Attributname	NB	JRip	J48	SV	RF	IG	DPlatz
16.	Abstieg1Gast	38	15	45	12	7	23	23,33
19.	Avg1PlatzGast	15	27	45	22	28	30	27,83
22.	Form4GastAusw	5	23	72	19	31	29	29,83

Tabelle 4.59: Top3-Nichtdifferenzattribute bei Ranking1

Wie an der Tabelle 4.59 zu erkennen ist, besteht die Top3-der Nichtdifferenzattribute komplett aus Gruppenattribut-Ausprägungen. Aus Interesse werden weiter unten daher noch die Top3-von manuell erstellten, sowie „Fußball Studio“- und Kombinierten Attributen betrachtet. Ebenso fällt auf, dass die Gastausprägungen deutlich die Überhand haben, daher werden auch noch weiter unten die Top3 der Heimausprägungen analysiert.

Top3-manuell erstellte Attribute:

Pl.	Attributname	NB	JRip	J48	SV	RF	IG	DPlatz
30.	AddUefa1JahrGast	22	31	45	94	29	38	43,17
31.	AddEwigeTabelleGast	33	18	51	44	22	95	43,83
38.	AddGermanyNatGast	10	49	72	101	42	49	53,83

Tabelle 4.60: Top3-manuell erstellte Attribute bei Ranking1

Man erkennt, dass die besten manuellen Attribute erst ab dem 30.Platz starten und von allen Attributen die Gastausprägung wieder deutlich besser zu sein scheint.

Top3-„Fußball Studio“- und kombinierten-Attribute:

Pl.	Attributname	NB	JRip	J48	SV	RF	IG	DPlatz
53.	TFHeim	168	33	26	29	11	136	67,17
60.	TDGast	59	74	51	57	85	108	72,33
72.	Gast	91	94	3	165	125	1	79,83

Tabelle 4.61: Top3-„Fußball Studio“- und kombinierte Attribute bei Ranking1

Es ist hier sichtbar, dass die meisten Attribute aus dieser Gruppe eher schwach abgeschnitten haben.

Top3-Heimausprägungen (PLG bedeutet Platz dazugehörige Gastausprägung):

Pl.	PLG	Attributname	NB	JRip	J48	SV	RF	IG	DPlatz
41.	38.	AddGermanyNatHeim	57	63	45	31	71	85	58,67
47.	92.	AddPositionHeim	53	46	127	40	42	74	63,67
53.	150.	TFHeim	168	33	26	29	11	136	67,17

Tabelle 4.62: Top3-Heimausprägungen bei Ranking1

Wie man unschwer an diesem Ranking erkennen kann, sind nicht alle Gastausprägungen als die Heimausprägungen besser, eine etwas genauere Analyse werde ich jedoch etwas später präsentieren.

10 schlechteste Attribute:

Pl.	Attributname	NB	JRip	J48	SV	RF	IG	DPlatz
160.	Versus3Heim	133	166	96	126	151	144	136
161.	Form1HeimHeim	147	135	145	134	128	131	136,67
162.	PlusminusHeim	167	153	79	126	154	154	138,83
163.	SiegeHeimTest	145	158	61	166	167	151	141,33
164.	Monat	152	143	101	166	154	154	145
165.	Versus5Heim	136	153	143	148	146	145	145,17
166.	Jahreszeit	164	143	82	164	166	154	145,5
167.	TabDiffForm	151	162	82	158	169	154	146
168.	TabDiffFormVerg	155	162	109	158	167	154	150,83
169.	VieleTore	169	153	169	169	157	154	161,83

Tabelle 4.63: 10 schlechtesten Attribute bei Ranking1

Unter den 10 schlechtesten Attributen finden sich die zeitlichen Attribute, die kombinierten Attribute, sowie viele Heimausprägungen wieder. Den letzten Platz des Attributfeldes belegt, wie erwartet, das Attribut „VieleTore“.

Somit bin ich am Ende des Rankings nach Plätzen angekommen und komme nun zu einer alternativen Rankingmethode, nämlich dem Ranking nach Folds (Ranking2).

Ranking nach Folds (5 Klassifizierer berücksichtigt, kein InfoGain)

Ranking nach Folds (Ranking2) wurde folgendermaßen erstellt:

-Die Foldergebnisse einzelner Wrapper von allen 5 Klassifizierern wurden addiert, gemittelt und nach diesem Durchschnitt sortiert

Da die Ergebnisse der einzelnen Wrapper schon sehr ausführlich in vorherigen Unterkapiteln diskutiert wurden, beschränke ich mich hier nur auf die Top10, die schlechtesten 10, sowie die Platzierungen von „Heim“ und „Gast“. Platzierungen beim Ranking2 werden im Vergleich zu den Platzierungen des ersten Rankings gesetzt.

Die Besten 10 beim Ranking2:

Pl.	Pl.R1	Attributname	NB	JRip	J48	SV	RF	DPlatz
1.	4.	DiffGermanyN	5,33	5,77	5,00	4,90	7,77	5,75 (57,5%)
2.	1.	DiffAvg1	5,57	6,77	2,73	6,60	5,57	5,45 (54,5%)
3.	3.	DiffNatPos	5,03	6,37	3,17	6,20	5,53	5,26 (52,6%)
4.	2.	DiffForm3	4,80	7,13	3,53	5,20	5,60	5,25 (52,5%)
5.	6.	DiffForm4	4,57	6,73	2,73	5,50	5,57	5,02 (50,2%)
6.	7.	DiffUefa5J	5,10	6,40	2,80	5,03	5,50	4,97 (49,7%)
7.	5.	DiffForm2	4,57	6,30	2,87	5,20	5,53	4,89 (48,9%)
7.	8.	DiffPunkteHeimGast	4,53	6,00	2,93	5,37	5,60	4,89 (48,9%)
9.	11.	DiffUefa1J	5,33	5,80	2,87	4,60	5,37	4,79 (47,9%)
10.	9.	DiffForm1	4,07	6,13	3,20	4,93	5,50	4,77 (47,7%)

Tabelle 4.64: Top10-Ranking nach Folds (Ranking 2)

Wie man erkennt, gibt es bei Ranking2 mit „DiffGermanyN“ einen anderen Tabellenführer als bei Ranking1. Dies hängt vor allem mit den hohen Ergebnissen von diesem Attribut bei J48 und Random Forest, sowie mit dem Weglassen von InfoGain zusammen. Ansonsten unterscheiden sich die

Ergebnisse jedoch nur in Details und in einzelnen wenigen Platzierungen. Insgesamt sind 9 von 10 Attributen sowohl im Ranking1 als auch im Ranking2 in der Top10 zu finden.

Platzierung der Vereinsnamen-Attribute bei Ranking mit Folds:

Pl.	Pl.R1	Attributname	NB	JRip	J48	SV	RF	DPlatz
94.	72	Gast	3,13	4,10	3,20	2,63	4,07	3,43 (34,3%)
134.	96.	Heim	2,77	3,63	3,20	2,37	3,90	3,17 (31,7%)

Tabelle 4.65:Vereinsnamen-Attribute bei Ranking nach Folds

Es ist zu bemerken, dass hier die Attribute noch deutlich schlechter abschneiden, als bei Ranking1. „Gast“ befindet sich in der zweiten Hälfte, „Heim“ befindet sich gar im unteren Drittel des Attributfeldes.

Die schlechtesten 10 beim Ranking nach Folds:

Pl.	Pl.R1	Attributname	NB	JRip	J48	SV	RF	DPlatz
160.	145.	RemisGast	1,53	3,53	2,47	3,20	3,73	2,89 (28,9%)
160.	157.	ADDAustriaGast	1,97	3,30	2,13	3,57	3,47	2,89 (28,9%)
162.	154.	DiffAustria	1,87	3,33	2,53	3,20	3,47	2,88 (28,8%)
163.	163.	SiegeHeimTest	2,47	3,43	2,40	2,53	3,40	2,85 (28,5%)
164.	167.	TabDiffForm	2,23	3,40	2,30	2,90	3,37	2,84 (28,4%)
165.	162.	PlusMinusHeim	1,50	3,50	2,33	3,13	3,57	2,81 (28,1%)
165.	164.	Monat	2,20	3,53	2,20	2,53	3,57	2,81 (28,1%)
167.	168.	TabDiffFormVerg	2,00	3,40	2,17	2,90	3,40	2,77 (27,7%)
168.	166.	Jahreszeit	1,53	3,53	2,30	2,67	3,43	2,69 (26,9%)
169.	169.	VieleTore	0,20	3,50	0,00	0,00	3,53	1,45 (14,5%)

Tabelle 4.66:Die schlechtesten 10 bei Ranking nach Folds

Hier ist der Unterschied zu Ranking1 deutlicher bei der als bei der Top10. Dennoch sind auch hier 7 von 10 Attributen gleich zum ersten Ranking.

Weitere Statistiken aus der Analyse

Nachfolgend erwähne ich noch ein paar Statistiken zu den einzelnen Wrappern. Dabei gehe ich hier vor allem auf die Folds und Attributauswahl ein, betrachte aber ausschließlich WrapperRSG und WrapperLFS. Auf die Betrachtung von WrapperBFBE wird verzichtet, da hier sehr oft die maximale Anzahl an Folds vergeben und immer mindestens 167 (von 169) Attributen ausgewählt wurden.

Folds bei Wrappern

Zuallererst gehe ich auf die Anzahl der Folds bei verschiedenen Wrappern von verschiedenen Klassifizierern ein:

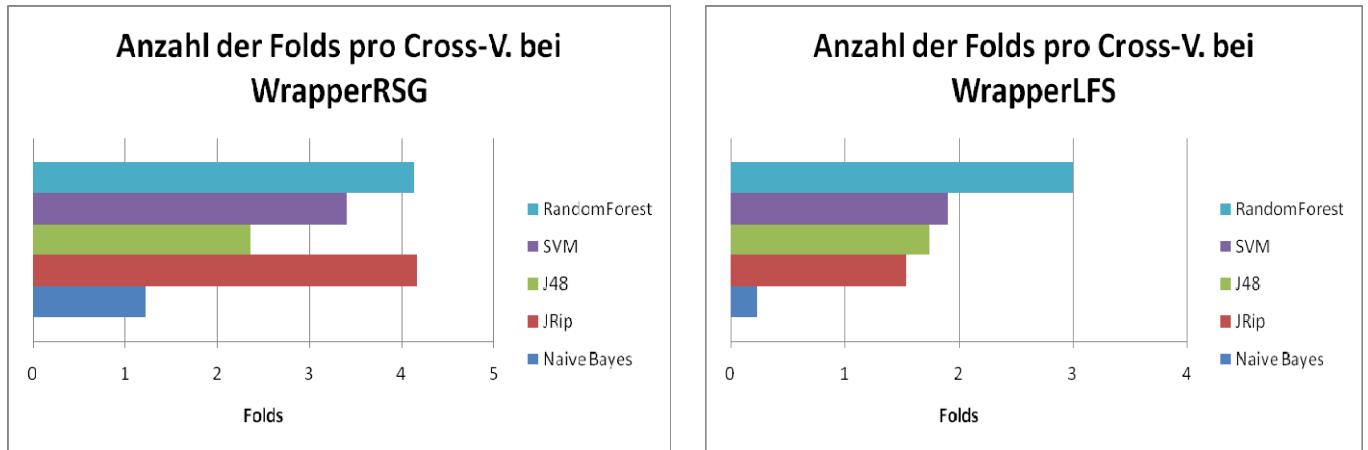


Bild 4.7: Foldverteilung (auf alle Attribute) pro CV bei WrapperRSG und WrapperLFS

Wie zu erkennen ist, wurden bei Wrapper mit Ranksearch von den Klassifizierern JRip und Random Forest die meisten Folds pro CrossValidation vergeben. Bemerkenswert ist, dass es ausgerechnet die beiden Klassifizierer sind bei denen WrapperBFBE nichts aussagende bzw. zu hohe Ergebnisse liefert.

Auch bei WrapperLFS vergibt Random Forest mit Abstand die meisten Folds, hier sind aber nachfolgend SVM, J48 und JRip nah beieinander.

Sowohl bei WrapperLFS als auch bei WrapperRSG hat Naive Bayes mit Abstand den niedrigsten Wert, der bei WrapperLFS sogar nahe 0 liegt.

Attributmenge bei Wrappern

Nun sehe ich mir noch die durchschnittliche Anzahl von ausgewählten Attributen pro CV an:

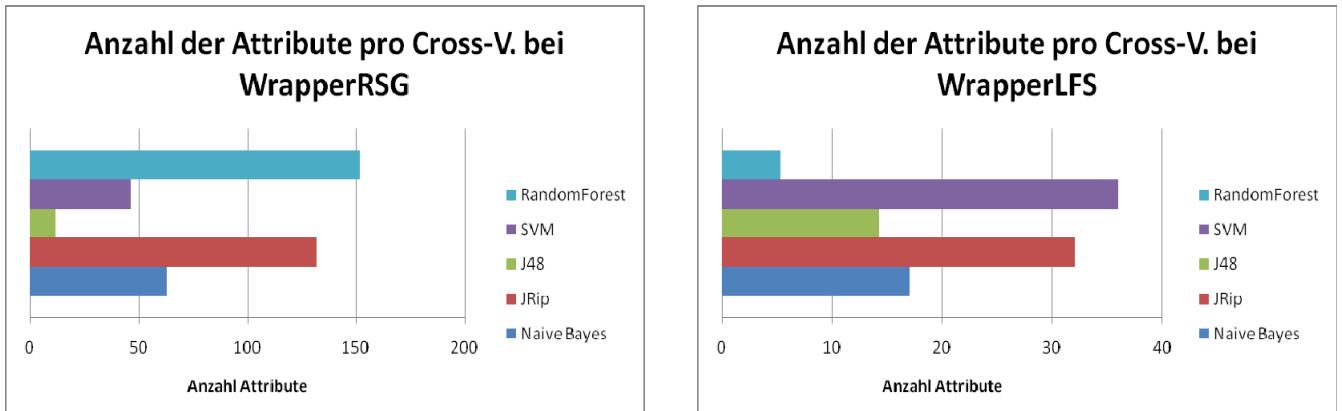


Bild 4.8: Attributverteilung pro CV bei WrapperRSG und WrapperLFS

Es ist zu bemerken, dass bei Wrapper RSG mit Abstand die meisten Attribute wieder bei Random Forest und JRip ausgewählt wurden. Bei WrapperLFS sieht das Bild diesmal jedoch komplett anders aus, hier führt SVM und JRip, bei Random Forest wurden mit Abstand die wenigsten Attribute ausgewählt.

Zusammenfassung der Ergebnisse

Durch die umfassenden Analysen in den vorangegangenen Unterkapiteln konnte ich folgende Ergebnisse festhalten:

Bei Ranking nach Tabellenplätzen

- Hier gewinnt das Attribut „DiffAvg1“ vor „DiffForm3“. Die Top10 besteht aus Differenzen bzw. Differenzen von Gruppenattributen
- Die beste Gruppe nach den Differenzen sind die Gruppenattribute, danach kommen manuell erstellten Attribute und am Ende zu finden sind die kombinierten Attribute und die „Fußball Studio“-Attribute
- Vereinsnamen-Attribute schneiden schwach ab. Dabei ist „Heim“ um ca 20 Plätze schlechter als „Gast“, das sich in der Mitte des Attributfeldes befindet
- Heimausprägungen schneiden deutlich schlechter als die Gastausprägungen ab, wobei es auch hier Ausnahmen gibt (siehe z.B. „TFHeim“)
- Unter den 10 schlechtesten Attributen finden sich die „üblichen Verdächtigen“ (wie u.a. zeitlichen Attribute, kombinierten Attribute, sowie viele Heimausprägungen). Am Ende des Attributfeldes befindet sich das Attribut „VieleTore“

Bei Ranking nach Folds

- Hier gewinnt „DiffGermanyN“, das bei dem Tabellenplatz-Ranking sich auf Platz 4 befindet. Ansonsten ist die Top10 von Ranking nach Folds eher ähnlich zu der Top10 von Ranking nach Tabellenplätzen (9 von 10 Attributen sind bei beiden in der Top10 dabei)
- „Heim“ und „Gast“-Vereinsnamenattribute sind deutlich schlechter platziert als beim ersten Ranking, was sicherlich mit dem Weglassen von InfoGain zusammenhängt
- Bei den schwächsten 10 ist der Unterschied zum ersten Ranking deutlicher bei der als bei der Top10. Aber auch hier sind immerhin noch 7 von 10 Attributen bei beiden Rankings gleich.

Weitere Statistiken

- Bei Folds wurden bei WrapperRSG bei Random Forest und JRip die meisten Folds pro CV vergeben, bei Naive Bayes die Wenigsten. Bei WrapperLFS weist ebenfalls Random Forest die meisten Folds auf, danach folgen SVM, J48 und JRip eng aufeinander, auch hier hat Naive Bayes den geringsten Wert
- Bei Attributen wurden bei WrapperRSG bei Random Forest und JRip die meisten Attribute pro CV-Durchgang ausgewählt. Bei WrapperLFS jedoch wurden bei Random Forest mit Abstand die wenigsten Attribute ausgewählt, die meisten Attribute wurden von JRip und SVM ausgewählt

Somit schließe ich die Analyse ab und beginne mit Rückschlüssen, die eine Basis für einen „BestOf“-Datensatz legen sollen.

4.5. Rückschlüsse aus der Analyse

In diesem Unterkapitel versuche ich basierend auf den Ergebnissen und Erkenntnissen der Analyse Rückschlüsse zu ziehen, die unter anderem später als Basis für eine Erstellung eines „BestOf“-Datensatzes dienen sollen.

Mit einem BestOf-Datensatz ist ein Datensatz gemeint, in dem ich meine Erkenntnisse bezüglich Differenzen, Rankings und Ausprägungen uvm. aus den Kapiteln 4.4 und 4.5 integriere und keine Attribute mit „doppelten“ Daten zulasse.

Hierfür will ich folgende Tests durchführen:

- Im Unterkapitel 4.5.1 gehe ich noch einmal kurz auf das Thema der Gruppenattribute ein, hier betrachte ich die besten Ausprägungen und Belegungen und vergleiche die Ergebnisse mit dem vorangegangenen Test
- Im Unterkapitel 4.5.2 schaue ich mir die Ausprägungen generell an. Hier vergleiche ich sowohl die „Heim“ und „Gast“-Ausprägungen, als auch Differenzen und andere Ausprägungen
- Schließlich werden ich im Unterkapitel 4.5.3 Ideen zur Filterung von schwachen Attributen erwähnt

4.5.1. Gruppenattribute unter Berücksichtigung der Analyse-Ergebnisse

In der Diplomarbeit wurden 4 Gruppenattribute erstellt: „Versus“, „Form“, „AvgPlatz“ und „Abstieg“. Im Kapitel 4.2 wurden die Belegungen dafür mittels InfoGain optimiert, nun will ich die Gruppenattribute in diesem Kapitel unter der Berücksichtigung der Ergebnisse aus der Wrapperanalyse (Kapitel 4.4) näher betrachten und die Besten hiervon auswählen.

Für die Betrachtung der Ergebnisse ziehe ich mir die beiden Rankings aus den vorangegangenen Unterkapiteln zu Rate (bei widersprüchlichen Ergebnissen hat Ranking 1 Vorrang, da es sowohl die Aussagen von InfoGain umfasst, als auch die Klassifizierer alle gleichbehandelt):

1. Versus-Gruppe

Zur Erinnerung:

Im Kapitel 4.2 wurde als beste Ausprägung festgestellt: VersusX

Die Reihenfolge der Belegungen war (nach Stärke geordnet):

Versus1=19, Versus2=18, Versus3=15, Versus4=12, Versus5=14

Die Ergebnisse aus dem Test im Kapitel 4.4. Ranking1:

Pl.	Attributname	NB	JRip	J48	SV	RF	IG	DPlatz
83.	Versus1	77	68	82	119	90	80	86
98.	Versus2	64	86	114	148	81	86	96,5
105.	Versus2Gast	68	95	82	111	121	119	99,33

Tabelle 4.67: Top3-Versus bei Ranking1

Ranking2 :

Pl.	Attributname	NB	JRip	J48	SV	RF	DPlatz
76.	Versus1	3,27	4,47	2,30	3,17	4,47	3,53 (35,3%)
88.	Versus2	3,43	4,20	2,13	3,00	4,53	3,46 (34,6%)
95.	Versus2Gast	3,37	4,07	2,30	3,20	4,10	3,41 (34,1%)

Tabelle 4.68: Vereinsnamen-Attribute bei Ranking2

Beide Tests zeigen übereinstimmend, dass sowohl die VersusX-Ausprägung, als auch Versus1-Belegung als die Beste bestätigt wurde. Somit wird für den BestOf-Datensatz, Versus1 stellvertretend für die Versus-Gruppenattribute übernommen.

2. Form-Gruppe

Zur Erinnerung:

Im Kapitel 4.2 wurde als beste Ausprägung festgestellt: DiffFormX

Die Reihenfolge der Belegungen war (nach Stärke geordnet):

Form1=17, Form2=19, Form3=13, Form4=12, Form5=18

Die Ergebnisse aus dem Test im Kapitel 4.4.

Ranking1:

Pl.	Attributname	NB	JRip	J48	SV	RF	IG	DPlatz
2.	DiffForm3	9	1	2	5	2	10	4,83
5.	DiffForm2	11	6	10	5	7	7	7,67
6.	DiffForm4	11	3	18	3	4	8	7,83

Tabelle 4.69: Top3-Form bei Ranking1

Ranking2:

Pl.	Attributname	NB	JRip	J48	SV	RF	DPlatz
4.	DiffForm3	4,80	7,13	3,53	5,20	5,60	5,25 (52,5%)
5.	DiffForm4	4,57	6,73	2,73	5,50	5,57	5,02 (50,2%)
7.	DiffForm2	4,57	6,30	2,87	5,20	5,53	4,89 (48,9%)

Tabelle 4.70: Top3-Form bei Ranking2

Beide Tests bestätigen übereinstimmend, dass DiffFormX die beste Ausprägung ist. Allerdings wird „DiffForm3“ von beiden als beste Belegung erkannt. Die bisherige beste Belegung „DiffForm1“ wird von den 5 Formbelegungen nur auf dem 4. Platz geführt und taucht nicht einmal in der Top3 auf. Somit wird für den BestOf-Datensatz, „DiffForm3“ stellvertretend für die Form-Gruppenattribute übernommen, statt dem bisherigen im Kapitel 4.2 festgestellten Favorit „DiffForm1“ (der Verlust an InfoGain-Wert ist mit 0,27 nicht sonderlich groß).

3. AvgPlatz-Gruppe:

Zur Erinnerung:

Im Kapitel 4.2 wurde als beste Ausprägung festgestellt: DiffAvgX

Die Reihenfolge der Belegungen war (nach Stärke geordnet):

AvgPlatz1=3, AvgPlatz2=7, AvgPlatz3=2, AvgPlatz4=4, AvgPlatz5=6

Die Ergebnisse aus dem Test im Kapitel 4.4.

Ranking 1:

Pl.	Attributname	NB	JRip	J48	SV	RF	IG	DPlatz
1.	DiffAvg1	1	2	18	1	4	3	4,83
14.	DiffAvg5	7	15	29	16	22	24	18,83
18.	DiffAvg3	38	11	72	9	4	15	24,83

Tabelle 4.71: Top3-AvgPlatz bei Ranking1

Ranking 2:

Pl.	Attributname	NB	JRip	J48	SV	RF	DPlatz
2.	DiffAvg1	5,57	6,77	2,73	6,60	5,57	5,45 (54,5%)
13.	DiffAvg5	4,97	5,87	2,63	4,50	5,27	4,65 (46,5%)
15.	DiffAvg3	4,00	6,03	2,37	4,90	5,57	4,57

Tabelle 4.72: Top3-AvgPlatz bei Ranking2

Beide Tests bestätigen übereinstimmend, dass DiffAvgX die beste Ausprägung ist. Ebenso ganz deutlich zu sehen ist, dass „DiffAvg1“ mit großem Abstand die beste Belegung ist. Somit wird für den BestOf-Datensatz, „DiffAvg1“ stellvertretend für die AvgPlatz-Gruppenattribute übernommen.

4. Abstieg-Gruppe

Zur Erinnerung:

Im Kapitel 4.2 wurde als beste Ausprägung festgestellt: AbstiegXGast

Die Reihenfolge der Belegungen war (nach Stärke geordnet):

Abstieg1=10, Abstieg2=11, Abstieg3=12, Abstieg4=13, Abstieg5=14

Die Ergebnisse aus dem Test im Kapitel 4.4.

Ranking 1:

Pl.	Attributname	NB	JRip	J48	SV	RF	IG	DPlatz
16.	Abstieg1Gast	38	15	45	12	7	23	23,33
23.	Abstieg2Gast	24	15	101	13	7	26	31
24.	Abstieg3Gast	22	18	79	22	21	27	31,5

Tabelle 4.73: Top3-Abstieg bei Ranking1

Ranking 2:

Pl.	Attributname	NB	JRip	J48	SV	RF	DPlatz
16.	Abstieg1Gast	4,00	5,87	2,50	4,73	5,53	4,53 (45,3%)
18.	Abstieg2Gast	4,23	5,87	2,20	4,60	5,53	4,49 (44,9%)
22.	Abstieg3Gast	4,27	5,80	2,33	4,23	5,30	4,39 (43,9%)

Tabelle 4.74: Top3-Abstieg bei Ranking2

Beide Tests bestätigen übereinstimmend, dass AbstiegXGast die beste Ausprägung ist, obwohl diese keine Differenzausprägung ist. Ebenso wird (wenn auch teilweise mit knappem Abstand) bestätigt, dass „Abstieg1Gast“ die beste Belegung ist. Somit wird für den BestOf-Datensatz, „Abstieg1Gast“ stellvertretend für die Abstieg-Gruppenattribute übernommen.

Zusammenfassung:

Somit wurden folgende Attributsausprägungen stellvertretend für alle Gruppenattribute für den BestOf-Datensatz übernommen: „Versus1“, „DiffForm3“, „DiffAvg1“, „Abstieg1Gast“.

Insgesamt betrachtet wurden die Ergebnisse aus dem Kapitel 4.2 durch die Wrapper-Analyse im Großen und Ganzen bestätigt, lediglich die DiffForm-Ausprägungen zeigen ein etwas anderes Ergebnis.

4.5.2. Betrachtung der Ausprägungen

Heim und Gast Ausprägungen

Zunächst vergleiche ich in der nachfolgenden Graphik (Bild 4.9) Heim- und Gastplatzierungen von jeweils einem Attribut. Insgesamt gibt es 55 Attribute, die jeweils in Heim und Gast aufgeteilt sind, also insgesamt 110 Ausprägungen (inklusive „Heim“ und „Gast“-Vereinsnamenattribute).

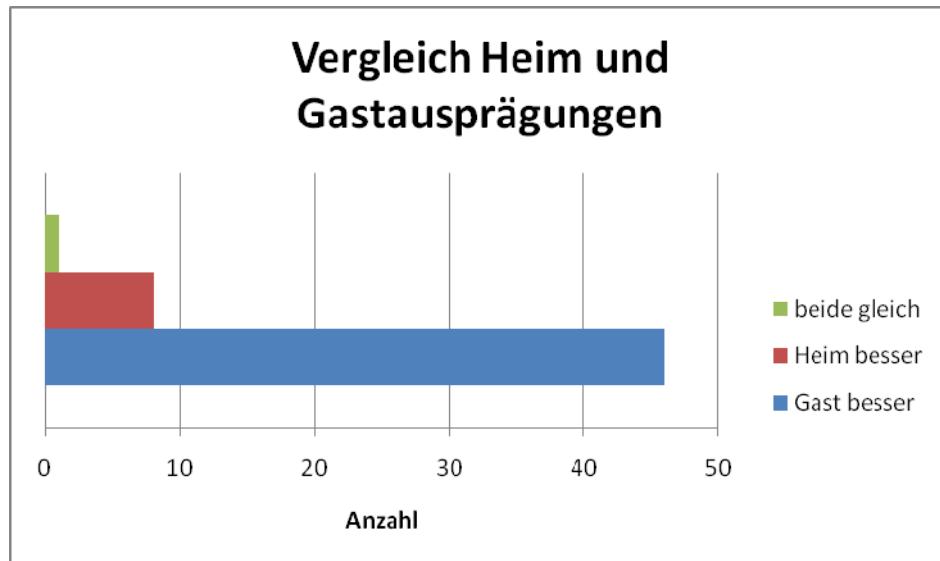


Bild 4.9: Vergleich Heim- und Gastausprägungen (Ranking 1)

Wie an der Graphik deutlich zu erkennen ist, sind Gastausprägungen deutlich besser.

Generell werde ich somit bei BestOf-Datensatz primär Gastausprägungen berücksichtigen. Natürlich ist dies aber auch noch vom Rankingplatz abhängig. Genauer gesagt, sollte der Rankingplatz von beiden gut sein, werden beide Ausprägungen übernommen, ansonsten, bei einem schlechten Rankingplatz kann es auch sein, dass keine Ausprägung übernommen wird.

Vergleich von Differenzen gegen andere Ausprägungen

Nun zu einem Vergleich von Differenzen und anderen Ausprägungen. Es gibt insgesamt 47 Differenzen, mit jeweils 2 oder mehr dazugehörigen Ausprägungen.

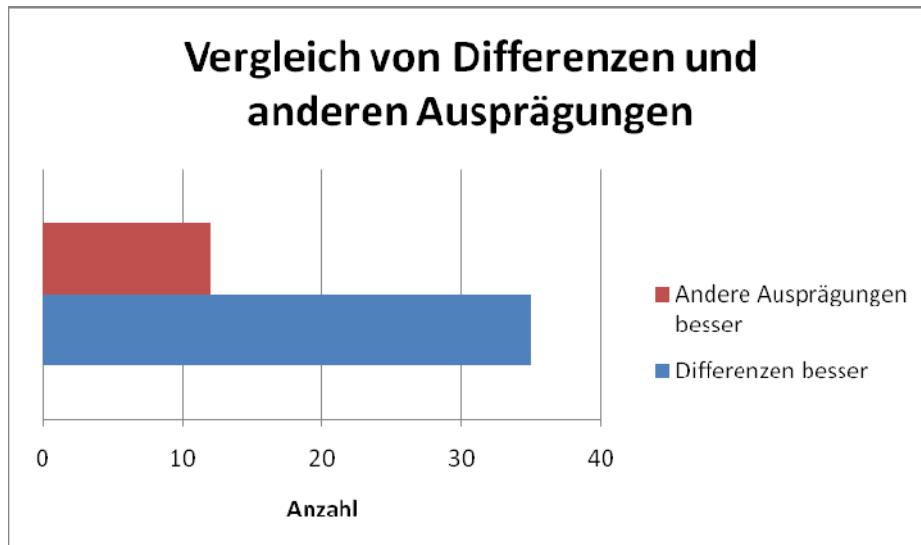


Bild 4.10: Vergleich Differenzen und andere Ausprägungen (Ranking 1)

Man erkennt, dass Differenzen in knapp 3/4 aller Fälle besser sind. Ausnahmen stellen dar: Die 5 Abstiegausprägungen (bei Gruppenattributauswahl bereits berücksichtigt), Remis-, Austria- und AustriaNat-, Brazil-, Stadiongröße-, EwigeTabelle- und 10Abstieg-Attribute. In allen übrigen Fällen sind Differenzen besser als andere Ausprägungen, von daher werden in den meisten Fällen ebendiese Ausprägungen durch die Differenzen ersetzt.

4.5.3. Diskussion über die schwachen Attribute

Sinngemäß ist es lohnend folgende Attribute aus dem Attributfeld insgesamt zu entfernen:

- Zeitliche Attribute (Monat, Jahreszeit)
- Viele schwache selbsterstellte Attribute (VieleTore, TabdiffForm uvm.)
- wenig aussagende Nationalmannschaftattribute (z.B. alles mit Austria bzw. AustriaNat)
- weitere Attribute aus der zweiten Attributhälfte des Gesamrankings (dies sollte man mit verschiedenen Mengen testen).

Zusammengefasst sollte man also bei der Bestimmung des BestOf-Datensatzes folgende Erkenntnisse berücksichtigen:

- Beste Ausprägungen der Gruppenattribute sind: Versus1, DiffForm3, DiffAvg1, Abstieg1Gast
- Gastausprägungen sind deutlich besser als die Heimausprägungen
- Differenzen sind in knapp 3/4 aller Fälle besser als andere Ausprägungen
- schwache Attribute sollten entfernt werden

Im nächsten Kapitel komme ich nun zur Auswahl der Datensätze für die praktischen Tests im 5.Kapitel.

4.6. Auswahl der Datensätze

1. BestOf-Datensatz

Ich führe also die Schritte durch, die im letzten Unterkapitel erwähnt wurden. Zu Erinnerung: Am Anfang hat der Datensatz 170 Attribute (mit dem Klassenattribut „Tendenz“)

1. Entfernung von allen nicht ausgewählten Gruppenattributen

Nach dieser Entfernung soll der Datensatz nur noch 4 (also je eins für jede Gruppe) aus den anfänglich 70 Gruppenattributen mit verschiedenen Ausprägungen und Belegungen haben.

Am Ende dieses Schrittes werden folgende Attribute in die Attributmenge übernommen:
Versus1, DiffForm3, DiffAvg1, Abstieg1Gast

Somit stehen am Ende dieses Schrittes 104 Attribute (von anfänglich 170) in der Attributmenge.

2. Differenzen vs andere Ausprägungen-Betrachtung

Jetzt werden von allen möglichen Ausprägungen der Attribute nur noch die Differenzen übrig gelassen. (Ausnahmen bilden die Attribute: Remis, Austria, AustriaNat, Brazil, Stadiongröße, EwigeTabelle, 10Abstieg; bei diesen Attributen wird jeweils beide Ausprägungen gelassen und die Differenz gelöscht)

Am Ende dieses Schrittes werden folgende Attribute in die Attributmenge übernommen:

Tabdiff, Siegdiff, Remisgast, Remisheim, Niederlagendiff, Punktediff, TDDiff, DifferenzPunkteHeimGast, DiffPktLS, DiffGstLS, DiffKassLS, DiffPosLS, ADDEwigeTabelleHeim, ADDEwigeTabelleGast, DiffNat, DiffNat2, DiffNatPos, DiffNattPos2, DiffNatPosMit, ADDStadiongrößeHeim, ADDStadiongrößeGast, ADD10AbstiegHeim, ADD10AbstiegGast, ADDBrazilHeim, ADDBrazilGast, DiffBrazilN, DiffGermanyN, ADDAustriaHeim, ADDAustriaGast, ADDAustriaNHeim, ADDAustriaNGast, DiffUEFA1J, DiffUEFA5J, DIFFHaeufPlaz

Somit stehen am Ende dieses Schrittes 57 Attribute (von anfänglich 170) in der Attributmenge

3. Eliminieren von schwachen Attributen und Attributsausprägungen:

In diesem Schritt betrachte ich Ranking1 aus dem Kapitel 4.4.7. Anhand dieses Rankings eliminieren wir schwache Attribute oder einzelne Attributsausprägungen. Generell wurde als untere Grenze für die „Eliminierung aufgrund der schlechteren Platzierung“ der 100. Platz genommen.

Am Ende dieses Schrittes werden folgende Attribute aus die Attributmenge eliminiert (in Klammern Gesamtplatzierung im Ranking, schlechtester der Gesamtplatzierung ist Platz 169):

VieleTore (169), TabDiffFormVerg (168), TabDiffForm (167), Jahreszeit(166), Monat (164), SiegeHeimTest (163), PlusMinusHeim (162), PlusMinusGast (152), Gaststark (158), ADDAustriaNatHeim(132), ADDAustriaNatGast (127), ADDAustriaHeim (139), ADDAustriaGast (157), RemisGast (145), RemisHeim (148), TorDiffForm (146), ADDStadiongrößeHeim (142), ADDStadiongrößeGast(131), TabGast (138), Tabvergleich (108), ADDBrazilHeim (125), TFGast(150), ADD10AbstiegHeim (134), ADDBrazilGast (102)

Somit befinden sich am Ende von diesem letzten Schritt 33 Attribute in der Attributmenge.

Am Ende besteht also der „BestOf“ Datensatz aus folgenden Attributen (in Klammern Platzierung in Ranking 1 und die zugehörige Gruppe aus dem 2.Kapitel):

1. *Heim* (96, FA), 2. *Gast* (72, FA), 3. *TabDiff* (65, Diff), 4. *Tabtendenz* (93, Komb), 5. *SiegDiff* (63, Diff),
6. *NiedDiff* (90, Diff), 7. *PunkteDiff* (50, Diff), 8. *TDDiff* (20, Diff), 9. *TFHeim* (53, Komb), 10. *Heimstark* (77, Komb), 11. *Versus1* (83, Grup), 12. *DiffForm3* (1, Grup), 13. *DiffAvg1* (2, Grup), 14.
ADD10AbstiegGast (69, Man), 15. *ADDEwigeTabelleGast* (31, Man), 16. *ADDEwigeTabelleHeim* (61, Man), 17. *Abstieg1Gast* (16, Grup), 18. *DiffPunkteHeimGast* (8, Diff), 19. *DiffPktLS* (21,Diff), 20.
DiffGsTLS (33,Diff), 21. *DiffKassTLS* (34,Diff), 22. *DiffPosLS* (17,Diff), 23. *DiffNat* (10,Diff), 24. *DifNat2* (15,Diff), 25. *DiffNatPos* (3,Diff), 26. *DiffNatPos2* (28,Diff), 27. *DiffNatPosMit* (13,Diff), 28. *DifBrazilN* (99, Diff), 29. *DiffGermanyN* (4,Diff), 30. *DiffUEFA1J* (11,Diff), 31. *DiffUEFA5J* (7,Diff), 32.
DiffHaeufPlaz (36,Diff), 33. *Tendenz* (Klassenattribut)

Dieser Datensatz wird im 5.Kapitel als **BestOf** bezeichnet.

2. Top-Datensätze

Hier werden lediglich die TopX der Ranking1-Tabelle als Datensatz übernommen, unabhängig ob es einige Ausprägungen der Gruppenattribute oder beispielsweise einige Ausprägungen von demselben Attribut sind.

Hier werden folgende Datensätze ins 5.Kapitel übernommen:

- *Top40 Datensatz* (später als **Top40** bezeichnet)
- *Top85 Datensatz* (genau die Hälfte des Attributfeldes, später als **Top85** bezeichnet)

Da in Top40 weder „Heim“ noch „Gast“ Attribut vorhanden ist, würde ich sehr gerne testen ob Vereinsnamenattribute eine größeren Unterschied in der Vorhersage ausmachen, daher kommen noch folgende 2 Datensätze dazu:

- *Top40 Datensatz mit „Heim“ und „Gast“-Attribut* (später als **Top40hg** bezeichnet)

3. Sonstige-Datensätze

Hier werde ich noch einige Datensätze hinzufügen um einige interessante Sachverhalte zu prüfen.

Auf jeden Fall wird der unveränderte Startdatensatz mit 170-Attributen ins 5.Kapitel mit übernommen.

- *Anfangsdatensatz* (später als **Komplett**-Datensatz bezeichnet)

Ebenso wäre ein Infogain Datensatz der verschiedenen Mengen an Attributen umfasst interessant:

- *InfoGain40 Datensatz* (später als **InfoGain40** bezeichnet)
- *InfoGain85 Datensatz* (genau die Hälfte des Attributfeldes, später als **InfoGain85** bezeichnet)

Dazu als Vergleich wären auch Datensätze aus dem Ranking2 interessant (Kapitel 4.4.7)

- *Ranking2 die Besten 40-Datensatz* (später als **Fold40** bezeichnet)
- *Ranking2-die Besten 85 Datensatz* (genau die Hälfte des Attributfeldes, später als **Fold85** bezeichnet)

Schließlich möchte ich noch einzelne Attributgruppen testen:

- Differenzen-Datensatz (ein Datensatz, der nur aus Differenzen besteht, später als **Differenz** bezeichnet)
- „Fußball Studio“-Attribute-Datensatz (ein Datensatz, der nur aus „Fußball Studio“-Attributen besteht, später als **FStudio** bezeichnet)
- manuell hinzugefügte-(Kicker) Attribute-Datensatz (ein Datensatz, der nur aus manuellen Attributen besteht, später als **Kicker** bezeichnet)

Somit werden für Tests im 5.Kapitel folgende 12 Datensätze übernommen:

1. BestOf
2. Top40
3. Top40hg
4. Top85
5. Komplett (Vorherdatensatz)
6. InfoGain40 (IG40)
7. InfoGain85 (IG85)
8. Fold40
9. Fold85
10. Differenz (Diff)
11. FStudio
12. Kicker

Diese Attribute werden im 5. Kapitel sowohl gegeneinander, als auch gegen viele verschiedene einfache Tippstrategien antreten. Bevor ich aber zum „Showdown“ im 5. Kapitel komme, betrachte ich noch die Ergebnisse für diese 11 Datensätze im Vergleich zu den Vorher-Gesamtergebnissen des Komplettdatensatzes am Anfang des Kapitels (Kapitel 4.3).

4.7. Vergleich der Datensätze

In diesem abschließenden Unterkapitel vergleiche ich noch einmal die Ergebnisse von 11 im letzten Kapitel neuerstellten Datensätzen mit dem bereits (im Kapitel 4.3) gemachten (Vorher-)Test vom Komplettdatensatz. Hierzu muss ich die Tabellen wegen der Größe in 5 Teilen je nach Klassifizierer splitten. Aus diesem Grund wird jeder Klassifizierer hier einzeln betrachtet (alle Angaben in %).

Die ausführlichen Ergebnisse der Tests kann man im Anhang E finden. Es folgen nun die größten Auffälligkeiten aus den Tests:

Naive Bayes:

- nur 2 Datensätze verbessern den Komplettdatensatz bei Naïve Bayes insgesamt („BestOf“ und „InfoGain40“)
- „BestOf“ schneidet bei Naive Bayes am Besten, der kleine Datensatz „FStudio“ deutlich am schlechtesten ab
- Bei der Größe der Trainingssätze, fällt nur auf, dass beim größten Datensatz nur der „schlechteste“ Datensatz „FStudio“ eine Verbesserung gegenüber dem „Komplett“-Datensatz einbrachte

JRip:

- Hier haben mehr als die Hälfte der Datensätze (7) die Leistung verbessert
- Der beste Datensatz bei JRip ist der kleinste Datensatz „FStudio“, der schlechteste Datensatz bei JRip ist „BestOf“ (also hier genau entgegengesetzte Ergebnis zu Naive Bayes)
- Bei Größe der Datensätze fällt nur auf, dass die 2 „Größe 5“-Datensätze (mit wenigen Ausnahmen) gut abschneiden

J48:

- Hier haben ebenso mehr als die Hälfte der Datensätze (7) geschafft die Leistung des „Komplett“-Datensatzes zu verbessern, dabei ist die Verbesserung oft beträchtlich (um ca 5%)
- Der beste Datensatz bei J48 ist „Top40hg“. Beachtenswert ist, dass dieser Datensatz bei allen 10 verschiedenen Trainings- und Testsetkombinationen eine Verbesserung erreichte
- Bei der Größe der Datensätze wurden keine besonderen Auffälligkeiten festgestellt, da bei J48 alle Test- und Trainingssetgrößen bei vielen Datensätzen positive Ergebnisse lieferten

SVM:

- Durchwegs haben alle Datensätze bei SVM das Ergebnis des „Komplett“-Datensatzes verbessert
- Die größte Verbesserung erreichte jedoch der „Differenzen-Datensatz“, die kleinste Verbesserung „InfoGain85“-Datensatz. Auffällig ist, dass die fast alle Datensätze SVM in 7 Fällen den „Komplett“-Datensatz verbesserten und in 3 Fällen verschlechterten
- Bei Größen der Datensätze fällt auf, dass bei Trainingssets der Größe 4 bei allen Datensätzen Verbesserungen erreicht wurden

Random Forest:

- Mehr als die Hälfte der Datensätze (8) haben das Ergebnis bei Random Forest gegenüber dem „Komplett“-Datensatz verbessert
- Die größte Verbesserung erreichte „Fold85“-Datensatz, der schlechteste Datensatz bei Random Forest war „FStudio“
- Bei Größe der Datensätze fällt ein positiver Trend bei Trainingssets der Länge 3 auf.

Auffälligkeiten aus der Perspektive der Datensätze:

Zum Schluss noch eine kleine Statistiktafel zu den Datensätzen, die beschreibt in wievielen Fällen eine Verbesserung bzw. eine Verschlechterung erreicht wurde:

Datensatzname	Durchschnitt ↑↓•	Gesamt ↑↓•
BestOf	4↑1↓	29↑17↓4•
Top40	3↑2↓	29↑21↓0•
Top40hg	3↑2↓	29↑21↓0•
Top85	3↑2↓	29↑19↓2•
InfoGain40	5↑0↓	29↑20↓1•
InfoGain85	4↑1↓	27↑23↓0•
Fold40	2↑3↓	24↑24↓2•
Fold85	3↑2↓	28↑22↓0•
Diff	3↑2↓	23↑22↓5•
FStudio	3↑2↓	30↑20↓0•
Kicker	2↑3↓	24↑25↓1•
Gesamt	35↑20↓	301↑234↓15•

Tabelle 4.75: Kurzauswertung der Ergebnisse aus den Vorher-Nachhertests

Überraschend ist, dass der kleinste und bei 2 von 5 am schlechtesten abschneidender Datensatz „FStudio“ die meisten Verbesserungen bietet. Ansonsten verbessert „InfoGain40“-Datensatz als einziger Datensatz bei allen 5 Klassifizierern. Ebenso ist es auffällig, dass bei fast allen Datensätzen mehr Verbesserungen als Verschlechterungen mit den neuen Datensätzen erreicht wurden (mit Ausnahme des „Kicker“ und des „Fold40“-Datensatzes). Somit wurde in 54,7% aller Testfälle eine Verbesserung erreicht.

Somit schließe ich mit den nachfolgenden Tabellen das 4.Kapitel ab und komme zum praktischen Teil.

Kapitel 5: Praktische Tests an den Datensätzen

5.1 Einleitung

In diesem Kapitel werde ich mit den im Kapitel 4 erarbeiteten Datensätzen und mit den eigens in diesem Kapitel hinzugefügten einfachen Tippstrategien praktische Tests starten und einige Benchmarks durchführen sowie Saisontabellen erstellen. Zuallererst starte ich dabei in diesem Unterkapitel mit der Einleitung, Verfahrenserklärung und Auswahl sowie Erläuterung von einfachen Tippstrategien, bevor im Kapitel 5.2 das große Ranking zwischen allen Datensätzen und Tippsstrategien aus unterschiedlichen Perspektiven durchgeführt wird. In diesem Ranking geht es dann einerseits um einfaches Zählen von erratenen Tendenzen und andererseits, in einem zweiten Benchmark, um das Erweitern des Rankings um die Berücksichtigung des „leichten Misses“.

Ein „leichter Miss“ bedeutet, dass eine leichte Abweichung vom richtigen Ergebnis erlaubt wird. Dabei wird angenommen, dass ein Unentschiedentipp beispielsweise näher an einem anderem Ausgang liegt (z.B. bei einem Heimsieg als Ergebnis), als ein Auswärtssiegertipp bei einem Heimsieg als Ergebnis. Zur Verdeutlichung habe ich ein Bild erstellt (Bild 5.1), das diese Abweichungen des Tips von tatsächlichem Ergebnis zeigen soll.

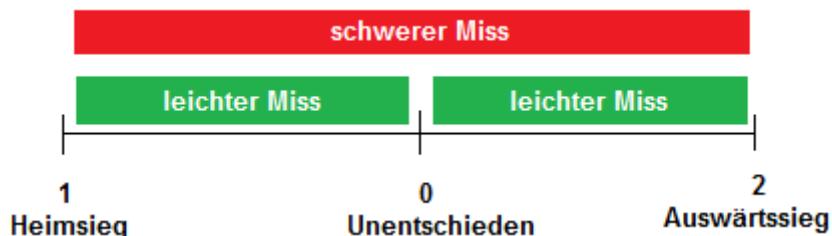


Bild 5.1: Leichter und schwerer Miss

Mit den besten 5 Datensätzen des jeweiligen Klassifizierers geht es dann im Kapitel weiter. Für diese baue ich in der entsprechenden Saison, im Kapitel 5.3 die Bundesliga-Vorhersagetabelle auf und vergleiche bei dieser in einem selbsterstelltem Ranking die Übereinstimmungen mit der realen Tabelle. Kapitel 5.4 beschäftigt sich mit den Vorhersagegenauigkeiten bei den Saisons 2002-2008 mit großen Trainingssets (Trainingssets mit Daten von allen Saisons ab 1963, also ab dem Start der ersten Fußball-Bundesligasaison bis heute). In diesem Kapitel plane ich außerdem 2 unterschiedlich große Datensätze für den Test zu benutzen und bei einer Saison die Tabelle mit allen Klassifizierern basierend auf diesem großen Trainingsset vorherzusagen. Im Kapitel 5.5 wird es, als ein Abschluss des Kapitels, eine qualitative Diskussion geben, in der ich die erhaltenen Ergebnisse zusammenfasse, diskutiere und auf die Vorhersageeigenheiten von jedem Klassifizierer einzugehen versuche. Mit Kapitel 5.6 schließt eine Vorhersage für die zukünftige Saison ab, die ich mit einem Datensatz, den ich basierend auf meinen Erkenntnissen aus der Qualitativen Diskussion auswähle, tätige.

5.1.1 Auswahl von Tippstrategien

Hier werde ich einige einfache Tippstrategien sammeln und erläutern, die später im Unterkapitel 5.2 gegen die Datensätze antreten werden. Viele von diesen Tippstrategien basieren auf meiner jahrelangen Erfahrung als Tippligaleiter. Es gab und gibt immer wieder einige Teilnehmer die diese Strategien verfolgen und anwenden.

1. Tippstrategie: „Immer Heimsieg“

Hier erfolgt schlichtweg immer ein „1-Tipp“, was sicherlich die einfachste Strategie darstellt. Beispielsweise endeten in der Saison 2008/2009 147 Spiele mit einem Heimsieg, was immerhin 48,04% aller Spiele ausmacht. Diese Strategie wird auch, basierend auf meinen Daten, von „ZeroR“ in Weka verwendet.

Diese Tippstrategie wird in späteren Unterkapiteln als „**TSH Heimsieg**“ bezeichnet.

2. Tippstrategie: „Tabellenplatz“

Eine der häufigsten Strategien, die „unerfahrene“ Tipper anwenden. Der höchste aktuelle Tabellenplatz gewinnt. Bei Ausnahmen in Form von gleichen Tabellenplätzen wird ein Unentschieden getippt. Am 1.Spieltag zählt der Tabellenplatz der letzten Saison (bei Aufsteigern-Tabellenplatz 16).

Beispiel

Eintracht Frankfurt (10.Platz) gegen Bayern München (1.Platz) → Tipp 2, da 10.Platz>1.Platz.
Hannover96 (6.Platz) gegen Hertha BSC Berlin (8.Platz) → Tipp 1, da 6.Platz>8.Platz.

Diese Tippstrategie wird in späteren Unterkapiteln als „**TST Tabellenplatz**“ bezeichnet.

3. Tippstrategie: „komplizierter Tabellenplatz“

Eine kleine Abwandlung von der „Tabellenplatz“-Strategie:
-Wenn der Tabellenplatz eines Teams um mehr als 3 höher ist als der Tabellenplatz des anderen Teams, dann gewinnt die entsprechende Mannschaft, sonst Unentschieden. Am 1.Spieltag zählt der Tabellenplatz der letzten Saison (bei Aufsteigern-Tabellenplatz 16).

Beispiel

Eintracht Frankfurt (10.Platz) gegen Bayern München (1.Platz) → Tipp 2, da 10.Platz>1.Platz
Hannover96 (6.Platz) gegen Hertha BSC Berlin (8.Platz) → Tipp 0, da die Differenz zwischen 8.Platz und 6.Platz=2<3

Diese Tippstrategie wird in späteren Unterkapiteln als „**TST Tabellenplatz2**“ bezeichnet.

4. Tippstrategie: „Versus-Letzte Saison“

So wie die Spiele der beiden Mannschaften in der letzten Saison ausgegangen sind, so gehen sie diese Saison auch wieder aus. Sollte mindestens eine von beiden Mannschaften Aufsteiger sein-> Niederlage für den Aufsteiger. Sollten 2 Mannschaften Aufsteiger sein, dann Heimsieg.

Beispiel

Eintracht Frankfurt (letzte Saison 1.Liga, gegen Bayern zuhause verloren) gegen Bayern München (letzte Saison ebenfalls 1.Liga, gegen Frankfurt auswärts gewonnen) → Tipp 2, (da letzte Saison auch „2“)
Mainz05 (Aufsteiger) gegen Hertha BSC Berlin (letzte Saison 1.Liga, gegen Mainz also nicht gespielt) → Tipp 2

Diese Tippstrategie wird in späteren Unterkapiteln als „**TS Versus**“ bezeichnet.

5. Tippstrategie: „Form-letztes Spiel“

Wenn das Heimteam das letzte Spiel gewonnen hat, dann gewinnt es auch dieses Spiel.

Wenn das Auswärtsteam letztes Spiel verloren hat, dann verliert es auch dieses Spiel.

sonst 0. Am ersten Spieltag zählt das letzte Spiel der letzten Saison (da sonst alle Spiele 0 enden)

Beispiel

Eintracht Frankfurt (letztes Spiel verloren) gegen Bayern München (letztes Spiel gewonnen) → Tipp 2

Mainz05 (letztes Spiel Unentschieden) gegen Hertha BSC Berlin (letztes Spiel gewonnen) → Tipp 0

Diese Tippstrategie wird in späteren Unterkapiteln als „**TSFormLS**“ bezeichnet.

6. Tippstrategie: „Form-5 Spiele“

Auch eine oft verwendete Strategie. Der Tipper schaut sich die letzten 5 Spiele beider Vereine an. Die Punktzahl aus diesen 5 Spielen wird addiert und verglichen. Der erfolgreichere Verein siegt, bei gleich erfolgreichen Vereinen Unentschieden.

Diese Tippstrategie wird in späteren Unterkapiteln als „**TSForm5**“ bezeichnet.

7. Tippstrategie: „Platzierung letzte Saison“

Ähnlich wie die Tippstrategie Nr.2, nur wird hier nicht der aktuelle Tabellenplatz betrachtet, sondern Platzierungen in der letzten Saison. Bei Aufsteiger gegen Nichtaufsteiger gewinnt auch hier der Nichtaufsteiger. Bei 2 Aufsteigern - Unentschieden.

Diese Tippstrategie wird in späteren Unterkapiteln als „**TSLS**“ bezeichnet.

8. Tippstrategie: „Durchschnittsplatzierung letzte 3 Saisons“

Hier wird der Durchschnitt der Platzierungen der letzten 3 Saisons des jeweiligen Vereins berechnet und verglichen. Die niedrigere Durchschnittsplatzierung siegt, bei Gleichheit erfolgt ein Unentschieden als Tipp.

Diese Tippstrategie wird in späteren Unterkapiteln als „**TS3LS**“ bezeichnet.

9. Tippstrategie: „Anzahl Siege“

Wenn die Anzahl der Siege einer Mannschaft bis zum aktuellen Spiel die Anzahl der Siege anderer Mannschaft übersteigt, dann gewinnt diese Mannschaft, sonst „0“. Am ersten Spieltag zählen die Ergebnisse der letzten Saison (da sonst alle Spiele Unentschieden enden würden). Dies bedeutet, dass die Aufsteiger am 1.Spieltag verlieren oder untereinander Unentschieden spielen (da beide am 1.Spieltag 0 Siege haben und keine Bundesliga-Statistiken aus der vergangenen Saison vorweisen können).

Diese Tippstrategie wird in späteren Unterkapiteln als „**TSSiege**“ bezeichnet.

10. Tippstrategie: „Anzahl Niederlagen“

Wenn die Anzahl der Niederlagen einer Mannschaft bis zum aktuellen Spiel die Anzahl der Niederlagen anderer Mannschaft übersteigt, dann verliert diese Mannschaft, sonst „0“. Am ersten Spieltag zählen die Ergebnisse der letzten Saison. Ausnahme stellen wieder die Aufsteiger dar, diese verlieren am 1.Spieltag oder spielen bei einem Spiel zweier Aufsteiger gegeneinander Unentschieden. Ferner wird es interessant zu sehen, was im direkten Vergleich der Tippstrategien „Anzahl Siege“ sowie „Anzahl Niederlagen“ mehr Punkte bringt.

Diese Tippstrategie wird in späteren Unterkapiteln als „**TSNied**“ bezeichnet

11. Tippstrategie: „Teamabhängig“

Dies ist ebenfalls eine sehr verbreitete Tippstrategie.

Teilnehmer kennt starke Teams wie „FC Bayern München“, „SV Werder Bremen“, „Bayer Leverkusen“, „VFB Stuttgart“, „FC Schalke 04“, „Hamburger SV“ und evtl. den letztjährigen Meister (sofern dieser nicht eins von den genannten Teams ist), diese tippt er immer als Sieger. Teilnehmer kennt ebenso schwache Teams, die immer unten zu finden sind wie „Arminia Bielefeld“, „1.FC Nürnberg“, „Eintracht Frankfurt“, „MSV Duisburg“, „VFL Bochum“ diese Teams, vorausgesetzt sie sind überhaupt in der 1.Liga, tippt er immer als Verlierer. Generell tippt dieser Tipper auch immer gegen Aufsteiger. In den restlichen Spielen, die von den obigen Bemerkungen nicht abgedeckt wurden, wird Heimspiel getippt.

Diese Tippstrategie wird in späteren Unterkapiteln als „**TSTeams**“ bezeichnet

12. Tippstrategie: „Zufallsstrategie“

Zusätzlich zu den Tippstrategien nehme noch ich einen Zufallstipper auf. Dieser tippt zufällig die Spiele mit den bekannten Tendenzen „0“, „1“ oder „2“. Statistisch gesehen sollten seine Ergebnisse bei knapp 33% liegen. Zufallstipp wird in späteren Unterkapiteln als „**Random**“ bezeichnet

Diese 12 Strategien (Inkl. Randomstrategie), werden zu den im Kapitel 4 erarbeiteten Datensätzen hinzugefügt und treten gegen diese in den Saisons 2005-2008 in nachfolgenden Benchmarks an.

5.1.2 Erläuterung des Teilnehmerfeldes und Ausprägungsformats

Somit lautet das Teilnehmerfeld für die beginnenden Benchmarks im nächsten Kapitel (kursiv Tippstrategien, nicht kursiv Datensätze, unterstrichen die Zufallstrategie):

1. BestOf
2. Top40
3. Top40hg
4. Top85
5. Komplett
6. InfoGain40 (IG40)
7. InfoGain85 (IG85)
8. Fold40
9. Fold85
10. Differenz (Diff)
11. FStudio
12. Kicker
13. *TSHeimsieg*
14. *TSTabellenplatz*
15. *TSTabellenplatz2*
16. *TSVersus*
17. *TSFormLS*
18. *TSForm5*
19. *TSLS*
20. *TS3LS*
21. *TSSiege*
22. *TSNied*
23. *TSTeams*
24. Random

Ich habe also 24 Teilnehmer für die beiden Benchmarks. Es wird jedoch von den Datensätzen einige Ausprägungen mehr geben, denn bei den Datensätzen werde ich (sofern möglich) mit der Trainingssetlänge variieren und unterschiedliche Klassifizierervorhersagen benutzen.

Beispiel:

- „BestOf_3S_Naive“ steht beispielsweise für den „BestOf“-Datensatz mit Trainingssetlänge von 3 Saisons und Naive Bayes als Klassifizierervorhersage auf dem Datensatz
- „Fold40_5S_SVM“ steht für einen „Fold40“-Datensatz mit 5 Saisons Trainingssetlänge und SVM als Klassifizierervorhersage auf dem Datensatz

Generell ist das Format des Namens eines Datensatzes so aufgebaut:

Name des Datensatzes _ Trainingssetlänge _ Klassifizierer

Name des Datensatzes: besteht aus den oben erwähnten Datensätzen

Trainingssetlänge: variiert von 3S=3 Saisons bis 6S=6 Saisons

Klassifizierer: hier werden, die bereits im Kapitel 4 verwendeten, sowie in Kapitel 3 erklärten Klassifizierer benutzt. Also: Naive Bayes, JRip, J48, SVM und Random Forest

Bevor ich mit den Benchmarks beginne, noch eine letzte Anmerkung. Da ich in den nachfolgenden Kapiteln aus Platzmangel nicht die gesamten Benchmarktabellen präsentieren kann, beschränke ich mich auch hier auf die interessantesten Ergebnisse und stelle die gesamten Ergebnisse im Anhang F zur Verfügung.

5.2 Benchmarks: „Datensätze versus Tippstrategien“

5.2.1 Erläuterung der Punkteverteilung

Bevor ich mit den interessantesten Ergebnissen des Tests fortfahre, sollte die Punkteverteilung in diesem Unterkapitel für die verschiedenen Verfahren erläutert werden. Dieses Kapitel berücksichtigt 2 verschiedene Rankings.

Das primäre (Benchmark1) Ranking ist:

- Tendenz erraten, 1 Punkt
- Tendenz nicht erraten, 0 Punkte

Die maximal mögliche Punktzahl beim ersten Ranking beträgt für jede Saison 306 Punkte (da 306 Spiele in einer Saison).

Das sekundäre Ranking (Benchmark2) berücksichtigt die „leichten Misses“ (siehe Bild 5.1).

Für das sekundäre Ranking werden folgende Punkteverteilungen verwendet:

- Tendenz erraten, 2 Punkte
- Tendenz knapp vorbei (leichter Miss), 1 Punkt (Tendenz getippt: „Heimsieg“ (bzw. „Auswärtssieg“) – Ergebnis: „Unentschieden“, oder umgekehrt Tendenz getippt: „Unentschieden“ – Ergebnis „Heimsieg“ (bzw. „Auswärtssieg“))
- Tendenz nicht erraten (schwerer Miss), 0 Punkte

Die maximal mögliche Punktzahl für jede Saison beträgt bei diesem Ranking 612 Punkte (306 Spiele*2 Punkte).

Dieses sekundäre Ranking wird nur am Rande betrachtet, da hier Punkte mit falschtippen erschummelt werden können und das primäre Ranking somit im Vordergrund steht. Es ist lediglich interessant zu wissen, wie die Klassifizierer dort im Vergleich zum 1.Benchmark abschneiden (ob sich das Feld komplett ändert oder die Datensätze, die bei Benchmark1 vorne sind auch bei Benchmark2 vorne bleiben).

5.2.2 Benchmarks: Betrachtung nach Saisons

Saison 2005

Mögliche Trainingssets für Testset Saison 2005: Saisons 2002-2004 (3 Saisons)

In der Saison 2005 waren neben den 10 Tippstrategien, alle Datensatzausprägungen mit der Trainingssetlänge von 3 Saisons möglich (da ich wie bereits erwähnt, von einigen Attributen nur auf Daten bis zur Saison 2002 zurückgreifen konnte). Somit bestand das Testfeld bei der Vorhersage für die Saison 2005 aus 72 Teilnehmern (12 Datensätzen mit der Trainingssetlänge von 3 Saisons mit je einem der 5 Klassifizierer getestet (=60) und den 12 Tippstrategien). Kommen wir nun zu den besten 10 für das Jahr 2005.

Top10

Platz	Name des Datensatzes oder Tippstrategie	Erzielte Punkte (max 306)
1.	FStudio_3S_Naive Bayes	151
2.	BestOF_3S_Naive Bayes	150
-.	InfoGain85_3S_Naive Bayes	150
-.	Fold40_3S_Naive Bayes	150
5.	Differenz_3S_SVM	149
-.	InfoGain40_3S_Naive Bayes	149
-.	Top40hg_3S_Naive Bayes	149
-.	Fold85_3S_Naive Bayes	149
-.	Komplett_3S_SVM	149
10.	Top40hg_3S_JRip	148
-.	Top85_3S_Naive Bayes	148

Tabelle 5.1: Top10 im Benchmark1-Saison 2005

Man erkennt, dass in dieser Saison die Klassifikationsmethode Naive Bayes am Besten voraussagt. Überraschend ist lediglich, dass die beste Vorhersage hier mit dem kleinsten und eigentlich bei Naive Bayes schwächsten Datensatz erreicht wird (FStudio-lediglich 18 Attribute). Ferner finden sich Datensätze von SVM und JRip in der Top10. Auffällig ist ebenso, dass ein Komplettdatensatz (von SVM) sich in der Top10 befindet. Nach Tippstrategien sucht man in der Top10 vergeblich. Generell erreicht die beste Vorhersage für die Saison 2005 etwa 49,4%.

Top3 Tippstrategien

Platz	Name des Datensatzes oder Tippstrategie	Erzielte Punkte (max 306)
13.	TSTeams	146
-.	TSForm5	146
19.	TSLS	144

Tabelle 5.2: Top3 der Tippstrategien im Benchmark1-Saison 2005

Die Tippstrategien haben in dieser Saison mittelmäßig abgeschnitten. So können die „Teams“ und „Form5“-Strategien sowie „Platz Letzte Saison“-Strategie gerade so mit den besten mithalten. Die restlichen Strategien sind in der Mitte oder gar am Ende des Feldes zu finden.

Schlechtesten 10

Platz	Name des Datensatzes oder Tippstrategie	Erzielte Punkte (max 306)
62.	Top85_3S_J48	126
-.	TSVersus	126
-.	TSNied	126
65.	InfoGain85_3S_J48	125
-.	Top40_3S_J48	125
67.	Kicker_3S_J48	120
68.	Komplett_3S_J48	119
-.	Differenz_3S_J48	119
70.	Fold40_3S_J48	111
71.	Fold85_3S_J48	106
72.	Random	89

Tabelle 5.3: Schlechtesten 10 im Benchmark1-Saison 2005

Auffällig ist, dass sich bei den schlechtesten 10 nur Datensätze von J48 befinden. Ferner komplettieren 3 Tippstrategien das Schlechtesten-Feld. Die letzte Platzierung von der Zufallsstrategie ist keine Überraschung, da eine Zufallsstrategie statistisch im Durchschnitt lediglich 33,3% Genauigkeit erzielt (was ca 102 Punkten entspricht), wobei hier der Wert der Zufallsstrategie mit ihrem Punktewert sogar noch unter dem Durchschnitt liegt (ca 29,1%).

Durchschnittsplatzierung der Klassifizierer im Vergleich zu Tippstrategien

Platz	Name des Klassifizierers oder Tippstrategie	Erzielte Punkte (max 306)
1.	Naive Bayes	147,33
2.	SVM	143,00
3.	JRip	138,08
4.	Tippstrategien ohne Random	135,82
5.	Random Forest	132,25
6.	Tippstrategien mit Random	131,92
7.	J48	126,33

Tabelle 5.4: Klassifizierer vs Tippstrategien-Saison 2005

Wie schon aus der Top10 zu erahnen war, schneidet Naive Bayes bei dieser Saison am Besten ab. Generell haben die Datensätze einen Durchschnittspunktewert von 137,4, was um etwa 1,5 Punkte besser ist als der Wert für Tippstrategien ohne Random und um etwa 5,5 Punkte besser ist, als die Tippstrategien mit Randomverfahren.

Durchschnittliche Punktzahl von Datensätzen im Vergleich zu Tippstrategien

Platz	Name des Datensatzes oder Tippstrategie	Erzielte Punkte (max 306)
1.	Top40hg	141,40
2.	FStudio	139,40
3.	InfoGain40	139,20
4.	BestOf	138,80
5.	InfoGain85	138,60
6.	Top85	138,00
7.	Differenz	137,80
8.	Top40	136,80
9.	Tippstrategien ohne Random	135,82
10.	Komplett	135,20
-.	Fold40	135,20
-.	Fold85	135,20
13.	Kicker	133,20
14.	Tippstrategien mit Random	131,92

Tabelle 5.5: Datensätze vs Tippstrategien-Saison 2005

Man erkennt, dass viele Datensätze besser sind als die Tippstrategien in Durchschnitt, sowohl mit als auch ohne Zufallsstrategie. In der Top3 sind 2 Strategien mit der Größe von circa 40 Attributen vertreten (Top40 mit Heim und Gast und auch die besten 40 nach InfoGain). Auch ist es interessant zu sehen, dass die Top40hg um etwa durchschnittlich ca 4,5 Punkte besser ist, als der „Top 40“ Datensatz.

Ergebnisse Ranking mit Abweichungsmöglichkeiten (Benchmark2)

Platz	Platz Bench1	Name des Datensatzes oder Tippstrategie	Erzielte Punkte (max 612)
1.	1.	FStudio_3S_Naive Bayes	402
2.	10.	Top40hg_3S_JRip	400
-.	2.	InfoGain85_3S_Naive Bayes	400
4.	5.	Top40hg_3S_Naive Bayes	399
5.	2.	BestOf_3S_Naive Bayes	398
-.	5.	InfoGain40_3S_Naive Bayes	398
-.	2.	Fold40_3S_Naive Bayes	398
-.	10.	Top85_3S_Naive Bayes	398
9.	5.	Fold85_3S_Naive Bayes	397
-.	5.	Komplett_3S_SVM	397

Tabelle 5.6: Top10 im Benchmark2-2005

Der Führende hat sich in diesem Benchmark2 gegenüber dem 1.Benchmark nicht verändert. Dahinter mischt sich das Feld ein wenig. Generell bleibt die Top10 im Großen und Ganzen gleich.

Zusammenfassung Saison 2005

- Am Besten schneidet hier Naive Bayes mit dem FStudio-Datensatz ab, auch insgesamt befinden sich fast nur Datensätze mit Naive Bayes als Klassifizierer in der Top10 (8 von 10)
- Die Tippstrategien schaffen es nicht in die Top10, es finden sich jedoch mit „TSTeams“, „TSForm5“ und „TSLS“ 3 Strategien in der Top20 wieder
- In der schlechtesten 10 sind 3 Tippstrategien (TSVersus, TSNied und Random) und ausschließlich Datensätze mit J48 als Klassifizierer zu finden. Zufallsklassifizierer schließt das Feld ab.
- Im Gesamtvergleich der Klassifizierer und Tippstrategien schneidet Naive Bayes-Klassifizierer vor SVM am Besten ab. J48 und Random Forest schneiden schlechter als Tippstrategien ab
- Im Gesamtvergleich der Datensätze und Tippstrategien schneiden viele Datensätze besser ab als die Tippstrategien, der beste Datensatz ist hier „Top40hg“ (das im Durchschnitt circa 4,5 Punkte besser ist als der ihm ähnliche Top40-Datensatz)
- Insgesamt sind die Datensätze durchschnittlich um ca 1,5 Punkte besser als die Tippstrategien ohne Random und um ca 5,5 Punkte besser als die Tippstrategien mit Random
- Bei zweiten Benchmark ist auch „FStudio_3S_Naive Bayes“ ganz vorne. Insgesamt ist das Feld von Benchmark2 sehr ähnlich zu Benchmark1

Saison 2006

Mögliche Trainingssets für Testset Saison 2006: Saisons 2003-2005 (3 Saisons), Saisons 2002-2005 (4 Saisons)

In der Saison 2006 waren Datensätze mit Trainingssetlängen von 3 und 4 Saisons möglich. Somit bestand das Testfeld bei der Vorhersage für die Saison 2005 aus 132 Teilnehmern. Im Folgenden werden die Ergebnisse für das Jahr 2006 dargestellt.

Top10

Platz	Name des Datensatzes oder Tippstrategie	Erzielte Punkte (max 306)
1.	BestOf_4S_SVM	150
-.	InfoGain40_4S_SVM	150
3.	Top40_4S_SVM	148
-.	Kicker_4S_SVM	148
5.	Fold40_4S_SVM	147
-.	Fold85_4S_SVM	147
7.	Fold40_3S_SVM	146
-.	Fold40_4S_JRip	146
-.	FStudio_4S_SVM	146
10.	Differenz_3S_SVM, InfoGain85_3S_JRip, BestOf_3S_SVM, InfoGain40_3S_JRip, InfoGain_40_4S_JRip	145

Tabelle 5.7: Top10 im Benchmark1-Saison 2006

Man erkennt, dass in dieser Saison SVM mit einer Trainingssetlänge von 4 Saisons triumphiert. Die meisten Punkte erzielen die Datensätze „BestOf“ und „InfoGain40“. Tippstrategien sind in dieser Saison ebenso wie in der Saison 2005 nicht unter den besten 10 zu finden. Generell erreicht die beste Vorhersage für die Saison 2005 etwa 49,0%.

Top3 Tippstrategien

Platz	Name des Datensatzes oder Tippstrategie	Erzielte Punkte (max 306)
22.	TSTeams	141
-.	TS3LS	141
45.	TSForm5, TSLS	136

Tabelle 5.8: Top3 der Tippstrategien im Benchmark1-Saison 2006

Die Tippstrategien haben hier noch schlechter abgeschnitten als in der Saison 2005. Dies liegt mit Sicherheit auch daran, dass diese Saison mehr Datensatzausprägungen zur Verfügung standen. Die 2 besten Tippstrategien sind hier „TSTeams“ und „TS3LS“. Beide schneiden aber generell etwas schlechter als in der Saison 2005 ab. Auf dem dritten Platz befinden sich die Strategien „TSForm5“ und „TSLS“ gefolgt von der einfachsten Strategie „TSHeimsieg“. Damit schneiden „TSLS“ und „TSHeimsieg“ etwas besser als in der letzten Saison ab.

Die schlechtesten Datensätze oder Tippstrategien

Platz	Name des Datensatzes oder Tippstrategie	Erzielte Punkte (max 306)
122.	Fold40_4S_J48	113
123.	TSTabellenplatz2	112
-.	Top40hg_4S_Random Forest	112
-.	InfoGain85_4S_RandomFor.	112
126.	Top40hg_3S_Random Forest	111
127.	TSVersus	109
128.	FStudio_4S_Random Forest	105
129.	Differenz_3S_J48	104
130.	Kicker_4S_J48	103
131.	Top85_4S_J48	97
132.	Random	96

Tabelle 5.9: Schlechtesten 10 im Benchmark1-Saison 2006

In Gegensatz zur letzten Saison befinden sich nicht nur J48 Datensätze unter den schlechtesten 10, sondern auch Random Forest-Datensätze. Ferner komplettieren auch 3 Tippstrategien das Schlechtesten-Feld. Die letzte Platzierung wird auch diesmal von der Random-Strategie belegt (allerdings nun mit einem etwas besseren Wert von 31,4%).

Durchschnittsplatzierung der Klassifizierer im Vergleich zu Tippstrategien

Platz	Name des Klassifizierers oder Tippstrategie	Erzielte Punkte (max 306)
1.	SVM	141,21
2.	JRip	138,38
3.	Naive Bayes	133,25
4.	Tippstrategien ohne Random	128,45
5.	Tippstrategien mit Random	125,75
6.	J48	124,54
7.	Random Forest	122,13

Tabelle 5.10: Klassifizierer vs Tippstrategien-Saison 2006

Wie schon aus der Top10 zu erahnen war, schneidet SVM bei dieser Saison am Besten ab. Generell haben die Datensätze mit einem Durchschnittspunktwert von 131,9 einen um ca 3,5 Punkte besseren Wert als der Wert für Tippstrategien ohne Random und einen um 6 Punkte besseren Wert gegenüber den Tippstrategien mit Randomverfahren.

Durchschnittliche Punktanzahl von Datensätzen im Vergleich zu Tippstrategien

Platz	Name des Datensatzes oder Tippstrategie	Erzielte Punkte (max 306)
1.	InfoGain40	136,30
2.	BestOf	135,70
3.	Top40	134,50
4.	Fold40	133,70
5.	Fold85	133,30
6.	InfoGain85	131,80
7.	Top40hg	131,60
8.	FStudio	130,00
9.	Komplett	129,50
10.	Top85	128,70
11.	Kicker	128,90
12.	Differenz	128,80
13.	Tippstrategien ohne Random	128,45
14.	Tippstrategien mit Random	125,75

Tabelle 5.11: Datensätze vs Tippstrategien-Saison 2006

Man erkennt, dass diesmal alle Datensätze besser sind als die Tippstrategien. Die beste Strategie ist „InfoGain40“-Strategie knapp vor „BestOf“-Strategie und der „Top40“-Strategie.

Zum Schluss noch ein kleiner Vergleich der einzelnen Trainingssetlängen. Hier erzielen die 3Saisons-Trainingssets mit 132,15 Punkten einen etwas besseren Durchschnitt als 4Saisons-Trainingssets mit 131,65 Punkten.

Ergebnisse Ranking mit Abweichungsmöglichkeiten (Benchmark2)

Platz	Platz Bench1	Name des Datensatzes oder Tippstrategie	Erzielte Punkte (max 612)
1.	36.	Top40_3S_J48	388
2.	1.	BestOf_4S_SVM	383
-.	36.	InfoGain85_3S_Random Forest	383
4.	3.	Kicker_4S_SVM	381
-.	8.	FStudio_4S_SVM	381
6.	14.	InfoGain40_3S_J48	380
-.	1.	InfoGain40_4S_SVM	380
8.	8.	Fold40_4S_JRip	379
9.	11.	BestOf_3S_SVM	378
10.	5.	Fold85_4S_SVM	377

Tabelle 5.12: Ergebnisse im Benchmark2-2006

Hier wurde das Ergebnis im Vergleich zum Benchmark1 richtig durcheinander gewürfelt. An erster Stelle ist dabei mit dem „J48“-Datensatz eine wirkliche Überraschung zu finden. Dieser Datensatz war im 1.Benchmark auf dem 36.Platz und wurde durch die vielen „leichten Misses“ nach vorne geschoben. Auch danach schaffen es noch ein weiterer J48- und sogar ein Random Forest-Datensatz in die Top10

des zweiten Benchmarks aufgenommen zu werden. Die Erstplatzierten des 1.Benchmarks sind auf Platz 2 bzw. Platz 6 zu finden.

Zusammenfassung Saison 2006

- Am Besten schneidet in dieser Saison SVM mit dem „BestOf“ und mit dem „InfoGain“-Datensatz ab. Auch insgesamt befinden sich überwiegend nur Datensätze mit SVM als Klassifizierer in der Top10. Die restlichen Plätze werden von JRip-Datensätzen belegt.
- Die Tippstrategien schaffen es in dieser Saison weder in die Top10, noch in die Top20.
- In der schlechtesten 10 sind 3 Tippstrategien und nur Datensätze mit J48 und Random Forest als Klassifizierer zu finden. Zufallsklassifizierer schließt das Feld wieder ab
- Im Gesamtvergleich der Klassifizierer und Tippstrategien schneidet SVM-Klassifizierer vor JRip am Besten ab. J48 und Random Forest landen im Vergleich sogar noch hinter den Tippstrategien im Feld
- Im Gesamtvergleich der Datensätze und Tippstrategien schneiden alle Datensätze besser als die Tippstrategien ab. Der „InfoGain40“-Datensatz liegt diesmal ganz vorne vor dem „BestOf“-Datensatz
- Bei Vergleich der Trainingssetlängen schneidet die Trainingssetlänge mit 3 Saisons etwas besser (ca 0,5 Punkte) ab, als die Trainingssetlänge von 4 Saisons
- Insgesamt sind die Datensätze durchschnittlich um ca 3,5 Punkte besser als die Tippstrategien ohne Random und um 6 Punkte besser als die Tippstrategien mit Random
- Im zweiten Benchmark gewinnt überraschenderweise ein J48 („Top40_3S_J48“)-Datensatz, der sich durch viele „leichte Misses“ nach vorne schieben könnte

Saison 2007

Mögliche Trainingssets für Testset Saison 2007: Saisons 2004-2006 (3 Saisons), Saisons 2003-2006 (4 Saisons), Saisons 2002-2006 (5 Saisons)

In der Saison 2007 stehen nun für die Datensätze die Trainingssetlängen von 3, 4 und 5 Saisons zur Verfügung. Somit besteht das Testfeld bei der Vorhersage für die Saison 2005 aus 192 Teilnehmern.

Hier eine Übersicht der Ergebnisse für die Saison 2007.

Top10

Platz	Name des Datensatzes oder Tippstrategie	Erzielte Punkte (max 306)
1.	Differenz_4S_SVM	159
-.	InfoGain85_3S_JRip	159
3.	Differenz_4S_Jrip	158
-.	TSTeams	158
-.	Top40_5S_SVM	158
6.	Fold85_4S_Naive Bayes	157
-.	Differenz_5S_SVM	157
8.	Fold85_3S_SVM	156
-.	Fold40_3S_SVM	156
10.	Top40_5S_JRip,Fold40_5S_JRip, Fold40_5S_SVM, InfoGain40_5S_Naive Bayes	155

Tabelle 5.13: Top10 im Benchmark1-Saison 2007

Auch hier sind SVM zusammen mit JRip die herausragenden Klassifizierer. Bemerkenswert ist, dass sich hier zum ersten Mal eine Tippstrategie in der Top10 befindet und dies sogar auf dem 3.Platz Generell erreicht die beste Vorhersage für die Saison 2007 etwa 52,0%.

Top3 Tippstrategien

Platz	Name des Datensatzes oder Tippstrategie	Erzielte Punkte (max 306)
3.	TSTeams	158
49.	TS3LS	150
102.	TSHeimsieg, TSLS	143

Tabelle 5.14: Top3 der Tippstrategien im Benchmark1-Saison 2007

Zum ersten Mal ist eine Tippstrategie in der Top10 und dann gleich auf dem 3.Platz. Danach folgt allerdings lange nichts mehr. Während „TS3LS“ es noch in die Top50 schafft finden sich die nächstbesten Strategien „TSLS“ und „TSHeimsieg“ bereits außerhalb der Top100 wieder.

Die schlechtesten Datensätze oder Tippstrategien

Platz	Name des Datensatzes oder Tippstrategie	Erzielte Punkte (max 306)
184.	Fold40_4S_J48	117
-.	Kicker_5S_J48	117
186.	Kicker_3S_J48	116
-.	Top40_5S_J48	116
188.	Top40_3S_J48	115
-.	Top85_3S_J48	115
190.	Fold40_3S_J48	114
191.	Fold85_3S_J48	111
192.	Fold85_3S_J48	109
-.	Differenz_5S_J48	109
194.	Random	100

Tabelle 5.15: Schlechtesten 10 im Benchmark1-Saison 2007

Hier besteht die schlechteste 10 wieder nur aus J48-Attributen und der Zufallsstrategie, die wie auch bisher das Feld abschließt (mit 32,7% Genauigkeit).

Durchschnittsplatzierung der Klassifizierer im Vergleich zu Tippstrategien

Platz	Name des Klassifizierers oder Tippstrategie	Erzielte Punkte (max 306)
1.	SVM	149,11
2.	JRip	149,06
3.	Naive Bayes	147,92
4.	Tippstrategien ohne Random	136,45
5.	Random Forest	134,47
6.	Tippstrategien mit Random	133,42
7.	J48	127,67

Tabelle 5.16: Klassifizierer vs Tippstrategien-Saison 2007

Diesmal ist das Ergebnis bei der Top3 ziemlich knapp, SVM ist lediglich um 0,05 Punkte besser als JRip und danach folgt mit etwa 1 Punkt Abstand Naive Bayes. Random Forest und J48 sind wieder schwächer als die Tippstrategien. Generell haben die Datensätze mit einem Durchschnittspunktwert von 141,64 einen um ca 5 Punkte besseren Wert, als den für Tippstrategien ohne Random und um etwa 8 Punkte besseren Wert gegenüber den Tippstrategien mit Randomverfahren.

Durchschnittliche Punktzahl von Datensätzen in Vergleich zu Tippstrategien

Platz	Name des Datensatzes oder Tippstrategie	Erzielte Punkte (max 306)
1.	InfoGain40	145,27
2.	Top40hg	144,40
3.	BestOf	143,13
4.	InfoGain85	142,53
5.	Top85	142,20
6.	Top40	141,80
7.	Differenz	141,67
8.	Fold40	141,47
9.	Fold85	141,27
10.	Kicker	140,20
11.	FStudio	138,40
12.	Komplett	137,40
13.	Tippstrategien ohne Random	136,45
14.	Tippstrategien mit Random	133,42

Tabelle 5.17: Datensätze vs Tippstrategien-Saison 2007

Auch in der dritten betrachteten Saison, sind die Datensätze besser als die Tippstrategien (sowohl mit als auch ohne Zufallsstrategie). Die beste Strategie ist hier „InfoGain40“ vor „Top40hg“.

Vergleich der unterschiedlichen Trainingssetlängen

Platz	Trainingssetlängen	Erzielte Punkte (max 306)
1.	4 Saisons	142,75
2.	5 Saisons	141,95
3.	3 Saisons	140,23

Tabelle 5.18: Analyse der Trainingssetlängen 2007

Auch hier, wie in der vorherigen Saison schneidet der Trainingsset mit der Größe von 4 Saisons am Besten ab, dicht gefolgt von 5 Saisons. Trainingssetlänge von 3 Saisons schneidet diesmal am schlechtesten ab.

Ergebnisse Ranking mit Abweichungsmöglichkeiten (Benchmark2)

Platz	Platz Bench1	Name des Datensatzes oder Tippstrategie	Erzielte Punkte (max 612)
1.	66.	Top40hg_3S_SVM	398
-.	1.	InfoGain85_3S_JRip	398
3.	1.	Differenz_4S_SVM	397
-.	5.	Differenz_5S_SVM	397
5.	7.	Fold85_3S_SVM	396
-.	5.	Fold85_4S_Naive Bayes	396
-.	35.	Komplett_4S_Naive Bayes	396
-.	9.	InfoGain40_5S_Naive Bayes	396
-.	2.	Top40_5S_SVM	396

Tabelle 5.19: Ergebnisse im Benchmark2-Saison 2007

Man sieht, dass der 66. aus dem ersten Benchmark zusammen mit dem 1. InfoGain85_3S_JRip im zweiten Benchmark das Feld anführt. Auch sonst sind viele Datensätze dabei, die im ersten Benchmark nicht mal in die Top50 gekommen sind und hier von den „leichten Misses“ profitieren.

Zusammenfassung Saison 2007

- In der Top10-Tabelle schneiden SVM und JRip am Besten ab, auch Naive Bayes ist in der Top10 vorhanden. Die besten Datensätze sind „Differenz_4S_SVM“ und „InfoGain85_3S_JRip“.
- Zum ersten Mal ist mit „TSTeam“ eine Tippstrategie in der Top10 und sogar auf Platz3, danach folgt aber erst auf Platz 49 folgt eine weitere Tippstrategie
- Unter den schlechtesten 10 finden sich ausschließlich Tippstrategien und Datensätze mit J48 wieder. Wie auch in den vorherigen 2 Saisons schließt der Zufallsklassifizierer das Feld ab.
- Im Gesamtvergleich der Klassifizierer und Tippstrategien schneidet SVM-Klassifizierer diesmal nur sehr knapp als erster vor JRip und Naive Bayes ab. J48 und Random Forest landen wieder hinter den Tippstrategien
- Im Gesamtvergleich stellen „InfoGain40“, „Top40hg“ und „BestOf“-Datensätze die Top3 zusammen, im Durchschnitt sind wieder alle Datensätze besser als die Tippstrategien
- Trainingssetlänge „4 Saisons“ sind auch hier etwas besser als die anderen Trainingssetlängen von „5“ und „3 Saisons“
- Insgesamt sind die Datensätze auch hier durchschnittlich um circa 5 Punkte besser als die Tippstrategien ohne Random und um circa 8 Punkte besser als die Tippstrategien mit Random
- Im zweiten Benchmark gewinnt (zusammen mit „InfoGain85_3S_JRip“) mit „Top40hg_3S_SVM“ ein Datensatz, der im Benchmark1 lediglich auf dem 66.Platz zu finden war

Saison 2008

Mögliche Trainingssets für Testset Saison 2008: Saisons 2005-2007 (3 Saisons), Saisons 2004-2007 (4 Saisons), Saisons 2003-2007 (5 Saisons), Saisons 2002-2007 (6 Saisons)

In der letzten betrachteten Saison 2008 stehen nun für die Datensätze die Trainingssetlängen von 3, 4, 5 und 6 Saisons zur Verfügung. Somit bestand das Testfeld bei der Vorhersage für die Saison 2005 aus der Maximalzahl von 252 Teilnehmern.

Nun zu den Ergebnissen für das Jahr 2008.

Top10

Platz	Name des Datensatzes oder Tippstrategie	Erzielte Punkte (max 306)
1.	Top40_6S_JRip	159
-.	InfoGain40_5S_JRip	159
-.	InfoGain85_4S_JRip	159
4.	Top40hg_3S_JRip	157
5.	Kicker_4S_JRip	156
6.	FStudio_4S_JRip	155
-.	Top40hg_6S_JRip	155
-.	FStudio_6S_JRip	155
-.	InfoGain40_6S_JRip	155
10.	Top85_4S_JRip	154
-.	Komplett_6S_JRip	154

Tabelle 5.20: Top10 im Benchmark1-Saison 2008

In dieser Saison findet ein Novum statt, alle Datensätze aus der Top10 stammen von einem Klassifizierer, nämlich von JRip. Die Sieger in dieser Saison sind „Top40_6S_Jrip“,

„InfoGain40_5S_JRip“ sowie „InfoGain85_4S_JRip“. Die erste nicht-JRip-Ausprägung befindet sich auf Platz 14 (Naive Bayes). Eine Tippstrategie ist wieder nicht unter den Top10 zu finden. Generell erreicht die beste Vorhersage für die Saison 2008 ebenfalls, wie bereits in der Vorsaison, etwa 52,0%.

Top3 Tippstrategien

Platz	Name des Datensatzes oder Tippstrategie	Erzielte Punkte (max 306)
31.	TSHeimsieg	147
39.	TSTeams	146
43.	TSTabellenplatz, TSFormLS	145

Tabelle 5.21: Top3 der Tippstrategien im Benchmark1-Saison 2008

Die einfachste Tippstrategie ist in dieser Saison auch die beste. TSTeams ist diesmal nur die zweitbeste Strategie, dicht gefolgt von TSTabellenplatz und TSFormLS.

Die schlechtesten Datensätze oder Tippstrategien

Platz	Name des Datensatzes oder Tippstrategie	Erzielte Punkte (max 306)
241.	Top40_3S_J48	112
-.	Komplett_3S_J48	112
-.	InfoGain85_4S_J48	112
-.	Fold40_6S_J48	112
-.	Fold85_4S_J48	112
246.	Differenz_4S_J48	111
-.	Fold_3S_J48	111
248	Random	110
249.	Differenz_3S_J48	109
250.	Top40_6S_J48	107
-.	Komplett_6S_J48	107
252.	Fold40_4S_J48	103

Tabelle 5.22: Schlechtesten 10 im Benchmark1-Saison 2008

Man findet hier wieder nur J48-Ausprägungen und die Tippstrategie Random vor. Überraschend ist lediglich, dass die Zufallsstrategie diesmal nicht den letzten Platz belegt und einen überdurchschnittlichen Wert von 35,9% Genauigkeit vorweist.

Durchschnittsplatzierung der Klassifizierer im Vergleich zu Tippstrategien

Platz	Name des Klassifizierers oder Tippstrategie	Erzielte Punkte (max 306)
1.	JRip	148,00
2.	Naive Bayes	141,85
3.	Tippstrategien ohne Random	139,55
4.	SVM	138,06
5.	Tippstrategien mit Random	137,08
6.	Random Forest	131,27
7.	J48	123,04

Tabelle 5.23: Klassifizierer vs Tippstrategien-Saison 2008

Hier ist das Ergebnis ziemlich klar, JRip führt mit fast 6 Punkten Vorsprung vor Naive Bayes. Auffällig ist, dass hier zum ersten Mal die Tippstrategien ohne Random mit SVM einen der Top3-Klassifizierer, überholen. Generell sind auch hier die Tippstrategien vor J48 und Random Forest zu finden.

Insgesamt haben die Datensätze einen Durchschnittspunktwert von 136,45. Diesmal weisen sie jedoch aufgrund von der enormen Menge der Datensätze und den schwachen Random Forest und J48-Ergebnissen einen um ca 3 Punkte schlechteren Wert als Tippstrategien ohne Random und einen um 0,5 Punkte schlechteren Wert als Tippstrategien mit Random auf.

Durchschnittliche Punktanzahl von Datensätzen im Vergleich zu Tippstrategien

Platz	Name des Datensatzes oder Tippstrategie	Erzielte Punkte (max 306)
1.	FStudio	141,55
2.	Tippstrategien ohne Random	139,55
3.	InfoGain40	137,85
4.	BestOf	137,70
5.	Top40hg	137,55
6.	Kicker	137,30
7.	Tippstrategien mit Random	135,20
8.	Differenz	136,55
9.	Fold85	136,25
10.	Top40	135,50
11.	InfoGain85	135,40
12.	Top85	135,05
13.	Komplett	133,95
14.	Fold40	132,70

Tabelle 5.24: Datensätze vs Tippstrategien-Saison 2008

In der vierten betrachteten Saison gewinnt der kleinste Datensatz „FStudio“ vor Tippstrategien ohne Random. Datensatz „InfoGain40“ belegt hier einen guten dritten Platz. Bemerkenswert ist, dass „Top40hg“, wieder vor dem „Top40“-Datensatz liegt (und dies mit ca 2 Punkten Vorsprung).

Vergleich der unterschiedlichen Trainingssetlängen

Platz	Trainingssetlängen	Erzielte Punkte (max 306)
1.	6 Saisons	137,98
2.	5 Saisons	137,55
3.	4 Saisons	136,42
4.	3 Saisons	133,83

Tabelle 5.25: Analyse der Trainingssetlänge-Saison 2008

Hier sind die Trainingssets absteigend nach Größe sortiert, was einen Unterschied zum Ranking der vorherigen Saisons ausmacht.

Ergebnisse Ranking mit Abweichungsmöglichkeiten (Benchmark2)

Platz	Platz Bench1	Name des Datensatzes oder Tippstrategie	Erzielte Punkte (max 612)
1.	31.	FStudio_3S_Naive Bayes	397
-.	12.	FStudio_5S_Naive Bayes	397
-.	28.	Top40hg_6S_J48	397
4.	2.	Top40hg_3SJrip	395
-.	1.	Top40_6S_JRip	395
6.	52.	FStudio_4S_Naive Bayes	394
7.	25.	Kicker_5S_J48	393
8.	5.	FStudio_4S_JRip	392
-.	1.	InfoGain40_5S_JRip	392
-.	1.	InfoGain85_4S_JRip	392

Tabelle 5.26: Ergebnisse Benchmark2-Saison 2008

Auch im letzten Datenset ist vieles durcheinander gewürfelt. Hier gewinnen 2 FStudio-Naive Bayes Datensätze und ein Top40hg_J48-Datensatz. Nur 4 Datensätze aus der Top10 vom ersten Benchmark sind hier bei Benchmark2 dabei. Der Erstplatzierte des 1.Benchmarks ist auf Platz 4 zusammen mit dem Zweitplatzierten zu finden.

Zusammenfassung Saison 2008

- In der Top10-Tabelle schneiden JRip-Datensätze deutlich am Besten ab. Es taucht kein anderer Klassifizierer oder Tippstrategie in der Top10 auf. Die besten Datensätze sind „Top40_6S_JRip“, „InfoGain40_5S_JRip“ und „InfoGain85_4S_JRip“
- Überraschend ist „TSHeimspiel“ die beste Tippstrategie (Gesamtplatz 31), erst danach kommt die bei bisherigen Saisons beste Strategie „TSTeam“ (39.)
- In der schlechtesten 10 sind wieder ausschließlich Tippstrategien und Datensätze mit J48 zu finden. Dabei belegt der Zufallsklassifizierer nicht den letzten, sondern den mit einem überdurchschnittlichen Ergebnis den fünftletzten Platz. Ein Datensatz „Fold40_4S_J48“ schließt das Feld ab
- Im Gesamtvergleich der Klassifizierer und Tippstrategien schneidet JRip-Klassifizierer mit Abstand am Besten ab, zweiter ist Naive Bayes und danach kommen Tippstrategien ohne Random vor SVM
- Im Gesamtvergleich der Datensätze und Tippstrategien liegt „FStudio“ vor „Tippstrategien ohne Random“, als Dritte ist die „InfoGain40“-Strategie zu finden
- Trainingssetlänge von 6 Saisons schneidet hier am Besten ab, die übrigen Trainingssetlängen sind absteigend bis 3 Saisons sortiert
- Diesmal sind die Datensätze aufgrund der größeren Anzahl und aufgrund des schwachen Abschneidens von J48 sowie Random Forest etwas schlechter als die Tippstrategien ohne Random (3,5 Punkte schwächer) sowie die Tippstrategien mit Random (etwa 0,5 Punkt schwächer).
- Im Benchmark2 gewinnen 2 Naive Bayes-FStudio-Datensätze und ein Top40hg_J48-Datensatz

Somit habe ich alle Saisons genaustens analysiert und komme nun zu der Betrachtung aus einer anderen Perspektive. Ich betrachte nun die Ergebnisse der Benchmarks aus der Perspektive der unterschiedlichen Trainingssetlängen.

5.2.3 Benchmarks: Betrachtung nach unterschieden Trainingssetlängen

Hier betrachte ich die nach Punkten gesamtbesten Trainingssets. Dabei sollte man berücksichtigen, dass die 3 Saison-Ausprägungen bei 4 Saisons vorkommen, 4 Saisons-Ausprägungen bei 3 Saisons, 5 Saisons-Ausprägungen bei 2 Ausprägungen und 6 Saison-Trainingsset nur bei einer Testset-Saison (Saison 2008) vorkommt. Daher sollten die Ergebnisse gesondert betrachtet werden.

3-Saisons-Trainingssets

Verwendete 3-Saison-Trainingssets: Saisons 2002-2004, Saisons 2003-2005, Saisons 2004-2006, Saisons 2005-2007

Top5

Platz	Name des Datensatzes oder Tippstrategie	Erzielte Punkte (max 306)
1.	Top40hg_3S_JRip	148,50
2.	Differenz_3S_SVM	147,75
-.	TSTeams	147,75
4.	InfoGain85_3S_JRip	147,50
5.	Fold40_3S_SVM	146,50

Tabelle 5.27: Top5 Benchmark1 bei Trainingssetlänge von 3 Saisons

Weitere Ergebnisse 3-Saison-Trainingssets

Beste Platzierung Naive Bayes-Datensatz: InfoGain40_3S_Naive Bayes (144,25, Platz 8)

Beste Platzierung JRip-Datensatz: Top40hg_3S_Jrip (148,50, Platz 1)

Beste Platzierung J48-Datensatz: InfoGain40_3S_J48 (141,5, Platz 22)

Beste Platzierung SVM-Datensatz: Differenz_3S_SVM (147,75, Platz 2)

Beste Platzierung Random Forest-Datensatz: BestOf_3S_Random Forest (134,5, Platz 38)

Beste Platzierung Tippstrategie: TSTeams (147,75, Platz 2)

Der beste Datensatz bei Trainingssets der Länge 3 Saisons ist ein „JRip-Top40hg“-Datensatz, gefolgt von der Tippstrategie „TSTeams“ sowie von 3 SVM-Datensätzen. Bei den besten Datensätzen ist es zu erkennen, dass auch Naive Bayes in der Top10 vorhanden ist, die anderen beiden Klassifizierer sich jedoch außerhalb der Top20 befinden.

Sehen wir uns noch die Top5 bei Benchmark2 an:

Platz	Name des Datensatzes	Erzielte Punkte (max 612)
1.	FStudio_3S_Naive Bayes	387,25
2.	BestOf_3S_Naive Bayes	384,24
3.	Top40hg_3S_JRip	382,50
4.	Komplett_3S_Naive Bayes	381,25
5.	Top40hg_3S_SVM	380,75

Tabelle 5.28: Top5 Benchmark2 bei Trainingssetlänge von 3 Saisons

Der Spitzensreiter des 1.Benchmarks ist auf Platz 3. Die Datensätze „FStudio“ und „BestOf“ von Naive Bayes führen das Feld an.

4-Saisons-Trainingssets

Verwendete 4-Saison-Trainingssets: Saisons 2002-2005, Saisons 2003-2006, Saisons 2004-2007

Top5

Platz	Name des Datensatzes oder Tippstrategie	Erzielte Punkte (max 306)
1.	InfoGain85_4S_JRip	149,67
2.	Differenz_4S_SVM	148,67
3.	TSTeams	148,33
4.	Kicker_4S_JRip	148,00
-.	Differenz_4S_JRip	148,00

Tabelle 5.29: Top5 Benchmark1 bei Trainingssetlänge von 4 Saisons

Weitere Ergebnisse 4-Saison-Trainingssets

Beste Platzierung Naive Bayes-Datensatz: Fold85_4S_Naive Bayes (144,00, Platz 15)

Beste Platzierung JRip-Datensatz: InfoGain85_4S_Jrip (149,67, Platz 1)

Beste Platzierung J48-Datensatz: FStudio_4S_J48 (137,00, Platz 37)

Beste Platzierung SVM-Datensatz: Differenz_4S_SVM (148,67, Platz 2)

Beste Platzierung Random Forest-Datensatz: Fold85_4S_Random Forest (138,33, Platz 33)

Beste Platzierung Tippstrategie: TSTeams (148,33, Platz 3)

Auch diesmal schneidet ein JRip-Datensatz am Besten ab und eine Tippstrategie (TSTeams) schafft es in die Top5.

Nun noch die Top5 bei Benchmark2 bei der Trainingssetlänge von 4 Saisons:

Platz	Name des Datensatzes oder Tippstrategie	Erzielte Punkte (max 612)
1.	InfoGain85_4S_JRip	382,33
2.	FStudio_4S_Naive Bayes	380,67
3.	Differenz_4S_Random Forest	380,33
4.	Komplett_4S_Naive Bayes	378,33
5.	BestOf_4S_Naive Bayes	377,67

Tabelle 5.30: Top5 Benchmark2 bei Trainingssetlänge von 4 Saisons

Auch hier gewinnt ein Naive Bayes-FStudio-Datensatz. Der beste Datensatz aus Benchmark1 ist auf Platz 4 im zweiten Benchmark.

5-Saisons-Trainingssets

Verwendete 5-Saison-Trainingssets: Saisons 2002-2006, Saisons 2003-2007

Top5

Platz	Name des Datensatzes oder Tippstrategie	Erzielte Punkte (max 306)
1.	InfoGain85_5S_JRip	154,50
2.	Fold85_5S_JRip	152,50
-.	Top40_5S_Jrip	152,50
4.	TSTeams	152,00
5.	FStudio_5S_Jrip	151,50

Tabelle 5.31: Top5 Benchmark1 bei Trainingssetlänge von 5 Saisons

Weitere Ergebnisse 5-Saison-Trainingssets

Beste Platzierung Naive Bayes-Datensatz: Kicker_5S_Naive Bayes (148,00, Platz 10)

Beste Platzierung Jrip-Datensatz: InfoGain85_5S_Jrip (153,00, Platz 1)

Beste Platzierung J48-Datensatz: FStudio_5S_J48 (144,00, Platz 27)

Beste Platzierung SVM-Datensatz: Differenz_5S_SVM (150,00 Platz 6)

Beste Platzierung Random Forest-Datensatz: Top40_5S_Random Forest (140,50, Platz 33)

Beste Platzierung Tippstrategie: TSTeams (152,00, Platz 4)

Diesmal befinden sich in der Top5 (bis auf Tippstrategien) nur Jrip-Datensätze, die Tippstrategie TSTeams schafft es hingegen diesmal auf den 4.Platz.

Jetzt noch die Top5 bei Benchmark2 bei der Trainingssetlänge 5 Saisons:

Platz	Name des Datensatzes oder Tippstrategie	Erzielte Punkte (max 612)
1.	FStudio_5S_Naive Bayes	387,50
2.	Komplett_5S_Naive Bayes	385,50
3.	InfoGain40_5S_JRip	385,00
4.	Top40_5S_Jrip	383,50
-.	BestOf_3S_Naive Bayes	383,50

Tabelle 5.32: Top5 Benchmark2 bei Trainingssetlänge von 5 Saisons

Zum dritten Mal in Folge gewinnt das kleinste FStudio-Datensatz den 2.Benchmark.

6-Saisons-Trainingssets

Verwendetes 6-Saison-Trainingsset: Saisons 2002-2007

Top5

Platz	Name des Datensatzes oder Tippstrategie	Erzielte Punkte (max 306)
1.	Top40_6S_Jrip	159
3.	Top40hg_6S_Jrip	155
-.	FStudio_6S_Jrip	155
-.	InfoGain40_6S_JRip	155
5.	Komplett_6S_JRip	154

Tabelle 5.33: Top5 Benchmark1 bei Trainingssetlänge von 6 Saisons

Weitere Ergebnisse 6-Saison-Trainingssets

Beste Platzierung Naive Bayes-Datensatz: FStudio_6S_Naive Bayes (151,00, Platz 16)

Beste Platzierung JRip-Datensatz: Top40_6S_Jrip (159,00, Platz 1)

Beste Platzierung J48-Datensatz: top40hg_6S_J48 (148,00, Platz 28)

Beste Platzierung SVM-Datensatz: Komplett_6S_SVM (145,00 Platz 43)

Beste Platzierung Random Forest-Datensatz: Top40_6S_Random Forest (142,00, Platz 78)

Beste Platzierung Tippstrategie: TSTeams (147,00, Platz 31)

Auch diesmal befinden sich in der Top10 gar nur JRip-Datensätze. Diesmal befindet sich keine Tippstrategie unter der Top5 (oder gar in der Top30).

Zum Schluss sollte noch die Top5 bei Benchmark2 bei der Trainingssetlänge von 6 Saisons analysiert werden:

Platz	Name des Datensatzes oder Tippstrategie	Erzielte Punkte (max 612)
1.	Top40hg_6S_J48	397
2.	Top40_6S_JRip	395
3.	InfoGain40_6S_JRip	388
-.	FStudio_6S_Naive Bayes	388
5.	BestOf_6S_J48	386

Tabelle 5.34: Top5 Benchmark2 bei Trainingssetlänge von 6 Saisons

Hier gewinnt Top40hg mit J48-Ausprägung. Dies ist also eine Ausprägung, die aufgrund von vielen „leichten Misses“ nach vorne geschoben wurde.

5.2.4 Gesamtüberblick und Zusammenfassung

Gesamtüberblick

Beim Gesamtüberblick gehe ich folgendermaßen vor:

- Zuerst werden die Gesamtergebnisse einzelner Datensätze analysiert und mit den Ergebnissen der Tippstrategien verglichen
- Ferner werden die Klassifizierer im Einzelnen betrachtet und dort ebenso ein Vergleich mit den Ergebnissen der Strategien angestellt
- Danach werden die Gesamtergebnisse des Benchmarks2 angesehen
- Zum Schluss werden die Ergebnisse der Tippstrategien untereinander analysiert und verglichen, bevor die Gesamtergebnisse zusammengefasst werden

Gesamtüberblick der einzelnen Datensätze und Vergleich zu Tippstrategien

Platz	Name des Datensatzes oder Tippstrategie	Erzielte Punkte (max 306)
1.	InfoGain40	139,65
2.	BestOf	138,83
3.	Top40hg	138,74
4.	FStudio	137,34
5.	Top40	137,15
6.	InfoGain85	137,08
7.	Fold85	136,50
8.	Differenz	136,20
9.	Top85	135,99
10.	Fold40	135,77
11.	Tippstrategien ohne Random	135,07
12.	Kicker	134,90
13.	Komplett	134,01
14.	Tippstrategien mit Random	132,04

Tabelle 5.35: Gesamtdurchschnittswerte von Datensätzen und Tippstrategien

Anhand dieser Tabelle lassen sich viele Aussagen und Erkenntnisse aus den gemachten Tests herauslesen. Der beste Datensatz ist InfoGain40, knapp vor dem „BestOf“ und dem „Top40hg“-Datensatz. Man erkennt, dass das Hinzufügen von „Heim“ und „Gast“-Attributen die Ergebnisse von „Top40“-Datensatz verbessert hat. Ebenso zu erkennen ist, dass der erste Datensatz ohne „Heim“ und „Gast“-Attribute sich „nur“ auf Platz 5 befindet. Generell ist zu sehen, dass mit Ausnahme des „Fold“-Datensatzes, kleinere Datensätze vor den größeren Datensätzen liegen und dass der größte Datensatz „Komplett“ den punktschlechtesten Datensatz darstellt und zusammen mit dem Kicker-Datensatz noch hinter den Tippstrategien liegt.

Bei Betrachtungen der Attributgruppen aus Kapitel 2, gewinnt die kleine FStudio-Gruppe vor Differenz und der Kicker-Gruppe. Schließlich kann man erkennen, dass fast alle Datensätze eine bessere Leistung erzielen als die Tippstrategien.

Gesamtanalyse der einzelnen Klassifizierer und Vergleich zu Tippstrategien

Platz	Name des Klassifizierers oder Tippstrategie	Erzielte Punkte (max 306)
1.	JRip	143,38
2.	SVM	142,85
3.	Naive Bayes	142,59
4.	Tippstrategien ohne Random	135,07
5.	Tippstrategien mit Random	132,04
6.	Random Forest	130,03
7.	J48	125,40

Tabelle 5.36: Gesamtdurchschnittswerte von Klassifizierern in Vergleich zu Tippstrategien

Man erkennt hier, das bereits aufgrund vieler einzelner Testergebnisse erwartetes Ergebnis, JRip ist knapp vor SVM und vor Naive Bayes. Abgeschlagen und hinter den Tippstrategien ist Random Forest sowie ganz am Schluss - J48. Bemerkenswert ist auch, dass der Abstand zwischen der Top3 und den dahinter platzierten Tippstrategien ohne Random über 7 Punkte beträgt.

Gesamtergebnisse des Benchmarks2

Platz	Platz B1	Name des Klassifizierers oder Tippstrategie	Erzielte Punkte (max 612)
1.	3.	Naive Bayes	379,77
2.	2.	SVM	375,54
3.	1.	JRip	370,10
4.	6.	Random Forest	367,08
5.	4.	Tippstrategien ohne Random	365,25
6.	7.	J48	361,98
7.	5.	Tippstrategien mit Random	361,88

Tabelle 5.37: Gesamtdurchschnittswerte von Klassifizierern in Vergleich zu Tippstrategien - Benchmark2

Man erkennt hier, dass in der Top3 Naive Bayes mit JRip die Plätze tauscht. Man kann dies mit dem Vorhersageverhalten von Naive Bayes und JRip zu erklären. Naive Bayes sagt ab und zu ein Remis voraus, was diesem zusätzliche Punkte im zweiten Benchmark einbringt. JRip findet dagegen meistens keine Regeln, die ein Remis abdecken. Ansonsten sind, bis auf eine leichte Verbesserung von Random Forest, keine großen Änderungen gegenüber dem ersten Benchmark zu bemerken.

Analyse der Tippstrategien

Platz	Name der Tippstrategie	Erzielte Punkte (max 306)
1.	TSTeams	147,75
2.	TS3LS	142,25
3.	TSLS	140,25
4.	TSHeimsieg	138,75
5.	TSTabellenplatz	138,00
6.	TSSiege	135,00
7.	TSForm5	134,00
8.	TSFormLS	132,75
9.	TSNied	129,75
10.	TSTabellenplatz2	127,25
11.	TSVersus	120,00
12.	Random	98,75

Tabelle 5.38: Gesamtdurchschnittstabelle der Tippstrategien

Man bemerkt, dass lediglich „TSTeams“ bei den besten Datensätzen mithalten kann, ohne jedoch irgendwo eine Spaltenplatzierung zu erreichen. Im Klartext heißt dies also, es war bei keinem einzigen Test eine Tippstrategie besser als alle Datensätze, es gab immer mehrere (mindestens 2, denn beste Platzierung von „TSTeams“ war der 3.Platz) Datensätze die immer besser waren. Ferner erkennt man, dass die Tabelle der letzten Saison („TSLS“) mehr aussagender war, als die Betrachtung der Tabellen der aktuellen Saison („TSTabellenplatz“, „TSTabellenplatz2“). Jedoch erkennt man auch, dass die Berechnung des Durchschnittes der letzten 3 Saisons („TS3LS“) noch mehr bringt, als nur eine Saison zu betrachten. Ebenso ist die Aussage herauszulesen, dass die Betrachtung der erzielten Siege („TSSiege“), deutlich besser ist als die Betrachtung der erzielten Niederlagen („TSNied“). Schließlich schneidet die einfache Heimsieg-(„TSHeimsieg“) und Tabellenplatz-Taktik („TSTabellenplatz“) besser ab, als beispielsweise die komplizierteren Versus- („TSVersus“) und Tabellenplatz („TSTabellenplatz2“)-Strategien.

Zusammenfassung

Nun die wichtigsten Erkenntnisse aus den Kapiteln 5.2.3 und 5.2.4 noch kurz zusammengefasst:

- Bei 3Saisons-Trainingssetlängen gewinnt der Datensatz „Top40hg_3S_JRip“ vor SVM-Datensätzen und der „TSTeams“-Tippstrategie. Bei Benchmark2 gewinnt „FStudio_3S_Naive Bayes“-Datensatz.
- Bei 4Saisons-Trainingssetlängen gewinnt mit dem Datensatz „InfoGain85_4S_JRip“ ebenfalls ein JRip-Datensatz. Im zweiten Benchmark gewinnt hier wieder „FStudio_Naive Bayes“ jedoch mit 4S Trainingssetlänge.
- 5Saisons-Trainingssetlängen werden ebenso von einem „InfoGain85-JRip“-Datensatz mit 5 Saisons-Ausprägungen angeführt. Im Benchmark2 gewinnt zum dritten Mal in Folge ein FStudio-Naive Bayes Datensatz.
- 6Saison-Trainingssetlänge wird von „Top40_6S_JRip“ gewonnen, bei Benchmark2 gewinnt „Top40hg_6S_J48“
- Bei der Betrachtung der einzelnen Datensätze über alle Saisons schneidet „InfoGain40“ am Besten ab. Zweitplatzierter wird „BestOf“ vor „Top40hg“. Beachtenswert ist zum Ersten, dass die ersten 4 Datensätze die Vereinsnamen-Attribute „Heim“ und „Gast“ beinhalten und damit „Top40hg“ auch besser abschneidet als sein Ebenbild ohne die Vereinsnamenattribute „Top40“ und zum Zweiten, dass (mit Ausnahme der „Fold“-Datensätze) kleinere Datensätze insgesamt ein besseres Ergebnis erzielen, als größere Datensätze. Schließlich erkennt man, dass Tippstrategien im Durchschnitt schlechter als die meisten Datensätze abgeschnitten haben
- Bei Gesamtanalyse der einzelnen Klassifizierer und Vergleich mit Tippstrategien schneiden die Top3-Klassifizierer mit Abstand am Besten ab. Hier wird es auch deutlich, dass die Vorhersagegenauigkeiten von „J48“ und „Random Forest“ sich eher auf einem mittelmäßigen Niveau bewegen
- Bei Analyse der Tippstrategien lautet das Ergebnis: „TSTeams“ vor „TS3LS“, „TSLS“ und vor dem einfachen „TSHeimsieg“. Generell könnte nur „TSTeams“ bei den besten Datensätzen vorne mithalten. Man erkennt ebenso, dass die Betrachtung der Tabellen von letzter Saison, etwas mehr bringt als das zu Rate ziehen der aktuellen Tabellen, ebenso wie, dass die Berechnung des Durchschnitts der Tabellenplätze der letzten 3 Saisons besser ist als eine einfache Saisonbetrachtung. Schließlich erkennt man, dass die Strategie bei der man Anzahl Siege vergleichen sollte in meinem Test etwas mehr gebracht hat, als die Strategie die Anzahl der Niederlagen zählt und vergleicht.

Somit habe ich meine Tests zu Datensätzen und Tippstrategien abgeschlossen. Es wurde erkannt, dass die Datensätze im Allgemeinen besser waren als die einzelnen Tippstrategien, es hat sich ebenso herauskristallisiert welche Datensätze zu welcher Saison passen.

5.2.5 Auswahl der Datensätze für die Saisonvorhersage

Zum Schluss zähle ich die Datensätze auf, die für die Erstellung der Saisontabellen im nächsten Kapitel verwendet werden. Die Auswahl erklärt sich von selbst. Es ist einfach der jeweils beste Datensatz für die Saison von jeden Klassifizierer im Benchmark1. Bei Punktgleichheit im Benchmark1, wird entweder zweite Benchmark zu Rate gezogen oder geguckt welcher Datensatz noch nicht getestet wurde, so dass möglichst alle Datensätze mindestens einmal zum Einsatz kommen.
(fettgedruckt sind die jeweils erstplatzierten Datensätze bei Saisontest)

Saison 2005:

Naive Bayes: **FStudio_3S_Naive Bayes (1.)**

JRip: Top40hg_3S_JRip (10.)

J48: InfoGain40_3S_J48 (13.)

SVM: Komplett_3S_SVM (5.)

Random Forest: Fold85_3S_Random Forest (34.)

Saison 2006:

Naive Bayes: Fold85_3S_Naive Bayes (29.)

JRip: Fold40_4S_JRip (7.)

J48: InfoGain40_3S_J48 (16.)

SVM: **BestOf_4S_SVM (1.)**

Random Forest: InfoGain85_4S_Random Forest (37.)

Saison 2007:

Naive Bayes: Fold85_4S_Naive Bayes (6.)

JRip: **InfoGain85_3S_JRip (1.)**

J48: Top40hg_3S_J48 (14.)

SVM: **Differenz_4S_SVM (1.)**

Random Forest: Top85_4S_Random Forest (92.)

Saison 2008:

Naive Bayes: FStudio_5S_Naive Bayes (14.)

JRip: **Top40_6S_JRip (1.)**

J48: Kicker_5S_J48 (24.)

SVM: Differenz_4S_SVM (31.)

Random Forest: Fold85_4S_Random Forest (27.)

Somit kommen alle 12 Datensätze zum Einsatz. Ebenso habe ich versucht für jeden Klassifizierer unterschiedliche Datensatzgrößen bei der Vorhersage auszuprobieren (bis auf J48, nach meinen bisherigen Erfahrungen, kommt J48 mit kleineren Datensatzgrößen deutlich besser zurecht).

5.3.Vorhersage der Saisontabelle mittels selbsterstellter Datensätze

In diesem Kapitel würde ich gerne für jede der Saison mit je 5 Klassifizierern die Tabelle der Fußball-Bundesliga vorhersagen. Dabei kann ich aus Platzgründen nicht immer die komplette Tabelle vorhersagen, die entsprechenden kompletten Vorhersagen sind im Anhang G zu finden. Ich gehe immer auf die einzelnen Bereiche der Tabelle ein und verlege Ähnlichkeitspunkte nach einem bestimmten selbst konstruierten Rankingschema. Sowohl den Begriff „Einzelne Bereiche der Tabelle“ als auch den Begriff „Rankingschema bzw. Ähnlichkeitspunkte“ möchte ich nun erst einmal erläutern.

Einzelne Bereiche der Tabelle

Die Bundesligatabelle nach einer Saison sagt aus:

- wer Meister geworden ist
- wer die jeweiligen internationalen Wettbewerbe (Champions League, Europa League) erreicht hat
- wer am Ende der Saison abgestiegen ist

Aufgrund von diesen Aussagen splitte ich die Tabelle und betrachte jeden Bereich einzeln. Ein Beispiel der Bundesligatabelle kann in der Tabelle 5.39 gesehen werden.

Platz	Vereinsname	Punkte
1.	VFL Wolfsburg	69
2.	FC Bayern München	67
3.	VFB Stuttgart	64
4.	Hertha BSC Berlin	63
5.	Hamburger SV	61
	...	
15.	Borussia Mönchengladbach	31
16.	Energie Cottbus	30
17.	Karlsruher SC	29
18.	Arminia Bielefeld	28

Tabelle 5.39: Beispieltabelle der 1.Bundesliga

Hier erkennt man auch, welche Plätze ich bei der Original- und Vorhersagetabelle in diesem Kapitel primär vergleichen möchte:

- Meister (dunkelgrün)
- Internationalen Wettbewerbe unabhängig von ihrer Art (olivgrün)
- Absteiger (rot, keine Relegation berücksichtigt, da die Relegation erst nach dem eigentlichen Beginn von meiner Diplomarbeit zum ersten Mal ausgespielt wurde)

Vergabe der Ähnlichkeitspunkte

Für jede der vorhergesagten Tabellen werde ich Ähnlichkeitspunkte vergeben. Diese bewerten wie sehr die vom Klassifizierer bzw. Datensatz vorhergesagte Tabelle der Originaltabelle ähnelt. Diese Punkte werden nach folgendem Schema verteilt:

- Meister erraten, 8 Punkte
- Absteiger mit Platzierung erraten 4 Punkte
- Absteiger erraten aber Platzierung falsch 3 Punkte
- einen Teilnehmer an internationalem Wettbewerb erraten, Platzierung richtig 3 Punkte
- einen Teilnehmer an internationalem Wettbewerb erraten aber Platzierung falsch 2 Punkte
- (Achtung, Meister zählt hier nicht zu den Teilnehmern an internationalen Wettbewerben, denn sonst gäbe es ja mehrmals Punkte beim Erraten)
- Eine Platzierung von einem Team außerhalb der oben erwähnten Plätze erraten, 1 Punkt
- Punktzahl zusätzlich zum Verein und seiner Platzierung erraten, 1 zusätzlicher Punkt
(Punktzahl erraten, Platzierung richtig aber Verein falsch, keine Punkte)

Aus dieser Tabelle wird es deutlich, wie sich meine Platzierungsprioritäten verteilen. Am wichtigsten ist es für mich den Meister zu erraten, danach kommt „die Absteiger zu erraten“. Wenn ein Klassifizierer jedoch 2 Absteiger mit Platzierung jeweils errät (oder alle 3 Absteiger ohne Platzierung) ist es für mich eine gleichzusetzende Leistung wie den Meister zu erraten. Als nächstes in der Prioritätsliste kommen die internationalen Wettbewerbe, hier kann bei Erraten von mehreren richtigen Teilnehmern bzw. deren Tabellenplätzen ebenfalls ein guter Wert

erzielt werden. Schließlich sollte es noch einen kleinen Bonus dafür geben, wenn man einen Verein außerhalb der markierten Plätze errät. Ebenso ist das Erraten der erzielten Punktzahl einen zusätzlichen Punkt wert (jedoch nicht wenn der Verein falsch ist, da die Punktezusammensetzung der Vorhersagetabelle mit an 100% grenzender Sicherheit eine andere ist als die Punkteverteilung der Originaltabelle).

Noch etwas zur Darstellungsart. In der Darstellungstabelle, werden die erratenen Daten fettgedruckt und grau untermalt. Bei nicht erratenen Vereinen, wird der reale Tabellenplatz des falsch vermuteten Vereins in Klammer gesetzt. Die Vorgehensweise ist nach Wichtigkeit sortiert, zuerst wird der Meister betrachtet, danach die Absteiger und danach die internationalen Plätze.

Beispiel (Originalbezugstabelle-Tabelle 5.39)

Angenommen ein Klassifizierer sagt folgende Tabelle voraus:

1. Bayern 69 2. Stuttgart 67 3. Hamburg 66 4. Wolfsburg 65 5.Dortmund 64 (dazwischen alles falsch)...13.Eintracht Frankfurt (richtig, Punktzahl falsch), (dazwischen wieder alles falsch) 15. Gladbach 31, 16. Cottbus 30, 17. Schalke04 17, 18. Hannover96 14

Auswertung des Beispiels:

- Für den Meister gibt es keine Punkte, da zwar die Punktzahl richtig aber der Verein falsch ist
 - Für internationale Wettbewerbe gibt es für Stuttgart und Hamburg (mit falschen Platzierungen) jeweils 2 Punkte, für Punktzahl gibt es keinen zusätzlichen Punkt, also 4 Punkte hier.
 - Für „Eintracht Frankfurt“ im Mittelfeld der Tabelle, außerhalb der markierten Plätze, gibt es 1 Punkt
 - Für 15. Gladbach gibt es 2 Punkte (Platzierung+Punktzahl)
 - Für 16. Cottbus gibt es 5 Punkte (Absteiger+Platzierung+Punktzahl)
- Somit werden als Gesamtpunktzahl 12 Punkte verbucht.

Letzte Anmerkung, sicherlich wird bei der Auswertung der Datensätze vorkommen, dass einzelne Vereine gleiche Punktzahl erzielen (und Tordifferenz können die Klassifizierer nicht vorhersagen). Ich werde dies zum Wohle des Klassifizierers/Datensatzes handhaben, das heisst wenn bei dem in der Tabelle 5.39 vorgestelltem Beispiel bei Klassifizierer als Platz 1 und 2 Wolfsburg und Bayern mit je 70 Punkten als Meister als Ergebnis erscheinen, werde ich einfach Platz 1 Wolfsburg 70 und Platz 2 Bayern 70 zählen und die Punkte für den erratenen Meister, sowie Platz 2 vergeben. Natürlich sollte ein Klassifizierer mit gleicher Anzahl an Punktzahlen jedoch nicht übertreiben, bei vielen punktgleichen Vereinen, werde ich die notwendigen Maßnahmen ergreifen, die bis zur Ausschließung des Klassifizierers aus der Betrachtung führen können.

Saison 2005

Ich komme nun zur Saison 2005. Diese Saison hat mit dem Meister „Bayern München“ einen Serienmeister, mit 2 Aufsteigern ein eher leicht vorhersagbares Absteigerfeld und auch keine Überraschungen bei „internationalen Teilnehmern“. Also eine tendenziell leicht vorherzusagende Saison.

In der Saison 2005 treten folgende Datensätze an:

Naive Bayes: **FStudio_3S_Naive Bayes (1. in dem TendenzRanking für Saison 2005)**

JRip: Top40hg_3S_JRip (10.)

J48: InfoGain40_3S_J48 (13.)

SVM: Komplett_3S_SVM (5.)

Random Forest: Fold85_3S_Random Forest (34.)

(fettgedruckt sind die Erstplatzierten in der Tendenztabelle der jeweiligen Saison)

Meister-Vorhersage

Endergebnis Saison 2005			Vorhersage: FStudio_3S_Naive Bayes			Vorhersage: Top40hg_3S_JRip		
P	Name	Pkt	P	Name	Pkt	P	Name	Pkt
1.	Bayern München	75	1.	Bayern München	102	1.	Bayern München	87
<hr/>								
Vorhersage: InfoGain40_3S_J48			Vorhersage: Komplett_3S_SVM			Vorhersage: Fold85_3S_Random Forest		
P	Name	Pkt	P	Name	Pkt	P	Name	Pkt
1.	Werder Bremen (2)	87	1.	Werder Bremen (2)	81	1.	Dortmund (7)	71

Tabelle 5.40: Meistertipp Saison2005

Der Meister der letzten Saison war Bayern München.

Wie man erkennt liegen nur 2 der 5 Klassifizierer richtig. Bei den anderen 3 liegt Bayern auf dem 2.Platz hinter Bremen. Bei 2 Klassifizierern, die den Tabellenführer erraten haben, stimmt jedoch die Punktzahl nicht. Vor allem bei Naive Bayes ist es auffällig, dass Bayern alle 34 Saisonspiele gewinnt und die Maximalpunktzahl holt, was in der Realität so gut wie gar nicht vorkommt. Ich werde in den nächsten Saisons prüfen ob dieses Phänomen am kleinen FStudio-Datensatz oder an der Vorhersageart von Naive Bayes liegt. Ebenso beachtenswert ist, dass alle 5 Klassifizierer dem Meister zu viele Punkte vergeben haben. Am nächsten von der Punktzahl ist hier noch SVM, der lediglich 6 Punkte drüber liegt.

Somit erhalten der Naive Bayes- und der JRip-Datensatz 8 Punkte für den richtigen Meister, die anderen 3 Klassifizierer gehen hier leer aus.

Absteiger-Vorhersage

Endergebnis Saison 2005			Vorhersage: FStudio_3S_Naive Bayes			Vorhersage: Top40hg_3S_JRip		
P	Name	Pkt	P	Name	Pkt	P	Name	Pkt
16.	Kaiserslautern	33	16.	MSV Duisburg	23	16.	MSV Duisburg	31
17.	Köln	30	17.	Kaiserslautern	22	17.	Köln	23
18.	Duisburg	27	18.	Eintr. Frankfurt (14)	18	18.	E. Frankfurt (14)	23
<hr/>								
Vorhersage: InfoGain40_3S_J48			Vorhersage: Komplett_3S_SVM			Vorhersage: Fold85_3S_Random Forest		
P	Name	Pkt	P	Name	Pkt	P	Name	Pkt
16.	E. Frankfurt (14)	26	16.	Nürnberg (8)	30	16.	E. Frankfurt (14)	30
17.	Nürnberg (8)	23	17.	Eintr.Frankfurt (14)	24	17.	Duisburg	27
18.	Duisburg	21	18.	Köln	20	18.	Nuernberg (8)	25

Tabelle 5.41: Absteigertipp Saison2005

Man erkennt, dass keiner von den Klassifizierern alle 3 Absteiger erraten hatte. 2 von 3, allerdings mit falscher Platzierung hatten Naive Bayes und JRip. Alle Klassifizierer hatten auch Eintracht Frankfurt als Absteiger gesehen. Dieser Verein könnte sich aber mit dem 14. Platz gerade so retten. Auch Nürnberg und Mainz wurden noch als Absteiger gehandelt, diese haben jedoch gute Leistungen in dieser Saison gebracht und kamen nicht in die Nähe der Abstiegsplätze. Die Aufsteiger dieser Saison waren 1. Duisburg, 2 Köln und 3. Frankfurt. Es ist daher nicht verwunderlich, dass alle Klassifizierer mindestens 2 (JRip sogar alle 3) Aufsteiger als Abstiegskandidaten sehen. Ebenso ist es auffällig, dass alle Klassifizierer den Absteigern zu wenige Punkte zugesprochen haben. Hier fällt wieder Naive Bayes mit FStudio besonders auf.

Punkteverteilungen für Absteigertipps lauten folgendermaßen:
 JRip - 7 Punkte (Absteiger mit Platzierung erraten+Absteiger ohne Platzierung)
 Naive Bayes – 6 Punkte (2 Absteiger ohne Platzierung)
 J48 – 4 Punkte (1 Absteiger mit Platzierung)
 SVM, Random Forest – 3 Punkte (1 Absteiger ohne Platzierung)

Internationale Wettbewerbe-Vorhersage

Endergebnis Saison 2005			Vorhersage: FStudio_3S_Naive Bayes			Vorhersage: Top40hg_3S_JRip		
P	Name	Pkt	P	Name	Pkt	P	Name	Pkt
2.	Werder Bremen	70	2.	Werder Bremen	96	2.	Hamburger SV	70
3.	Hamburger SV	68	3.	Hamburger SV	82	3.	Schalke 04	69
4.	Schalke 04	61	4.	Schalke 04	81	4.	Werd. Bremen	63
5.	Leverkusen	52	5.	VFB Stuttgart (9)	71	5.	B. Dortmund (7)	60

Vorhersage: InfoGain40_3S_J48			Vorhersage: Komplett_3S_SVM			Vorhersage: Fold85_3S_Random Forest		
P	Name	Pkt	P	Name	Pkt	P	Name	Pkt
2.	B. München (1)	81	2.	B. München (1)	78	2.	W. Bremen (1)	71
3.	Leverkusen	78	3.	Stuttgart (9)	66	3.	Hamburger SV	71
4.	Schalke 04	78	4.	Leverkusen	63	4.	Hertha (6)	68
5.	Stuttgart (9)	76	5.	M'gladbach (10)	55	5.	Leverkusen	62

Tabelle 5.42: Tipp Teilnehmer internationale Wettbewerbe Saison 2005

Hier ist deutlich zu erkennen, dass Naive Bayes und Random Forest-Datensätze in dieser Kategorie fast schon zum Orakel aufgestiegen sind. Damit hat FStudio die Plätze 1-4 in dieser Saison komplett richtig. Alle 4 Teilnehmer der int.Wettbewerbe hat jedoch kein Klassifizierer erraten (obwohl alle 4 bei unterschiedlichen Klassifizierern genannt wurden). Überraschend ist die Aussage „Mönchengladbach“ bei SVM, hier haben wohl die manuell eingetragenen Attribute wie Länderspiele deutscher Nationalspieler eine Rolle gespielt (bei Gladbach spielten Altnationalspieler Neuville, Böhme und Ziege in dieser Saison mit). Die Punkteverteilungen sind bei J48, Random Forest und vor allem Naive Bayes wieder zu hoch. Werder Bremen erzielte bei Naive Bayes 96 Punkte, dies bedeutet, dass dieser Verein bei Naive Bayes alle Spiele bis auf die beiden Spiele gegen Bayern München (die ja bekanntlich 102 Punkte hatten) gewann. Ansonsten bewegen sich die Punktzahlen bei SVM und JRip in der Nähe der realen Punktzahl ohne diese jedoch richtig zu treffen. Bei JRip wurde einmal die Punktzahl des Tabellenzweiten erraten, was ich auch kursiv hervorgehoben habe. Allerdings geschah dies leider bei falschem Verein, daher gibt es hier keine Punkte.

Punkteverteilungen für Absteigertipps lauten folgendermaßen:
 Naive Bayes, Random Forest - 9 Punkte (3 int.Teilnehmer mit Platzierung)
 JRip – 6 Punkte (3 int.Teilnehmer ohne Platzierung)
 J48 – 5 Punkte (1 int.T. mit und 1 ohne Platzierung)
 SVM – 2 Punkte (1 int.T. ohne Platzierung)

Gesamtüberblick Vorhersage zur Saison 2005

Insgesamt ergibt sich für die Saison 2005 folgendes Bild:

P.	Klassifizierer	M	Abs	Int	Rest	Pkt	Gesamt (max 60)
1.	FStudio_3S_Naive Bayes (1.)	8	6	9	3	0	26
2.	Top40hg_3S_JRip (10.)	8	7	6	0	0	21
3.	Fold85_3S_RandForest (34.)	0	3	9	1	0	13
4.	InfoGain40_3S_J48 (13.)	0	4	5	0	0	9
5.	Komplett_3S_SVM (5.)	0	3	2	1	0	6

Tabelle 5.43: Vergleich der Datensätze-Saison 2005

Legende

In Klammern hinter den Datensätzen, Plätze bei Tendenztests für die Saison 2005 im Kapitel 5.2

M = Meistertipp (max 8 Punkte)

Abs = Absteigertipp (max 12 Punkte)

Int = Tipp der Internationalen Tabellenplätze (max 12 Punkte)

Rest = Punkte aus den restlichen Tipps. Die restlichen Punkte ergeben sich durch die von den Klassifizierern erratenen Platzierungen außerhalb der detailliert betrachteten Bereiche (max 10 Punkte)

Pkt = Zusatzpunkte durch erratene Punkte zusammen mit Tabellenplatz (max 18)

Wie aus der Tabelle zu erkennen, triumphiert Naive Bayes in dieser Saison mit einer beachtlichen Punktzahl. Durch das Erraten von 3 zusätzlichen Tabellenplätzen (Gladbach, Mainz, Bielefeld) in dieser Saison schafft es der kleine FStudio-Naive Bayes Datensatz 7 von 18 Tabellenplätzen komplett richtig vorherzusagen. Lediglich die Punktzahlen waren bei Naive Bayes am weitesten von allen Klassifizierern von der Realität entfernt. Den zweiten Platz sichert sich JRip, der als einziger Klassifizierer auf einer Augenhöhe mit dem Naive Bayes-Datensatz mithalten könnte. Dritter wird Random Forest. Eine totale Katastrophe waren dagegen die Vorhersagen von SVM mit dem Komplettdatensatz. Ein Beispiel dafür neben der Punktzahl in der Gesamttabelle ist, dass es in der restlichen (nicht betrachteten) Tabelle bei SVM mit Duisburg und Kaiserslautern 2 Absteiger auf die Plätze 6-9 schafften.

Saison 2006

In dieser Saison gab es mit VFB Stuttgart einen Überraschungsmeister, nur einen Aufsteiger als Absteiger und Traditionsmünder Bayern München als „nur“ Tabellenvierten (ansonsten jedoch bei int. Wettbewerbern keine Überraschungen). Diese Saison ist tendenziell schwerer vorherzusagen als die Vorherige.

In der Saison 2006 treten folgende Datensätze an:

Naive Bayes: Fold85_3S_Naive Bayes (29. Platz im Tendenzranking für die Saison 2006)

JRip: Fold40_4S_JRip (7.)

J48: InfoGain40_3S_J48 (16.)

SVM: **BestOf_4S_SVM (1.)**

Random Forest: InfoGain85_4S_Random Forest (37.)

(fettgedruckt sind die Erstplatzierten in der Tendenztabelle)

Meister-Vorhersage

Endergebnis Saison 2006			Vorhersage: Fold85_3S_Naive Bayes			Vorhersage: Fold40_4S_JRip		
P	Name	Pkt	P	Name	Pkt	P	Name	Pkt
1.	Stuttgart	70	1.	B. München (4)	100	1.	Hamburger SV (7)	95
<hr/>								
Vorhersage: InfoGain40_3S_J48			Vorhersage: BestOf_4S_SVM			Vorhersage: InfoGain85_3S_RandForest		
P	Name	Pkt	P	Name	Pkt	P	Name	Pkt
1.	Schalke 04 (2)	91	1.	Werder Bremen (3)	85	1.	B.München (4)	79

Tabelle 5.44: Meistertipp Saison2006

Diese Saison hatte mit VFB Stuttgart eine Überraschung als Saisonsieger und prompt rechnete keiner der Klassifizierer damit. Am nächsten war noch J48 mit dem Tabellenzweiten Schalke04 als Meistervorhersage. Ebenfalls haben alle Klassifizierer wieder zuviele Punkte an den Besten vergeben. Naive Bayes zeichnet sich wieder durch eine dreistellige Punktzahl für Bayern München aus. Diesmal (in Gegensatz zur Saison 2005) gewinnt Bayern jedoch nur noch 33 von 34 Spielen und spielt eins Unentschieden. Auch JRip, J48 und SVM zeichnen sich durch zu hohe Punktzahlen aus. Lediglich bei Random Forest ist der Meister lediglich nur 9 Punkte von der Punktzahl in der Realität entfernt.

Somit erhält keiner der Klassifizierer hier Punkte für den Meistertipp.

Absteiger-Vorhersage

Endergebnis Saison 2006			Vorhersage: Fold85_3S_Naive Bayes			Vorhersage: Fold40_4S_JRip		
P	Name	Pkt	P	Name	Pkt	P	Name	Pkt
16.	Mainz	34	16.	Bochum (8)	19	16.	Mainz	33
17.	Aachen	34	17.	Cottbus (13)	07	17.	Bochum (8)	33
18.	Gladbach	26	18.	Aachen	04	18.	Aachen	30
<hr/>								
Vorhersage: InfoGain40_3S_J48			Vorhersage: BestOf_4S_SVM			Vorhersage: InfoGain85_3S_RandForest		
P	Name	Pkt	P	Name	Pkt	P	Name	Pkt
16.	Bochum (8)	26	16.	Nürnberg (6)	30	16.	Aachen	30
17.	Aachen	26	17.	Eintr.Frankfurt (14)	21	17.	Bochum (8)	28
18.	Cottbus (13)	25	18.	Cottbus (13)	14	18.	Cottbus (13)	22

Tabelle 5.45: Absteigertipp Saison2006

In dieser Saison waren die Aufsteiger Bochum, Aachen und Cottbus. Es ist daher nicht verwunderlich, dass z.B. Naive Bayes, J48 und Random Forest auf diese 3 als die Schwächsten tippen. Am Besten bei den Absteigern schneidet wieder, wie in der Saison 2005, JRip ab. JRip hat als einziger Klassifizierer 2 der 3 Aufsteiger erraten, auch von den Punktzahlen lag JRip in der Nähe. Am schlechtesten schneidet hier SVM ab. Hier wurde kein einziger Absteiger erraten. Dennoch ist meiner Meinung nach der Tipp von SVM kein schlechter gewesen. Als einziger auf keinen der 3 Aufsteiger zu tippen und stattdessen, 3 andere schwächere Teams unten zu sehen, bedarf schon an großem Mut (auch wenn es bei Nürnberg dieser Saison ein sehr großes Verfehlten ist, generell ist Nürnberg dennoch unten anzusiedeln, was auch nächste Saison bestätigte). Bei den Punktzahlen fällt auch die Starrheit von Naive Bayes auf, hier haben die Aufsteiger Cottbus und Aachen sogar Punkte im einstelligen Bereich erhalten). Überraschend ist ebenso, dass keiner der Klassifizierer den eigentlichen Tabellenletzten

Gladbach als Absteiger getippt hat (bei 3 Klassifizierern war Gladbach als 10. platziert, bei einem als 9. und bei einem als 12.).

Punkteverteilung hier ist also:

JRip – 7 Punkte (1 Absteiger mit Platzierung und 1 Absteiger ohne Platzierung erraten)

J48 – 4 Punkte (1 Absteiger mit Platzierung)

Random Forest und Naive Bayes – 3 Punkte (1 Absteiger ohne Platzierung)

SVM – 0 Punkte (nichts erraten)

Internationale Wettbewerbe-Vorhersage

Endergebnis Saison 2006			Vorhersage: Fold85_3S_Naive Bayes			Vorhersage: Fold40_4S_JRip		
P	Name	Pkt	P	Name	Pkt	P	Name	Pkt
2.	Schalke 04	68	2.	Werder Bremen	95	2.	W.Bremen	95
3.	Werder Bremen	66	3.	Schalke 04	91	3.	B. München	88
4.	Bayern München	60	4.	Leverkusen	79	4.	Schalke 04	60
5.	Leverkusen	51	5.	Hamburger SV	76	5.	Leverkusen	57

Vorhersage: InfoGain40_3S_J48			Vorhersage: BestOf_4S_SVM			Vorhersage: InfoGain85_3S_RandForest		
P	Name	Pkt	P	Name	Pkt	P	Name	Pkt
2.	Werder Bremen	83	2.	B. München	84	2.	W. Bremen	79
3.	Bayern München	71	3.	Stuttgart (9)	83	3.	Hamburg (7)	74
4.	Stuttgart (1)	68	4.	Schalke 04	76	4.	Schalke 04	72
5.	Leverkusen	62	5.	Leverkusen	71	5.	Leverkusen	58

Tabelle 5.46: Tipp Teilnehmer internationale Wettbewerbe Saison2006

Am Besten schneidet auch hier JRip ab, hier wurden sogar alle 4 Teilnehmer erraten. Leider wurden dabei die Plätze 2-4 vertauscht, sonst wäre die Vorhersage perfekt. Auch die Punktzahl des Viertplatzierten stimmt bei JRip, jedoch nicht der Viertplatzierte selbst. Die anderen 4 Klassifizierer erraten hier alle 3 von 4 Teilnehmern. Bemerkenswert ist auch, dass 4 von 5 Klassifizierern Leverkusen als Tabellenfünften erraten haben.

Die Punkteverteilungen für int. Wettbewerbe in der Saison2006 lauten somit folgendermaßen:

JRip – 9 Punkte (1 Teilnehmer mit Platzierung und 3 Teilnehmer ohne Platzierung erraten)

Random Forest, SVM und J48 – 7 Punkte (1 Teilnehmer mit Platzierung und 2 ohne Platzierung)

Naive Bayes – 6 Punkte (3 Teilnehmer ohne Platzierungen)

Gesamtüberblick Vorhersage zur Saison 2006

Insgesamt ergibt sich für die Saison 2006 folgendes Bild:

P.	Klassifizierer	M	Abs	Int	Rest	Pkt	Gesamt (max 60)
1.	Fold40_4S_JRip (7.)	0	7	9	2	0	18
2.	InfoGain85_3S_RandForest (37.)	0	3	7	3	0	13
3.	InfoGain40_3S_J48 (16.)	0	4	7	1	0	12
4.	Fold85_3S_Naive Bayes (29.)	0	3	6	0	0	9
5.	BestOf_4S_SVM (1.)	0	0	7	1	0	8

Tabelle 5.47: Vergleich der Datensätze-Saison2006

Legende

In Klammern hinter den Datensätzen, Plätze bei Tendenztests für die Saison 2005 im Kapitel 5.2

M = Meistertipp (max 8 Punkte)

Abs = Absteigertipp (max 12 Punkte)

Int = Tipp der Internationalen Tabellenplätze (max 12 Punkte)

Rest = Punkte aus den restlichen Tipps. Die restlichen Punkte ergeben sich durch die von den Klassifizierern erratenen Platzierungen außerhalb der detailliert betrachteten Bereiche (max 10 Punkte)

Pkt = Zusatzpunkte durch erratene Punkte zusammen mit Tabellenplatz (max 18)

Diese Saison gewinnt der JRip-Datensatz. Vor allem zeichnete sich dieser Datensatz dadurch aus, dass es alle Teilnehmer der internationalen Wettbewerbe erraten hat. Insgesamt sind die Ergebnisse bei allen Klassifizierern aufgrund des verfehlten Meistertipps nicht berauschend, dennoch haben die „internationalen Teilnehmer“ noch vieles gerettet. Letzter wurde wieder der SVM-Datensatz, auch der Naive Bayes-Datensatz war diese Saison eher schlecht. Überraschend ist, dass der Datensatz der bei Tendenzranking am Besten abgeschnitten hat (BestOf_4S_SVM), bei Tabellenberechnung den letzten Platz belegt hat (ohne den Meister und auch nur einen Absteiger zu erraten).

Saison 2007

In dieser Saison gibt es mit Bayern München einen Traditionsmüster. Bei Absteigern steigen 2 Aufsteiger wieder ab, bei drittem Absteiger gibt es jedoch mit dem Pokalsieger und Sechstplatzierten letzter Saison Nürnberg eine Überraschung. Ebenso eine Überraschung gibt es mit VFL Wolfsburg als Teilnehmer der int. Wettbewerbe. Insofern ist es eine mittelschwere Saison.

Folgende Datensätze treten in der Saison 2007 an:

Naive Bayes: Fold85_4S_Naive Bayes (6.Platz im Tendenzranking für die Saison 2007)

JRip: **InfoGain85_3S_JRip (1.)**

J48: Top40hg_3S_J48 (14.)

SVM: **Differenz_4S_SVM (1.)**

Random Forest: Top85_4S_Random Forest (92.)

(fettgedruckt sind die Erstplatzierten im Tendenzranking)

Meister-Vorhersage

Endergebnis Saison 2007			Vorhersage: Fold85_3S_Naive Bayes			Vorhersage: InfoGain85_3S_JRip		
P	Name	Pkt	P	Name	Pkt	P	Name	Pkt
1.	Bayern München	76	1.	Bayern München	98	1.	Bayern München	83
Vorhersage: Top40hg_3S_J48			Vorhersage: Differenz_4S_SVM			Vorhersage: Top85_4S_Random Forest		
P	Name	Pkt	P	Name	Pkt	P	Name	Pkt
1.	Bayern München	81	1.	Werder Bremen (2)	97	1.	Werder Bremen (2)	91

Tabelle 5.48: Meistertipp Saison 2007

In dieser Saison wurde bei 3/5 Klassifizierern der deutsche Meister erraten. Die beiden anderen Klassifizierer tippten auf den zweitplatzierten Werder Bremen. Die Punktzahl lag nur bei JRip und J48 relativ in der Nähe. Bei Naive Bayes ist die Punktzahl mit 98 wieder weit daneben. Beachtenswert ist, dass SVM Bremen mit Hilfe eines Datensatzes ohne Vereinsnamen (Differenz) mit 97 Punkten auf Platz eins gesetzt hat.

Somit werden für den Meistertipp folgende Punktzahlen vergeben:

Naive Bayes, JRip und J48 – 8 Punkte (richtiger Meistertipp), Random Forest und SVM – 0 Punkte

Absteiger-Vorhersage

Endergebnis Saison 2007			Vorhersage: Fold85_3S_Naive Bayes			Vorhersage: InfoGain85_3S_JRip		
P	Name	Pkt	P	Name	Pkt	P	Name	Pkt
16.	Nürnberg	31	16.	Rostock	14	16.	Bielefeld (15)	33
17.	Rostock	30	17.	Cottbus (14)	12	17.	Rostock	33
18.	Duisburg	29	18.	Duisburg	05	18.	Cottbus (14)	30

Vorhersage: Top40hg_3S_J48			Vorhersage: Differenz_4S_SVM			Vorhersage: Top85_4S_Random Forest		
P	Name	Pkt	P	Name	Pkt	P	Name	Pkt
16.	Rostock	24	16.	Nürnberg	30	16.	Duisburg	26
17.	Karlsruher SC (11)	21	17.	Wolfsburg (5)	28	17.	Rostock	25
18.	Duisburg	21	18.	Cottbus (14)	25	18.	Karlsruhe (11)	19

Tabelle 5.49: Absteigertipp Saison 2007

Die Aufsteiger dieser Saison waren: 1. Karlsruhe, 2. Rostock 3. Duisburg

Diesmal tippten J48 und Random Forest auf die alle 3 Aufsteiger und lagen gar nicht so weit daneben. Den dritten Absteiger Nürnberg hat nur SVM erraten, dies aber leider mehr durch Glück, da in der Endtabelle des Klassifizierers 5 Teams (von Platz 12-16) Punktgleich sind und da ich in Zweifel für den Klassifizierer auswähle, da der Klassifizierer keine Tordifferenzen kennt, wurde Nürnberg auf Platz 16 gesetzt.

Generell liegen die Punktzahlen von JRip, SVM und eher noch Random Forest im normalen Bereich.

Bei Naive Bayes (wie auch bisher) und J48 sind die Punktzahlen bei den Absteigern zu niedrig.

Überraschend ist auch der Tipp von SVM, der Wolfsburg als einen Abstiegskandidaten sieht.

Die Punkteverteilungen:

Naive Bayes, J48 und Random Forest – 7 Punkte (1 Absteiger mit und 1 ohne Platzierung erraten)
SVM und JRip – 4 Punkte (1 Absteiger mit Platzierung).

Internationale Wettbewerbe-Vorhersage

Endergebnis Saison 2007			Vorhersage: Fold85_3S_Naive Bayes			Vorhersage: InfoGain85_3S_JRip		
P	Name	Pkt	P	Name	Pkt	P	Name	Pkt
2.	Werder Bremen	66	2.	Werder Bremen	96	2.	Schalke 04	83
3.	Schalke 04	64	3.	Stuttgart (6)	81	3.	Hamburger SV	75
4.	Hamburger SV	54	4.	Schalke 04	80	4.	W. Bremen	74
5.	Wolfsburg	54	5.	Leverkusen (7)	80	5.	Leverkusen	70

Vorhersage: Top40hg_3S_J48			Vorhersage: Differenz_4S_SVM			Vorhersage: Top85_4S_Random Forest		
P	Name	Pkt	P	Name	Pkt	P	Name	Pkt
2.	Stuttgart (6)	81	2.	B. München (1)	93	2.	B. München (1)	86
3.	Schalke 04	81	3.	Leverkusen (7)	90	3.	Stuttgart (6)	75
4.	Werder Bremen	78	4.	Stuttgart (6)	88	4.	Leverkusen (7)	62
5.	Leverkusen (7)	75	5.	Schalke 04	72	5.	Hannover (8)	61

Tabelle 5.50: Tipp Teilnehmer internationale Wettbewerbe Saison 2007

Diesmal waren die Ergebnisse bei den Teilnehmern der „int. Wettbewerbe“ sehr durchwachsen. Lediglich JRip hat 3 von 4 Teilnehmern, allerdings allesamt mit falscher Platzierung, erraten. VFL Wolfsburg hatte keiner der Klassifizierer auf der Rechnung. Zum ersten Mal gibt es auch mit Random Forest einen Klassifizierer, der alle 4 internationale Wettbewerb Teilnehmer falsch vorhergesagt hat. Die Punkte von allen Klassifizierern sind diesmal zu hoch angesetzt.

Die Punkteverteilungen sind:

JRip – 6 Punkte (3 Teilnehmer ohne Platzierung erraten)

Naive Bayes, J48 – 5 Punkte (1 Teilnehmer mit Platzierung und 1 ohne Platzierung erraten)

SVM – 2 Punkte (1 Teilnehmer ohne Platzierung erraten)

Random Forest – 0 Punkte (nichts erraten)

Gesamtüberblick Saison 2007

P.	Klassifizierer	M	Abs	Int	Rest	Pkt	Gesamt (max 60)
1.	Fold85_3S_Naive Bayes (6.)	8	7	5	1	0	21
-.	Top40hg_3S_J48 (14.)	8	7	5	1	0	21
3.	InfoGain85_3S_JRip (1.)	8	4	6	1	0	19
4.	Differenz_4S_SVM (1.)	0	4	2	3	0	9
5.	Top85_4S_Random Forest (92.)	0	7	0	1	0	8

Tabelle 5.51: Vergleich der Datensätze-Saison2007

Legende

In Klammern hinter den Datensätzen, Plätze bei Tendenztests für die Saison 2005 im Kapitel 5.2

M = Meistertipp (max 8 Punkte)

Abs = Absteigertipp (max 12 Punkte)

Int = Tipp der Internationalen Tabellenplätze (max 12 Punkte)

Rest = Punkte aus den restlichen Tipps. Die restlichen Punkte ergeben sich durch die von den Klassifizierern erratenen Platzierungen außerhalb der detailliert betrachteten Bereiche (max 10 Punkte)

Pkt = Zusatzpunkte durch erratene Punkte zusammen mit Tabellenplatz (max 18)

Die Saison 2007 wird von gleich 2 Datensätzen gewonnen. Beide Datensätze haben den Meister und 2 Absteiger erraten und holten knapp vor JRip den Saisonsieg. Dabei ist hier erstmals ein J48-Datensatz auf Platz1. SVM-Datensatz schneidet wieder schlecht ab aber diesmal gab es noch einen mit Random Forest einen Klassifizierer der noch schlechter ist. Wieder sieht man, dass die beiden tendenzstärksten Datensätze nicht die vorderen Plätze belegen.

Saison 2008

Die letzte Saison ist ein Highlight. Meiner Meinung ist diese Saison mit der am schwersten vorherzusagenden Tabelle. Wolfsburg als Meister und Hoffenheim als Aufsteiger. Dazu die vorher eher mittelmäßige Hertha in den internationalen Wettbewerben. Zusätzlich dazu ist kein Aufsteiger in dieser Saison abgestiegen.

In dieser Saison treten folgende Datensätze an (Fettgedruckt sind die Erstplatzierten im Tendenzranking 2008):

Naive Bayes: FStudio_5S_Naive Bayes (14.Platz im Tendenzranking für die Saison 2008)

JRip: **Top40_6S_JRip (1.)**

J48: Kicker_5S_J48 (24.)

SVM: Differenz_4S_SVM (31.)

Random Forest: Fold85_4S_Random Forest (27.)

Meister-Vorhersage

Endergebnis Saison 2008			Vorhersage: FStudio_5S_Naive Bayes			Vorhersage: Top40_6S_JRip		
P	Name	Pkt	P	Name	Pkt	P	Name	Pkt
1.	Wolfsburg	69	1.	B. München (2)	93	1.	B. München (2)	78
<hr/>								
Vorhersage: Kicker_5S_J48			Vorhersage: Differenz_4S_SVM			Vorhersage: Top85_4S_Random Forest		
P	Name	Pkt	P	Name	Pkt	P	Name	Pkt
1.	Stuttgart (3)	86	1.	B. München (2)	99	1.	B. München (2)	94

Tabelle 5.52: Meistertipp Saison2008

Wie bereits erwartet, setzte keiner der Klassifizierer Wolfsburg auf Platz 1. Beste Platzierung des Meisters in der Saison 2008 war bei den Klassifizierern ein 3. Platz bei JRip. Bei den restlichen Datensätzen wurde eine Platzierung rund um den Bereich 7-12 Platz vorhergesagt. Die Punkteanzahl ist bei allen Klassifizierern zu hoch. Lediglich bei JRip liegt es noch im annehmbaren Bereich. Überraschend ist, dass SVM mit dem Differenz-Datensatz (ein Datensatz ohne die Vereinsnamen-Attribute) wieder eine sehr hohe Punktzahl für den Meister voraussagt, diesmal die höchste von allen Klassifizierern.

Für den Meistertipp werden somit keine Punkte vergeben.

Absteiger-Vorhersage

Endergebnis Saison 2008			Vorhersage: FStudio_5S_Naive Bayes			Vorhersage: Top40_6S_JRip		
P	Name	Pkt	P	Name	Pkt	P	Name	Pkt
16.	Cottbus	30	16.	Cottbus	20	16.	Cottbus	36
17.	Karlsruher SC	29	17.	Karlsruher SC	17	17.	Köln (12)	36
18.	Bielefeld	28	18.	M'Gladbach (15)	10	18.	Hoffenheim (7)	33
<hr/>								
Vorhersage: Kicker_5S_J48			Vorhersage: Differenz_4S_SVM			Vorhersage: Top85_4S_Random Forest		
P	Name	Pkt	P	Name	Pkt	P	Name	Pkt
16.	E.Frankfurt (13)	30	16.	Cottbus	24	16.	Hannover (11)	27
17.	Bielefeld	26	17.	Köln (12)	24	17.	E.Frankfurt (13)	24
18.	Dortmund (6)	23	18.	Karlsruher SC	23	18.	Köln (12)	21

Tabelle 5.53: Absteigertipp Saison2008

Die Aufsteiger diese Saison waren: 1. Gladbach 2. Hoffenheim 3. Köln

Generell ist zu erkennen, dass Naive Bayes hier am Besten abgeschnitten hat. Random Forest erkennt keinen einzigen Absteiger und liegt auch ziemlich weit daneben. Auch amüsant ist der Tipp für Dortmund als Tabellenletzten von J48-Datensatz (trotz der erratenen Punktzahl des „falschen“ Vereins). Auffällig ist, dass nur JRip-Klassifizierer mehr als 1 Aufsteiger als Absteiger diesmal vorhergesagt hat. JRip ist auch der einzige Klassifizierer der Hoffenheim zu den Absteigern zählte. Alle anderen Klassifizierer waren bei derart starken Aufsteigern vorsichtig, lediglich Köln sahen 4 von 5 Klassifizierer als Abstiegskandidat.

Die Punkteverteilung:

Naive Bayes – 8 Punkte (2 Absteiger mit Platzierung erraten)
 SVM – 7 Punkte (1 Absteiger mit Platzierung und 1 Absteiger ohne Platzierung)
 JRip – 4 Punkte (1 Absteiger mit Platzierung)
 J48 – 3 Punkte (1 Absteiger ohne Platzierung)
 Random Forest – 0 (nichts erraten)

Internationale Wettbewerbe-Vorhersage

Endergebnis Saison 2008			Vorhersage: FStudio_5S_Naive Bayes			Vorhersage: Top40_6S_JRip		
P	Name	Pkt	P	Name	Pkt	P	Name	Pkt
2.	Bayern München	67	2.	Leverkusen (9)	77	2.	Schalke 04 (8)	73
3.	Stuttgart	64	3.	Hamburger SV	76	3.	Wolfsburg (1)	63
4.	Hertha BSC	63	4.	Hoffenheim (7)	73	4.	Hamburger SV	62
5.	Hamburger SV	61	5.	Hertha BSC	71	5.	Leverkusen (9)	59

Vorhersage: Kicker_5S_J48			Vorhersage: Differenz_4S_SVM			Vorhersage: Top85_4S_Random Forest		
P	Name	Pkt	P	Name	Pkt	P	Name	Pkt
2.	Bayern München	72	2.	Schalke 04 (8)	90	2.	Schalke (8)	83
3.	Leverkusen (7)	69	3.	W.Bremen (10)	89	3.	W.Bremen (10)	82
4.	W. Bremen (10)	66	4.	Leverkusen (7)	82	4.	Stuttgart	75
5.	Schalke 04 (8)	66	5.	Hamburger SV	72	5.	Leverkusen (9)	75

Tabelle 5.54: Tipp Teilnehmer internationale Wettbewerbe Saison2008

Auch diesmal sind die Ergebnisse bei den Vorhersagen der Teilnehmer von internationalen Wettbewerben eher schwach. Naive Bayes errät 2 von 4 Teilnehmern ohne Platzierung und tippt sogar noch mutig auf Hoffenheim auf den int. Plätzen. SVM und J48 erraten 1 Teilnehmer mit und JRip sowie Random Forest ohne Platzierung. Die Punkte von Random Forest und SVM sind zu hoch, bei JRip und bei J48 in akzeptabler Höhe angesetzt.

Die Punkteverteilungen sind:

Naive Bayes – 4 Punkte (2 Teilnehmer ohne Platzierung erraten)
 SVM, J48 – 3 Punkte (1 Teilnehmer mit Platzierung erraten)
 JRip und Random Forest – 2 Punkte (1 Teilnehmer ohne Platzierung erraten)

Gesamtüberblick Saison 2008

P.	Klassifizierer	M	Abs	Int	Rest	Pkt	Gesamt (max 60)
1.	FStudio_5S_Naive Bayes (14.)	0	8	4	1	0	13
2.	Differenz_4S_SVM (31.)	0	7	3	1	0	11
3.	Kicker_5S_J48 (24.)	0	3	3	0	0	6
-.	Top40_6S_JRip (1.)	0	4	2	0	0	6
5.	Top85_4S_Random Forest (27.)	0	0	2	0	0	2

Tabelle 5.55: Vergleich der Datensätze-Saison2008

Legende

In Klammern hinter den Datensätzen, Plätze bei Tendenztests für die Saison 2005 im Kapitel 5.2

M = Meistertipp (max 8 Punkte)

Abs = Absteigertipp (max 12 Punkte)

Int = Tipp der Internationalen Tabellenplätze (max 12 Punkte)

Rest = Punkte aus den restlichen Tipps. Die restlichen Punkte ergeben sich durch die von den Klassifizierern erratenen Platzierungen außerhalb der detailliert betrachteten Bereiche (max 10 Pkt.)
Pkt = Zusatzpunkte durch erratene Punkte zusammen mit Tabellenplatz (max 18)

Wie vermutet liegen bei dieser „schwereren“ Saison die Punktzahlen deutlich niedriger als bei vorherigen Saisons. Vorne ändert sich wenig, auch diesmal gewinnt Naive Bayes. Auf Platz 2 ist diesmal aber SVM zu finden, die beste Platzierung des Klassifizierers (auch wenn die Punktzahl mit 11 eher niedrig ist). Ganz hinten ist wieder Random Forest mit lediglich 2 Punkten zu finden.

Gesamtüberblick

Nun betrachte ich die Klassifizierer in einem Gesamtüberblick (Angaben sind in Punkten, in Klammern sind die Platzierungen bei den jeweiligen Saisons bzw. bei Gesamt die Durchschnittsplatzierung):

P.	Klassifizierer	2005	2006	2007	2008	Gesamt (max 240)
1.	Naive Bayes	26 (1)	9 (4)	21 (1)	13 (1)	69 (1,75)
2.	JRip	21 (2)	18 (1)	19 (3)	6 (3)	64 (2,25)
3.	J48	9 (4)	12 (3)	21 (1)	6 (3)	48 (2,75)
4.	Random Forest	13 (3)	13 (2)	8 (5)	2 (5)	36 (3,75)
5.	SVM	6 (5)	8 (5)	9 (4)	11 (2)	34 (4,00)

Tabelle 5.56: Gesamttabelle des Tabellenvorhersagerankings

Sowohl nach Punkteverteilungen als auch nach Platzierungen ist das Bild eindeutig. Naive Bayes baut mit Abstand die bestaussagenden Tabellen. In 3 von 4 Saisons war die Tabelle von Naive Bayes die von allen 5 Klassifizierern die am meisten Aussagendste. Das Problem für Naive Bayes sind lediglich die Punkteverteilungen, bei diesen liegt der Klassifizierer bei Meister immer relativ weit drüber, sowie bei Absteigern relativ weit drunter. Danach folgt JRip, der in 1 von 4 Saisons vorne lag und von Punktzahlen sogar realistischere Punkteverteilungen als Naive Bayes zu bieten hat. Es folgen J48 und Random Forest. Letzter wird SVM, das in fast allen Saisons hinten im Ranking zu finden war.

5.4. Vorhersage der aktuellen Saison anhand von großen Trainingssets

In diesem Kapitel möchte ich Trainingssets von den Daten, seit Beginn der 1.Bundesliga in 1963 erstellen und prüfen, wie genau die neueren Saisons damit vorhergesagt werden können. Hierzu werde ich 2 Datensätze verwenden.

1. Der größtmögliche Datensatz, dieser Datensatz wird als „Komplett1963“ bezeichnet
2. Ein deutlich kleinerer Datensatz, eine Art „BestOf“, ich nenne es „BestOf1963“-Datensatz

Am Ende des Kapitels werde ich (jedoch in sehr knapper nichttabellarischer Form) mit den Daten eine Vorhersage der Tabelle von einer Saison tätigen und die Trends knapp diskutieren. Dabei wird für jeden Klassifizierer der Datensatz aus den beiden oberen ausgewählt, mit dem der Klassifizierer besser abgeschnitten hat. Die kompletten Ergebnisse aus den Vorhersagen mit großen Trainingssets finden Sie im Anhang H.

Konstruktion des größtmöglichen Datensatzes

Nicht alle Daten von meiner Attributsammlung liegen seit der Saison 1963 vor. Bevor ich also mit der Analyse beginne, sollte ich die manuell eingegebenen Daten, die lediglich seit der Saison 2002 vorliegen und darauf bezogenen Differenzen löschen. Somit bleiben 110 Attribute für diesen Datensatz übrig.

Konstruktion des BestOf1963-Datensatzes

Hier wende ich meine Erkenntnisse aus dem Kapitel 4.6 an:

1. Ich verwende statt aller Gruppenattribute nur die Besten aus den Tests im Kapitel 4.
2. Der „Differenzen vs andere Ausprägungen“-Schritt wird hier nicht so umfassend wie im 4.Kapitel ausfallen, da viele Differenzen für die Daten von Saison 1963-2007 gar nicht vorliegen.
3. Eliminieren von schwachen Attributen und Attributsausprägungen wird wie im Kapitel 4 stattfinden, bis auf einen Unterschied, dass die Schranke für die Eliminierung etwas weiter nach unten gesenkt wird. Dies geschieht mit der Absicht einem sehr kleinen Datensatz eventuell noch paar Attribute mehr hinzuzufügen. (von Platz 100 auf Platz 110, was aber lediglich in einem zusätzlichem Attribut (Tabvergleich) resultiert).

Somit liegen für BestOf1963-Datensatz folgende Attribute vor:

1. Heim (96, FA), 2. Gast (72, FA), 3. TabDiff (65, Diff), 4. Tabtendenz (93, Komb), 5. SiegDiff (63, Diff), 6. NiedDiff (90, Diff), 7. PunkteDiff (50, Diff), 8. TDDiff (20, Diff), 9. TFHeim (53, Komb), 10. Heimstark (77, Komb), 11. Versus1 (83, Grup), 12. DiffForm3 (1, Grup), 13. DiffAvg1 (2, Grup), 14. Abstieg1Gast (16, Grup), 15. DiffHaeufPlaz (36, Diff), 16. Tabvergleich (108, Komb), 17. Tendenz (Klassenattribut)

Kommen wir nun zu den Ergebnissen unserer Tests (in Klammern die Angabe der Gesamtzahl erratener Tendenzen):

1. Komplett1963 (alle Werteangaben in Prozent)

Klassifizierer Trset.—Testset	Naive Bayes	JRip	J48	SVM	Random Forest
1963/2001---2002	46,08 (141)	50,00 (153)	39,54 (121)	47,06 (144)	44,12 (135)
1963/2002---2003	50,98 (156)	54,25 (166)	46,08 (141)	52,94 (162)	50,65 (155)
1963/2003---2004	48,37 (148)	48,37 (148)	47,71 (146)	50,65 (155)	45,10 (138)
1963/2004---2005	47,71 (146)	45,42 (139)	43,79 (134)	44,44 (136)	41,83 (128)
1963/2005---2006	43,79 (134)	43,79 (134)	42,48 (130)	43,46 (133)	39,54 (121)
1963/2006---2007	48,03 (147)	50,00 (153)	42,81 (131)	48,69 (149)	44,12 (135)
1963/2007---2008	45,75 (140)	49,34 (151)	49,02 (150)	49,35 (151)	46,73 (143)
Gesamtdurchschn.	47,24	48,74	44,49	48,12	44,58

Tabelle 5.57: Ergebnisse des Komplett1963-Datensatzes

Wie man deutlich erkennt sind die hier Werte fast überall besser als bei entsprechendem Komplett-Datensatz (hier vergleiche ich die Werte ab Saison2005). Besonders gut sind die Werte bei JRip, besonders der Wert von über 54% bei Saison 2003 ist beeindruckend. Generell scheint die Saison 2003 deutlich vorhersagbarer als die anderen Saisons zu sein, dies wird sowohl bei Naive Bayes, SVM, als auch bei JRip und auch Random Forest deutlich. Bei „Komplett1963“ erzielt JRip das beste Ergebnis unter allen Klassifizierern.

2. BestOf1963 (alle Werteangaben in Prozent, die Pfeile bedeuten Verbesserung bzw. Verschlechterung gegenüber dem „Komplett1963“-Datensatz)

Klassifizierer Trset.—Testset	Naive Bayes	JRip	J48	SVM	Random Forest
1963/2001---2002	43,14 (132) ↓	48,37 (148) ↓	45,10 (138) ↑	50,65 (155) ↑	40,52 (124) ↓
1963/2002---2003	50,61 (155) ↓	54,25 (166) ⚡	51,96 (159) ↑	52,94 (162) ⚡	49,67 (152) ↓
1963/2003---2004	49,34 (151) ↑	49,02 (150) ↑	50,00 (153) ↑	50,65 (155) ⚡	41,18 (126) ↓
1963/2004---2005	45,42 (139) ↓	43,79 (134) ↓	43,14 (132) ↓	44,77 (137) ↑	43,14 (132) ↑
1963/2005---2006	44,77 (137) ↑	43,46 (133) ↓	43,79 (134) ↑	43,46 (133) ⚡	43,46 (133) ↑
1963/2006---2007	49,35 (151) ↑	47,71 (146) ↓	48,69 (149) ↑	49,35 (151) ↑	41,50 (127) ↓
1963/2007---2008	50,00 (153) ↑	48,04 (147) ↓	51,63 (158) ↑	49,35 (151) ⚡	39,87 (122) ↓
Gesamtdurchschn.	47,52 ↑	47,81 ↓	47,76 ↑	48,74 ↑	42,76 ↓

Tabelle 5.58: Ergebnisse des BestOf1963-Datensatzes

Bei einem kleineren Datensatz wird J48 besser, dieser Trend wurde von mir fast durchwegs über die gesamte Diplomarbeit bemerkt. Auch die Verschlechterung von JRip ist zu bemerken. Während aber die Verschlechterung von JRip sich eher in einem marginalen Bereich bewegt, ist die Verbesserung von J48 mehr als deutlich. Solche Spitzenwerte für J48 wie bei diesem Datensatz wurden sonst nirgends erreicht. SVM verbessert sich etwas und erzielt in „BestOf1963“-Datensatz das beste Ergebnis unter allen Klassifizierern. Random Forest fällt hingegen gegenüber dem größerem Datensatz stark ab, wobei bei diesem Klassifizierer von beiden Datensätzen keine besonders guten Ergebnisse erzielt wurden.

Zum Schluss würde ich noch gerne ziemlich knapp und kurz eine Saison mit den großen Trainingssets vorhersagen. Es ist interessant wie sich die Klassifizierer bei solch einem großen Trainingsset verhalten. Zum Testen nehme ich die am schwersten vorherzusagende Saison 2008.

Da „BestOf1963“ eine Verbesserung bei Naive Bayes, J48 und SVM brachte, wird es von diesen Attributen für die Vorhersage verwendet. Die anderen beiden Klassifizierer JRip und Random Forest verwenden „Komplett1963“ als Datensatz.

Vorhersage der Saison 2008 mit Saison 1963-2007 als Trainingsset

Aus Platzgründen werde ich hier, wie bereits oben erwähnt, auf die Klassifizierer einzeln eingehen und keine tabellarische Übersichten präsentieren.

Naive Bayes

Naive Bayes sagt wieder strikt für die Bayern voraus. D.h. der Serienmeister erhält die maximale Punktzahl von 102 Punkten mit einem Sieg in jedem Spiel. Als Absteiger wurde 1 Aufsteiger Mönchengladbach ausgewählt sowie 2 Vereine die zuletzt relativ kurz in der Liga waren also Karlsruhe und Cottbus. Dabei waren die Punkteverteilungen Naive Bayes-typisch sehr gering (Gladbach 15, Karlsruhe 14 und Cottbus 4 Punkte). Im internationalen Wettbewerb stehen mit HSV, Schalke, Stuttgart und Leverkusen 4 Traditionsvereine, die zuletzt häufiger auf diesen Plätzen zu finden waren. Überraschenderweise belegt mit Bremen ein weiterer Traditionsverein nur den 7.Tabellenplatz. Ebenso überraschend ist, dass der Liganeuling Hoffenheim trotz fehlender Tradition, dafür aber scheinbar mit einem vielsagenden Kader von Naive Bayes Datensatz auf Platz 10 geschätzt wurde. Ansonsten sind bis auf die bei Naive Bayes typisch „starren“ Punkteverteilungen keine Probleme aufgefallen.

JRip

Auch hier gewinnt Bayern München die Meisterschaft. Nach Bayern sind jedoch punktgleiche Grüppchen zu erkennen. Plätze 2,3 und 4 sind diese mit 3 Traditionsvereinen Bremen, Schalke und Stuttgart gut besetzt aber danach... Platz 5-9, Platz 10-12, Platz 13-16 und Platz 17-18, all dies sind punktgleiche Grüppchen und seit der Gruppe um Platz 5-9 trennen jede Gruppe genau 3 Punkte Abstand. Dies sind die Anzeichen dafür, dass JRip möglicherweise nicht sehr gut mit einem enorm großem Bundesligatrainingsset mit vielen zum Teil kontroversen Daten zureckkommt und für die Vorhersage eine sehr einfache Regel auswählt.

J48

J48 fiel dadurch auf, dass hier „Schalke 04“ und nicht „Bayern München“ Meister wurde. Ferner sind die Punktverteilungen sehr realistisch. Auch bei den Platzierungen gibt es bis auf den Meister nichts Ungewöhnliches. Bei den Absteigern wurde Cottbus als einziger Absteiger mitsamt der Platzierung erraten. Bei den internationalen Wettbewerbsteilnehmern wurden Bayern und HSV erraten.

SVM

SVM lieferte das seltsamste Ergebnis von allen Klassifizierern. Dieses lautet: 1. Bayern 102 Punkte (Der Rekordmeister gewinnt somit alle Spiele). Die anderen Vereine auf Plätzen 2-18 haben allesamt 48 Punkte (jeder gewinnt nur seine Heimspiele, bis eben auf das Spiel gegen Bayern. D.h. 16 Heimspiele (ohne Bayernspiel) = 3 Punkte = 48 Punkte).

Scheinbar war diese Strategie aus der Sicht von SVM die Optimalste um die Saison vorherzusagen, was darauf schließen lässt, dass zu viele zum Teil kontroverse Daten SVM eher verwirrt als geholfen haben. Ein Versuch den Fehler mit einem „Komplett1963“-Datensatz (statt den verwendeten „BestOf1963“-Datensatz) zu beheben scheiterte kläglich. Mit dem „Komplett1963“ tätigt SVM die Vorhersage, dass alle Vereine in der Saison 2008 51 Punkte erzielen (d.h. jeder gewinnt seine Heimspiele, ohne Beachtung von Bayern).

Random Forest

Auch hier ist das Ergebnis sehr realistisch, zu dem wurden alle 3 Aufsteiger (wenn auch in falscher Reihenfolge) erraten. Bayern wurde fälschlicherweise auch hier als Meister getippt. Bei den internationalen Wettbewerbsteilnehmern wurde lediglich Stuttgart erraten.

Insgesamt waren hier somit 3 Aussagen (Naive Bayes, J48 und Random Forest) realistisch und als eine gute Vorhersage zu bewerten. JRip und Random Forest lieferten in diesem Fall eher unverwertbare und fragwürdige Ergebnisse.

Somit schließe ich dieses Thema ab und beginne im nächsten Unterkapitel mit der abschließenden qualitativen Diskussion.

5.5.Qualitative Diskussion

In diesem Unterkapitel diskutiere ich über die Ergebnisse meiner Diplomarbeit. Dabei möchte ich die Diskussion in 3 Ebenen aufteilen. Die erste Ebene, die zugleich die Wichtigste ist sind die Klassifizierer. Hier gehe ich auf jeden einzelnen Klassifizierer ein, diskutiere die Ergebnisse, Vorhersageverhalten sowie –authentizität, das Abschneiden bei dem Test gegen Tippstrategien, bei Tabellenbildung, sowie die Reaktion auf einen sehr großen Trainingsset. Die zweite Ebene bilden die Attribute selber, hier versuche ich anhand der Ergebnisse aus dem 4.Kapitel eine Aussage zu treffen, welche Attribute nun besser und welche schlechter sind. Die dritte Ebene bilden schließlich die am Ende des 4.Kapitels erstellten Datensätze und Tippstrategien. Hier möchte ich ebenso auf deren Abschneiden im 5.Kapitel eingehen und diskutieren.

5.5.1 Diskussion über Klassifizierer

In diesem Unterkapitel diskutiere ich über die in der gesamten Diplomarbeit verwendeten Klassifizierer: Naive Bayes, JRip, J48, SVM und Random Forest. Hierzu erwähne ich neben der Zusammenfassung 2 zusätzliche Versuche bzw. Tests, die ich für die Klassifizierer angewendet habe.

Im ersten Versuch habe ich über einige Spieltage einer Bundesligasaison 2 verschiedene Datensätze für die Vorhersage benutzt, einmal den Komplettdatensatz und einmal den modifizierten Datensatz mit um ca 1/3 weniger Attributen. In beiden Fällen wurde verglichen wie die beiden Tipps eines Klassifizierers sich voneinander unterscheiden.

Im zweiten nachfolgenden Versuch wurden die oben im ersten Versuch erwähnten Spieltage mit einem oder beiden Datensets mit jedem Klassifizierer vorhergesagt und die Tipps auf die Realitätsnähe geprüft.

Naive Bayes

„Naive Bayes“ ist ein statistischer Klassifizierer, der bisher als schnell und „trotz“ seiner Schnelligkeit auch als genau bekannt war (näheres dazu im Kapitel 3.1.2).

Ergebnisse der Attributanalyse im Kapitel 4

Auf dem Komplettdatensatz vor der großen Analyse erzielte Naive Bayes den zweitbesten Wert, der noch über dem Durchschnitt von ZeroR (einfacher „Tendenz=1“ Tipp) liegt, was bedeutet, dass Naive Bayes bereits vor der Attributanalyse annehmbare Vorhersagen treffen konnte.

In Gegensatz zu vielen anderen Klassifizierern erzielten bei Naive Bayes alle 3 Wrapper (vor allem der WrapperBFBE) annehmbare Ergebnisse. Zwar waren die Folds bei WrapperLFS sehr gering, insgesamt entsprach aber das Ergebnis auch hier tendenziell dem Gesamtergebnis. Die Aussagen waren deutlich, die Folds im Durchschnitt mittelhoch. Ebenso positiv war die Arbeitsgeschwindigkeit von Weka bei diesem Klassifizierer. Eine BestFirst Backward Elimination war in weniger als einem Tag zu schaffen, während viele andere Klassifizierer mehrere Tage dafür brauchten (und dies teilweise ohne adäquate Ergebnisse herauszubekommen, siehe WrapperBFBE-Ergebnisse bei JRip und Random Forest)

Noch ein paar Worte zu den Ergebnissen von Naive Bayes: Differenzen und vor allem Differenzen von Gruppenattributen schnitten bei Naive Bayes wie auch bei anderen Klassifizierern gut ab. Ebenso war eine Formausprägung in der Top10 zu finden. Auch bei den schwachen Attributen traf Naive Bayes klare und nachvollziehbare Aussagen. Unter den schwächsten Attributen, waren zeitlichen Attribute, kombinierte Attribute und selbstgemachte Attribute, Remisattribute sowie die eher wenig

aussagenden Attribute über österreichische Spieler sowie Nationalspieler anzutreffen. Lediglich das Attribut „TFHeim“, das aussagt ob die Mannschaft zuhause ein Tabellenführer ist, schnitt bei Naive Bayes deutlich schlechter als bei allen anderen Attributen ab. Ebenso machte Naive Bayes eine Tendenz zu Gastausprägungen deutlich, die später von vielen anderen Klassifizierern unterstützt wurde.

Bei Vorher-Nachher Vergleich, wird ein Nachteil von Naive Bayes, der im weiteren Verlauf der Diplomarbeit mehrmals auffiel, deutlich: „Die Starrheit“. Damit ist gemeint, dass wenn Naive Bayes eine Aussage traf, er dann große Schwierigkeiten hatte, diese bei Notwendigkeit zu korrigieren. Diese „Starrheit“ (was aber andererseits als „Stabilität“ positiv ausgelegt werden kann) machte sich auch bei Datensatztests deutlich. Ich habe 11 verschiedenste Datensätze im Test gegen den „Komplett“-Datensatz antreten lassen und lediglich 2 Datensätze schafften Naive Bayes Ergebnisse marginal zu verbessern. Die anderen Datensätze verschlechtern Naive Bayes ebenso minimal (Bei 8 von 9 verschlechtert sich der Durchschnittswert um weniger als 1%). Lediglich der kleinste Datensatz „FStudio“ lässt Naive Bayes etwas „pendeln“. Damit meine ich, dass FStudio sowohl im positiven als auch im negativen Bereich (leider aber etwas mehr im negativen Bereich) starke Abweichungen und Schwankungen von „Komplett“-Werten verursacht.

Ergebnisse bei praktischen Tests im Kapitel 5

Beim Antreten gegen die Tippstrategien, hatte Naive Bayes bei Trainingssetgröße von 3 Saisons das beste Ergebnis. Auch bei sonstigen Ergebnissen lag Naive Bayes immer in der Top3 (meistens jedoch auf dem 3.Platz hinter JRip und SVM) und war im Durchschnitt immer besser als die Tippstrategien. Lediglich die Tippstrategie „TSTeams“ überholte die Naive Bayes Ausprägungen in einigen wenigen Tests.

Bei der Auswahl der Datensätze für die Kreation der Saisontabelle wurde zweimal der oben erwähnte „pendelnde“ Datensatz „FStudio“ und alternativ dazu zweimal mit „Fold85“ ein großer Datensatz ausgewählt. Bezuglich der Tabellenplatzierungen waren die Ergebnisse gut, nicht aber bezüglich der Punkteverteilungen. Damit meine ich, dass Naive Bayes viele Platzierungen einzelner Vereine erraten hat und auch insgesamt Naive Bayes der beste Klassifizierer im Tabellenvorhersageranking für die Saisons 2005-2008 war. Die Vergabe der Punktzahlen jedoch war immer sehr unrealistisch und unterstützte die oben erwähnte These von der „Starrheit“ von Naive Bayes. Hier erreichte Bayern München als siegreiche Traditionsmannschaft schonmal eine dreistellige Punktzahl am Ende der Saison, andererseits landeten Mannschaften mit einem unbekannten Kader sowie mit geringer Tradition schon mal im einstelligen Punktebereich. Dieser Vorhersagegedanke: „Spitzenverein gewinnt immer, Abstiegskandidat verliert immer“ unterstreicht, wie bereits erwähnt, einerseits die Starrheit andererseits jedoch die „Menschlichkeit“ von Naive Bayes. Denn genau so tippen sehr viele Teilnehmer einer Tippliga. Da ist der Hauptgedanke: „FC Bayern der gewinnt sowieso immer und beispielsweise Cottbus holt bestimmt keine Punkte“. Ich wage jetzt auch die These, dass wenn man bei vielen Tippern eine Vorhersagetabelle nach Ihren Tipps aufstellen würde, eine sehr ähnliche Tabelle zu Naive Bayes herauskäme.

Bei sehr großem Trainingsset zeigte Naive Bayes noch etwas bessere Ergebnisse bei vorherigen Tests und könnte diese Ergebnisse sogar mit einer Verkleinerung des Attributsets etwas steigern (allerdings alles im marginalen Bereich von im Durchschnitt ca 0,5%). Die realistischen Tabellenvorhersagen wurden genauso bestätigt, wie die unrealistischen Punktzahlen.

Vorhersagen von Naive Bayes

Wie bereits bei den Tabellen erwähnt, war Naive Bayes bei den Vorhersagen eines Spieltags eher „menschlich“ und „professionell“. Dies bedeutet, dass alle Aussagen nachvollziehbar aber sehr „risikoscheu“ sind. Es kommt bei Naive Bayes nur selten vor, dass auf einen Außenseiter getippt wird.

Dabei ist auch nicht primär der aktuelle Tabellenstand entscheidend sondern eher die Vergangenheit. Beispielsweise würde „Naive Bayes“ bei einem Spiel eines schwächeren Traditionvereins (z.B. Werder Bremen Saison 2008/2009) gegen einen starken Aufsteiger (z.B. Hoffenheim, Hinrunde Saison 2008/2009) den Traditionsverein bevorzugen oder zumindest im „schlimmsten Falle“ auf ein Unentschieden tippen. Ebenso ein Beispiel hierfür ist der „1-Tipp“ bei einem Heimspiel des bei Saisonstart schwächeren Rekordmeisters (2 Punkte aus 3 Spielen) gegen den eher gut gestarteten aktuellen Meister (6 Punkte aus 3 Spielen). Dabei ist Naive Bayes nicht „unentschiedenscheu“, sollten die Vereine viele Ähnlichkeiten aufzeigen scheut sich Naive Bayes nicht davor auf ein Unentschieden zu setzen. Auch hier wird jedoch die Starrheit deutlich, Naive Bayes ändert nur selten seine Vorhersagen in Abhängigkeit von aktuellen Daten oder kleineren Datensätzen (siehe Ergebnisse der Versuche).

Ergebnisse in den Versuchen

Das Resultat im ersten Versuch bei Naive Bayes waren 2 sehr ähnliche bis gleiche Vorhersagen die sich nur in sehr geringen Details unterschieden. Auch waren die meisten Änderungen nur von kosmetischer Natur, d.h. von einer „1“ (bzw „2“) zu einer „0“ und umgekehrt und keinesfalls von einer „1“ zu einer „2“.

Im zweiten Versuch schienen die Tipps von Naive Bayes am realistischsten von allen Klassifizierern zu sein. Auf jeden Fall könnte ich bei allen Tipps nachvollziehen wieso diese getätigt wurden. Viele richteten sich nach Tabelle aber auch die Heimstärke eines Vereins oder gar die Heimschwäche (z.B. Leverkusen in der Saison 2008 im Heimaushilfstadion aus Düsseldorf) wurden oft berücksichtigt.

Fazit Naive Bayes

Der Leser mag schon bestimmt gemerkt haben, dass der Naive Bayes-Klassifizierer während der gesamten Diplomarbeit mir ziemlich am Herzen lag. Diese Affinität begründe ich damit, dass ich einerseits mit der Schnelligkeit und (trotz der Schnelligkeit) guten Genauigkeit von Naive Bayes im Verlauf der gesamten Diplomarbeit mehr als zufrieden war und andererseits ich die Vorhersagen vom Naive Bayes-Klassifizierer bereits für einen KI-Teilnehmer in meiner Tippliga [**Tipp**] verwende (mit bisher einigermaßen guten Ergebnissen). Zu einem negativen Punkt muss ich aber bei Naive Bayes trotz der enormen Vorschusslorbeeren noch mal kommen, die „Starrheit“ und die „Unbeweglichkeit“ bei kurzfristigen Ereignissen stellte bei Naive Bayes ein Problem dar, das sich einerseits in „menschlich unbiogsamen“ Tipps für Favoriten resultierten und andererseits in Punkteverteilungen bei Tabellenvorhersagen deutlich machen. Dennoch würde ich die Ergebnisse von Naive Bayes, trotz dieser Malheurs auf gut schätzen.

JRip

Mit JRip wurde ein Regellerner in das Feld der Klassifizierer mit aufgenommen, dass während der gesamten Diplomarbeit mit guten, oft unter allen Klassifizierern besten Werten überzeugte. Genaue Information zu JRip finden Sie im Kapitel 3.1.3.

Ergebnisse der Attributanalyse im Kapitel 4

Beim Testen auf dem Komplettdatensatz erzielte JRip-Klassifizierer den besten Wert und hatte als einziger Klassifizierer beim Vortest Abstecher über die 50%-Marke. Damit lag JRip über dem Wert von ZeroR, was ebenfalls wie bei Naive Bayes auf bereits annehmbare Aussagen des Komplettdatensatzes hindeutet.

Bei der Hauptanalyse waren die Aussagen von 2 der 3 Wrapper sehr vielsagend. Vor allem der Ranksearch-Wrapper lieferte sehr interessante und eindeutige Aussagen. Auch bei WrapperLFS

waren die Aussagen eher eindeutig, lediglich die Folds waren wie bei Naive Bayes deutlich niedriger. Die Aussagen von BestFirst Backward Elimination Wrapper waren bei JRip leider absolut nichtsaussagend, da viel zu hoch. Fast alle Attribute wurden hier in 10/10, einige wenige in 9/10 Folds ausgewählt, was die Folds in der Gesamtwertung auch künstlich nach oben gepuscht hat aber sonst keine Auswirkungen auf das Gesamtergebnis hatte.

Beim Gesamtergebnis der Hauptanalyse gewinnen auch bei JRip die Differenzen und vor allem Differenzen von Gruppenattributen. Im Gegensatz zu Naive Bayes jedoch besteht hier die komplette Top10 nur aus Differenzen. Bei schwachen Attributen gibt es 2 sehr große Überraschungen, die Abwesenheit unter den 10 schlechtesten Attributen des schlechtesten Attributs „VieleTore“ und der „zeitlichen Attribute“. Ansonsten besteht das Feld der schwächsten Attribute bei JRip aus vielen „alten Bekannten“. Die Tendenz zur Überlegenheit der Gastausprägungen, die von Naive Bayes bereits festgestellt wurde, wurde von JRip bestätigt. Die Geschwindigkeit bei JRip war eher unterdurchschnittlich, für einen WrapperBFBE-Test brauchte ein leistungsstarker Computer mehrere Tage.

Nun zu Vorher-Nachher Vergleich. Trotz des bereits unter allen Klassifizierern besten Wertes von JRip bei dem Vorher-Vergleich haben 7 von 11 Datensätzen diesen Wert sogar noch steigern können. Auch hier waren die Verbesserungen und Verschlechterungen jedoch eher im Bereich von unter 1%. Es wurde auch bei einem Datensatz nie mehr als bei 6 von 10 Trainingssets eine Verbesserung erreicht. Am Besten schnitt hier der kleine Datensatz FStudio ab, der sich bei JRip gleich mit 3 Trainingssets über die Grenze von 50% verbesserte.

Ergebnisse bei praktischen Tests im Kapitel 5

JRip hat sich in Tests gegen die Tippstrategien als deutlich besser gezeigt und hat sich auf dem ersten Platz im gesamten Test unter den Klassifizierern positioniert. Bei Saisons 2007 und 2008 stellte JRip den jeweils besten Datensatz, wobei in der Saison 2008 sogar die gesamte Top10 nur aus JRip-Datensätzen bestand. JRip hatte somit im Verlauf der gesamten Tests mindestens einen Datensatz, der immer besser als die beste Tippstrategie war. Insgesamt war in den einzelnen Saisons JRip immer unter den Top3 der Klassifizierer und hat sich auch in der Gesamtwertung knapp als Klassifizierer Nummer 1 durchgesetzt. Bei den Tabellenvorhersagen wurde JRip hinter Naive Bayes nur zweiter, in 2 Saisons wurde JRip sogar mit einem relativ schwachen Ergebnis nur dritter (in einer Saison dagegen erster). Bei den Tabellenpunkten war es deutlich realistischer als Naive Bayes, es wurden keine dreistelligen oder einstelligen Punktzahlen vergeben, bei den Absteigern waren die Punktzahlen in Ordnung, bei Meisterschaftskandidaten etwas zu hoch (eine Saison deutlich zu hoch wo gleich 2 Mannschaften 95 Punkte erzielten). Die Bevorzugung von Bayern München war vorhanden, zeigte sich aber nicht so deutlich wie bei Naive Bayes.

Beim Testen auf großen Trainingssets waren die Ergebnisse von JRip eher mittelmäßig. Es bildeten sich Grüppchen in der Tabelle, so dass die Tabelle bis auf den ersten Platz (wo natürlich Bayern München verweilte) eher wenig Aussagekraft hatte.

Vorhersagen von JRip

Nun aber (neben den Problemen bei WrapperBFBE) zum größten Manko von JRip, den Vorhersagen. JRip ist ein Regellerner und seine Vorhersagen unterliegen logischerweise oft einer oder mehreren bestimmten Regelmengen mit entsprechenden Defaultteil und dieser Defaultteil heisst meistens bzw. in meinen Tests sogar immer „sonst 1“. Was im Allgemeinen bedeutet, dass die Tipps von JRip meist so aussehen, dass nur wenn etwas durch eine Regel von JRip abgedeckt wurde, im Vorhersagen keine „1“ steht, ansonsten jedoch in allen restlichen Spielen eine „1“ getippt wird. Eine Unentschiedenvorhersage geschieht bei JRip in ganz seltenen Fällen, wenn eine oder mehrere Regeln hierfür gefunden werden. So sehen die Spieldate mit 6-9 Heimtipps sehr künstlich aus und

entsprechen eher nicht den menschlichen Tipps. Oft tippen so Teilnehmer, die entweder sehr wenig Ahnung von Fußball oder sehr wenig Zeit haben und das Tippformular sehr schnell ausfüllen wollen. Das JRip trotzdem so gute Ergebnisse hat, liegt mit Sicherheit daran, dass erstens der Heimtipp an sich gar nicht so schlecht ist (siehe ZeroR im Kapitel 4.3) und zweitens wenn JRip schon eine Regel in das Regelset aufnimmt die von ZeroR abweicht, diese mit Sicherheit gute Ergebnisse in letzten Saisons geliefert hat und auch für diese Saison geeignet zu sein scheint (ein durchaus gutes Beispiel hierfür kann die einfache Regel sein: IF bayernmünchen=Gast dann „2“).

Dabei ist JRip auch eher auf seine Weise „starr“. Zwar kann man diesen Klassifizierer nicht mit der „Starrheit“ von Naive Bayes vergleichen aber dennoch, solange bis JRip keine „gute“ Regel gefunden hat, tippt er eben auf seine „1“. Sollte JRip eine neue Regel finden, dann ändert er entsprechend der Abdeckung dieser Regel einige Werte in „2“ oder gar „0“. Dennoch bleibt aber die Mehrheit bei JRip gleich, da aufgrund der kontroversen Ergebnisse im Trainingsset es nur sehr schwer ist für JRip eine Regel zu finden die viele Beispiele abdeckt und je größer der Trainingsset desto kontroverser und schwerer wird es für JRip. Dies könnte auch die Erklärung für die eher schwachen Ergebnisse von JRip bei sehr großem Trainingsset sein. Bei über 10000 Attributen, hat JRip einfach „Tendenz=1“ als Regel ausgewählt und eventuell noch 3 bis 4 kleine Regeln, so dass viele Vereine bei der Tabellenerstellung nah beisammen lagen, bzw. Gruppen bildeten die nur durch kleine Punktzahl getrennt waren. Vereine aus Gruppe 1 erfüllten die große Regel „Tendenz=1“ und die meisten der kleinen Regeln, die Vereine in Gruppe dahinter erfüllten eine kleine Regel weniger usw.

Ergebnisse in den Versuchen

Beim dem 1. Versuch bei JRip haben die Datensätze eher gleiche Ergebnisse gezeigt. Einige Tendenzen wurden verändert, weil auf den jeweiligen Daten eine oder mehrere zusätzliche Regeln gefunden wurden.

Beim zweiten Versuch waren wie erwartet fast alle Tipps auf „1“ gesetzt, nur in Ausnahmefällen war ein „2“ oder ein „0“-Tipp zu finden, was zwar meistens realistisch, als Tipps jedoch sehr künstlich erscheint.

Fazit JRip

JRip erzielte starke Ergebnisse und war meistens der beste Klassifizierer. Auch wurden die einfachen Tippstrategien besiegt, und Realismus bei der Punktverteilung bei den Tabellen erreicht. Leider wurden aber diese Ergebnisse auf Kosten der „Menschlichkeit und Realismus“ der Tipps gewonnen. Wenn man die Vorhersagen mit denen eines menschlichen Tipplers vergleichen würde, waren diese eher Tipps von einem sehr pragmatischen und zweckorientierten Tipper. Damit meine ich, wo JRip oder der entsprechende Tipper sicher war, stellte es seine „2“ oder sogar „0“, bei Unsicherheiten immer „1“. Es scheint also, dass es JRip gar nicht so wichtig war, alle Spiele zu erraten, sondern eben bei möglichst vielen Spielen richtig zu liegen. Genau dies ist der Unterschied zu „eher menschlichem“ Naive Bayes und der Grund wieso ich die Tipps von Naive Bayes den Tipps von JRip, trotz teilweise besseren Ergebnissen von JRip, vorziehe.

Insgesamt würde aber ich die Leistung von JRip aufgrund der guten Ergebnisse bei vielen Tests auf gut schätzen.

J48

J48-Algorithmus basiert auf Entscheidungsbaum-Verfahren und sollte eine Alternative zu den anderen Klassifizierern darstellen. Ob dies insgesamt gelungen ist, fasse ich nun in folgenden Abschnitt zusammen. Erklärung zu J48 können Sie im Kapitel 3.1.4 finden.

Ergebnisse der Attributanalyse im Kapitel 4

Beim Vorhertest auf dem Komplettdatensatz zeigte J48 ein sehr schwaches Ergebnis. Es war mit großem Abstand das schwächste Attribut, hatte auch nur bei 1 von 10 Datensätzen einen Genauigkeits-Prozentwert von über 40% erzielen können. Alle Prozentwerte lagen weit unter dem ZeroR-Wert.

Bei J48 waren die Ergebnisse vom WrapperBFBE annehmbar, dafür aber die Foldwerte von den anderen beiden Wrappern und vor allem bei WrapperLFS deutlich niedriger als bei vielen anderen Klassifizierern. Lediglich „DiffGermanyN“ wurde als einziges Attribut von WrapperLFS sehr stark bevorzugt, was dieses Attribut auch an die Spitze der J48-Tests führte. Die schwachen Ergebnisse der anderen Attribute führten jedoch auch dazu, dass bei den Ergebnissen sehr niedrige Folds an Attribute vergeben wurden und die Abstände zwischen den Durchschnittsfolds einzelner Attribute sehr niedrig waren. Als Resultat waren somit bei vielen guten Attributen niedrigere Platzierungen rausgekommen als bei vorherigen Klassifizierern (0,1 Fold konnte bei J48 bereits um mehrere Plätze nach hinten absteigen).

Bei den Ergebnissen von J48 gewann, wie bereits erläutert, mit einem ziemlich großen Abstand das Attribut „DiffGermanyN“. Insgesamt waren in der Top10 zwar Differenzen vorhanden aber diesmal waren auch einige manuell hinzugefügte Attribute und Vereinsnamenattribute „Heim“ und „Gast“ dort anzutreffen. Auch bei den schwächsten Attributen gab es Überraschungen, wie beispielsweise komplett Ausprägungen von einzelnen Gruppenattributen. Ebenso waren einige bekannte Attribute unter den Schlechtesten zu finden (VieleTore z.B.). Bezüglich der Heim- und Gastausprägungen könnte in J48 keine deutliche Tendenz erkannt und daher auch der Trend der von JRip und Naive Bayes gesetzt wurde, bei J48 nicht bestätigt werden.

Beim Vorher-Nachher Vergleich mit kleineren Datensätzen verbesserte sich J48 deutlich. 7 der 11 Attribute verbesserten das Vorherergebnis des Komplettdatensatzes beträchtlich. Teilweise erreichte J48 eine Verbesserung gegenüber dem Komplettdatensatz von über 5%. Auffällig war auch, dass die großen Verbesserungen überwiegend mit den kleineren 40er oder gar dem kleinsten FStudio-Datensatz erreicht wurden, während größere Datensätze die Klassifikation durch J48 eher wenig verbesserten oder gar verschlechterten (Ausnahme „Fold40“-Datensatz, siehe 2.Auffälligkeit). Zweite Auffälligkeit ist, dass alle Datensätze ohne „Heim“ und „Gast“-Attribut oder nur mit einem „Gast“-Attribut das Ergebnis bei J48-Datensatz verschlechtert haben. Hier gab es keine Ausnahme, die Vereinsnamenattribute waren für J48 also von höchster Bedeutung. So kam es also, dass Datensätze, die diese 2 Auffälligkeiten in sich beinhalten, gut abschneiden (Top40hg, InfoGain40, FStudio sind die besten Datensätze, die sowohl eine Datensatzgröße <= 40 haben, als auch die „Heim“ und „Gast“-Attribute beinhalten).

Ergebnisse bei praktischen Tests im Kapitel 5

Beim Antreten gegen Tippstrategien hatte J48 sehr schlechte Ergebnisse erzielt. Der Klassifizierer war bei Benchmark1 kein einziges Mal in der Top10, immer einige oder sogar viele Datensätze unter den schlechtesten 10. Insgesamt hatte J48 das durchschnittsschlechteste Ergebnis von allen Klassifizierern. Lediglich im Benchmark2, mit Akzeptierung von „leichten Misses“, waren einige J48-Datensätze weiter vorne. Ebenso verlor J48 im Durchschnitt gegen die Tippstrategien. Bei der Tabellenbildung, war J48 dagegen gut. Es überholte SVM deutlich war einige Male sogar vor JRip und

in einer Saison (2007) sogar gemeinsam mit Naive Bayes auf Platz 1. Insgesamt belegte J48 hier den 3.Platz. Die Ergebnisse bei den Tabellen selbst waren bei J48 ziemlich überraschend. Traditionsverein Bayern München belegte kein einziges Mal alleine den ersten Rang. Lediglich in einer Saison wurde Bayern zusammen mit 2 anderen Vereinen als Meister vorhergesagt. Von Punktzahlen waren die Vorhersagen insgesamt in Ordnung.

Bei sehr großem Trainingsset erzielte J48 deutlich bessere Ergebnisse. Dabei bestätigte J48 die Tendenz für kleine Attributsets, in dem es den kleinen BestOf1963-Datensatz dem Komplett1963-Datensatz bevorzugte und darauf starke Ergebnisse (die stärksten für J48 während der gesamten Diplomarbeit) erzielte.

Vorhersagen von J48

Die Vorhersagen von J48 sollte man in 2 verschiedene Teile aufsplitten: Vorhersagen auf großen und auf kleinen Datensätzen. Auf großen Datensätzen wie dem Komplettdatensatz, gleicht J48 fast schon einem Zufallstipper, völliges Chaos und Mangel an Realismus, oft genau das Gegenteil von den Tipps von JRip und Naive Bayes und die darausfolgenden Genauigkeiten von teilweise unter 40%. Auf kleineren Datensätzen jedoch sind die Tipps im Großen und Ganzen annehmbar und haben durchaus Potenzial. Zwar gibt es auch dann immer noch einige nicht nachvollziehbare Tipps aber diese sind dann eher in der Unterzahl.

Ergebnisse in den Versuchen

Im ersten Versuch machte J48 im „Komplett“ sowie im kleineren Datensatz etwas unterschiedliche Vorhersagen. Allerdings war der „kleinere“ Datensatz mit 104 Attributen immer noch ziemlich groß, was immer noch in vielen „komischen“ und nicht nachvollziehbaren Tipps mündete.

Im zweiten Versuch, bei den Saisonvorhersagen, tippte J48 völlig zufällig und unverständlich, was aber, wie bereits erwähnt, an der Größe der Datensätze liegen könnte.

Fazit J48

Generell zeigte J48 bei den Tests eher durchwachsene Ergebnisse, was zu einem Rückschluss führen könnte, dass dieser Klassifizierer für zukünftige Analysen außen vor gelassen werden sollte. Allerdings zeigte J48 mit sinkender Anzahl an Attributen großes Steigerungspotential, was ihn für weitere Betrachtungen eventuell sehr interessant machen könnte. Dennoch aufgrund der Vielzahl von eher schwachen Ergebnissen in den vorherigen Tests zählt J48 ganz sicher nicht zur ersten Wahl für die Ergebnisvorhersagen.

SVM

Support Vektor Maschinen gelten als eines der Genauesten jedoch aber auch als eines der komplexesten und auch zeitaufwendigsten Verfahren. Ob die Genauigkeit in meinem Fall zutrifft und ob sich das aufwendige und vor allem zeitintensive Verfahren lohnt, sollte analysiert werden. Erklärung zu SVM kann im Kapitel 3.1.5 gefunden werden.

Ergebnisse der Attributanalyse im Kapitel 4

Beim Test auf dem Komplettdatensatz zeigte SVM eher durchwachsene Ergebnisse. Die Vorhersagegenauigkeit lag bei Anfang-Mitte 40%. Die Ergebnisse waren hinter JRip und Naive Bayes auf dem 3.Platz und lagen noch meistens hinter ZeroR.

Bei SVM erzielten 2 von 3 Wrappern (WrapperRSG und WrapperLFS) annehmbare Ergebnisse. Beim dritten Wrapper (WrapperBFBE) waren die Ergebnisse an der Grenze zum Annehmbaren. Bei

WrapperBFBE wurden wieder zu viele Folds vergeben, jedoch waren es nicht soviele, dass man keinen Sinn aus den Ergebnissen herauslesen könnte (wie bei JRip und auch später bei Random Forest). Wie bereits berichtet und bereits bekannt war SVM sehr zeitaufwendig, einfachere Verfahren dauerten bei SVM bereits mehrere Stunden und WrapperBFBE wäre ohne die Hilfe von „Cluster-Server“ und der separaten Cross-Validation des KE-Fachgebietes der TU Darmstadt zeitlich nicht zu schaffen.

Bei den Ergebnissen von SVM ist hervorzuheben, dass auch hier alle 10 Plätze von Differenzen belegt wurden. Insbesondere Differenzausprägungen von Gruppenattributen waren oft in der Top10 vertreten. Bei den schwächsten Attributen von SVM, fanden sich die Vereinsnamen-Attribute wieder, was darauf schließen lässt, dass diese für SVM von geringer Bedeutung sein könnten. Bei den restlichen schwachen Attributen gibt es keine Überraschungen. Auffällig ist, dass auch bei SVM die Gastausprägungen wieder bessere Ergebnisse erzielten als die Heimausprägungen.

Bei Vorher-Nachteil Vergleich wurde ein sehr eindeutiges aber auch komisches Ergebnis erzielt. Alle Datensätze verbesserten die Ergebnisse von SVM gegenüber dem Komplettdatensatz deutlich, der beste Datensatz sogar um fast 4%. Verwunderlich war, dass fast alle Datensätze waren in 7 gleichen Fällen besser und bei 3 gleichen Training- und Testsets schlechter waren. Scheinbar wurde bei diesen 3 „schlechteren“ Testfällen ein Wert nahe dem maximalen für SVM in dieser Saison bereits mit dem Komplettdatensatz erreicht. Die Datensätze mit den meisten Verbesserungen waren diejenigen ohne Vereinsnamen (also genau entgegengesetzter Fall zu J48). Die Datensätze mit Vereinsnamen-Attributen erzielten zwar Verbesserungen, jedoch waren diese im Allgemeinen nicht so hoch. Es schien so als hätten sich die Vereinsnamen-Attribute auf die Verbesserung stagnierend ausgewirkt (was der Vergleich von Top40hg und Top40-Datensätzen bei SVM bestätigt).

Ergebnisse von praktischen Tests im Kapitel 5

Bei den Tests gegen die Tippstrategien waren die Ergebnisse von SVM durchwegs gut, was in Saisons 2006 und 2007 sogar in den besten Rankingplätzen für SVM-Datensätze resultierte. Bei Saison 2006, bestand sogar die gesamte Top10 fast nur aus SVM-Datensätzen. In Vergleich zu Tippstrategien hatte SVM, ebenso wie JRip, immer mindestens eine bessere Datensatzausprägung im Feld und auch im Durchschnitt war SVM, bis auf eine Ausnahme in der Saison 2008, immer in der Top3. Im Gesamtdurchschnitt belegt SVM knapp hinter JRip den zweiten Platz.

Bei der Auswahl der Datensätze für die Tabellenvorhersagen wurde zweimal der beste Datensatz ohne Vereinsnamen ausgewählt, einmal der Komplettdatensatz, der in der Saison 2005, als beste Ausprägung von SVM abgeschnitten hat (eins von den 7 immer besseren Fällen, oben bereits erwähnt) und einmal der BestOf-Datensatz (mit Vereinsnamen). Die Ergebnisse des Tabellenaufbaus waren jedoch ernüchternd. Der bisher zu den besten gehörende Klassifizierer, tippte reihenweise die falschen Absteiger und Meister und gab bei Unsicherheit mehreren Teams die gleiche Punktzahl. Sogar die Differenzdatensätze (ohne Vereinsnamen), die bei den Vorher-Nachhertests gute Ergebnisse erzielten und bei Benchmark1 im Kapitel 5 in mehreren Saisons ganz vorne waren, brachten keine Verbesserung. SVM landete unter allen Klassifizierern auf dem letzten Platz. Wurde in 2 von 4 Saisons letzter, in einer vorletzter und lediglich in einer Saison zweiter (allerdings mit einer schwachen Punktzahl, da in dieser Saison alle Klassifizierer schwächelten).

Sein nächstes Fiasko hatte der Klassifizierer bei Tabellenaufbau mit großen Trainingssets. Hier erzielte SVM bei den Genauigkeiten gute Werte. Bei dem BestOf1963-Datensatz sogar den besten Wert unter allen Klassifizierern. Als es zum Tabellenaufbau kam, wurde jedoch das Problem erkannt. SVM tippte alle Spiele bis auf das jeweilige Bayern-Spiel auf „1“. Alle Spiele mit Bayern wurden auch auf Bayern getippt. Anscheinend war die Größe des Trainingssets zu viel für SVM, so dass dieser Klassifizierer auf solche einfache Lösung zurückgreifen musste. Nach der Vergrößerung des

Attributsets, tippte SVM dann nur noch auf „1“ und unterschied nicht mehr zwischen Bayern und den anderen (somit haben alle Teams 51 Punkte erhalten).

Vorhersagen von SVM

Bei den Vorhersagen selbst, kann man nur sehr schwer Aussagen treffen, denn viele Datensätze von SVM liegen ohne Vereinsnamen vor und die Vorhersagen selbst lassen sich dort daher nur schwer rauslesen. Bei den Datensätzen mit Vereinsnamen scheinen die Tipps zum größten Teil logisch zu sein und auch vom ganzen Spektrum („0“, „1“, „2“) ist etwas zu erkennen. Leider scheinen die Tipps aber zu keinem guten Gesamtergebnis zu führen, da die dadurch entstandenen Tabellen allesamt falsch waren.

Ergebnisse in den Versuchen

Beim ersten Versuch waren die Vorhersagen von SVM wie bei Naive Bayes eher starr, der Klassifizierer änderte die Vorhersagen bei verschiedenen Datensatzgrößen nur in äußersten Fällen.

Beim zweiten Versuch schnitten die Tipps von SVM hinter Naive Bayes am realistischsten ab. Es gab zwar hier und da ein paar unverständliche Tipps, im Großen und Ganzen waren die Vorhersagen jedoch nachzuvollziehen.

Fazit SVM

SVM stellte sich in dem Test, wie erwartet, als ein guter Klassifizierer mit guten Ergebnissen. Jedoch sind die Tabellen von SVM nicht so erfolgreich gewesen und ob sich der Aufwand der Ressourcen- und Zeitintensivität für die annähernd gleiche Genauigkeit wie z.B. bei JRip und Naive Bayes lohnt, bleibt dem Nutzer selbst zu entscheiden. Meiner Meinung nach lohnt sich, nach Ansicht der Ergebnisse, dieser Aufwand nach momentanem Stand nicht. Allerdings steht SVM natürlich als eine gute Alternative zu JRip und Naive Bayes zur Verfügung. Die Ergebnisse des Tests mit SVM waren bis auf die Tabellenbildung ebenso zufriedenstellend.

Random Forest

Random Forest ist eine Methode die Entscheidungsbäume und Ensemble-Methoden kombiniert. Also eine Methode, die eine schöne Ergänzung zu der bisherigen Klassifizierersammlung darstellt. Ob diese Methode bessere Ergebnisse als anderen Methoden zeigt, wird im Folgenden betrachtet. Erklärung zu Random Forest kann im Kapitel 3.1.6 gefunden werden.

Ergebnisse der Attributanalyse im Kapitel 4

Im Vorhertest auf dem Komplettdatensatz erzielte Random Forest eine Genauigkeit gerade noch im Bereich der 40%. Insgesamt lag es damit noch vor J48 auf dem vorletzten Platz. Die Ergebnisse waren damit deutlich unter dem ZeroR-Datensatz.

Bei der Hauptanalyse gab es hier die gleichen Probleme mit WrapperBFBE wie bei JRip. Dieser Wrapper lieferte zu hohe Fold-Ergebnisse. Die Ergebnisse der anderen beiden Wrappern waren annehmbar. Durch die gleichen Probleme wie JRip sind auch hier die Folds in der Gesamtwertung deutlich höher. Die Arbeitsgeschwindigkeit von Random Forest war positiv. Zwar kam man hier nicht an die Geschwindigkeit an Naive Bayes heran. Jedoch war die Geschwindigkeit der Attributtests besser als die von J48, JRip und natürlich SVM.

Nun zu den Ergebnissen von Random Forest:

Hier fanden sich neben den Differenzattributen bzw. Differenzen von Gruppenattributen auch 2 Ausprägungen von Abstiegsattributen in der Top10 wieder. Bei den schlechtesten Attributen konnte man sowohl die zeitlichen als auch andere schwächere Attribute am Ende des Feldes erkennen. Das einzige Attribut, das man unter den schwächsten 10 nicht entdecken konnte war „VieleTore“. Die Tendenz zu Gastausprägungen wurde von Random Forest bestätigt.

Beim Vorher-Nachher Vergleich, wurde Random Forest von 8 Datensätzen verbessert, was von allen Klassifizierern hinter SVM die meiste Verbesserungsanzahl darstellt. Hier konnte jedoch keine Regelmäßigkeit bei den besseren oder schlechteren Datensätzen festgestellt werden, da die Verbesserungen sowohl bei großen als auch kleinen Datensätzen erfolgten. Ebenso war auch keine Tendenz bei den Datensätzen mit „Heim“ und „Gast“-Attributen zu erkennen. Der beste Datensatz war „Fold85“, er verbesserte den Komplettdatensatz um 2,5 Prozent.

Ergebnisse bei praktischen Tests im Kapitel 5

Die Ergebnisse von Random Forest beim Test gegen die Tippstrategien waren, wie bei J48, eher schlecht. Der Klassifizierer war meistens hinter den Tippstrategien jedoch noch vor J48. Bei Tabellenvorhersagen, war Random Forest bei den ersten beiden Saisons mittelmäßig und bei den beiden anderen Saisons sehr schwach. Die 2 Punkte in der Saison 2008 waren ein Negativrekord. Den Meister hat Random Forest kein einziges Mal erraten und landet somit in Gesamttabellenwertung auf Platz 4, knapp vor dem katastrophalen SVM. Die Punkteverteilungen waren teilweise zu hoch, meistens jedoch realistisch.

Bei Tests mit großen Datensätzen zeigte Random Forest ebenso ein mittelmäßiges bis schlechtes Ergebnis und konnte auch durch die kleineren Datensätze nicht verbessert werden. Bei der Tabellenbildung waren die Punktzahlen und der Tabellenaufbau selbst akzeptierbar.

Vorhersagen von Random Forest

Bei den Vorhersagen von Random Forest erkennt man nicht immer einen roten Faden. Es ist zu erkennen, dass die Traditionsviere vor den Aufsteigern oft bevorzugt werden und der Klassifizierer nicht mit Unentschieden- oder Auswärtssiegetipps geizt. Allerdings waren die Vorhersagen bei den beiden Versuchen gegen Ende der letzten Saison eher zufällig (siehe nächster Absatz).

Ergebnisse in den Versuchen

Im ersten Versuch zeigte Random Forest oft mit dem großen Datensatz komplett andere Tipps als mit dem kleineren Datensatz.

Auch der Sinn von den Tipps im zweiten Versuch war zumindest anzuzweifeln.

Fazit Random Forest

Random Forest erzielt im Verlauf der Diplomarbeit durchwegs mittelmäßige bis schwache Ergebnisse. Ebenso zeigt dieser Klassifizierer keine klaren Tendenzen bei den Verbesserungsmöglichkeiten, wobei die Analyse des 4. Kapitels die Ergebnisse des Klassifizierers zumindest etwas verbessert habe. Generell hat sich der Klassifizierer durch seine eher mittelmäßige Ergebnisse nicht empfohlen. Zwar erzielte J48 gleiche oder teilweise sogar etwas schlechtere Resultate, jedoch wurde bei J48 ein Verbesserungspotenzial deutlich. Dieses Verbesserungspotenzial war bei Random Forest im Verlauf der gesamten Diplomarbeit nur begrenzt anzutreffen. Generell würde ich die Ergebnisse von Random Forest auf mittelmäßig bis schlecht schätzen.

5.5.2 Betrachtung der Attribute

Im Großen und Ganzen sind die Attributrankings so ausgefallen wie ich es erwartet habe. Beispielsweise habe ich erwartet, dass Differenzen ein deutlich stärkeres Ergebnis als 2 (Heim und Gast) Ausprägungen eines Attributs vorweisen können, da der Klassifizierer hier mit einem Wert eine Art Maßstab erhält und damit besser umgehen kann als mit 2 unterschiedlichen Werten. Ebenso war zu erwarten, dass die Anzahl der deutschen Nationalspieler ein Maßstab an Qualität bilden kann, was auch in der Realität zu sehen ist (Bayern, Leverkusen uvm. bauen auf deutsche Nationalspieler). Sorgen habe ich mir hier lediglich um die Verzerrung der Ergebnisse durch Altnationalspieler gemacht, die Ihre Leistung nicht mehr so abrufen können wie früher, jedoch mit Ihren zahlreichen Einsätzen in der Vergangenheit trotzdem Ihren Club im Ranking nach vorne schießen könnten. Noch stärker erwartet habe ich die Attribute rund um die Top9-Länder und brasilianische Nationalspieler, da diese Nationalmannschaften die stärksten Mannschaften der Welt bilden. Ich kann mir dies nur so erklären, dass einerseits von vielen Ländern nur wenige Nationalspieler vorhanden waren (z.B. Spanien) und andererseits viele Spieler, wie bereits erwähnt, Altnationalspieler waren und leider nur wenig Leistung brachten oder sich in Deutschland nicht adaptieren könnten. Während jedoch Top9 die Erwartungen dennoch größtenteils erfüllt hat (DiffNat, DiffNatPos), landeten die brasilianischen Nationalspieler eher weiter hinten im Feld. Diese relativ schwache Platzierung kann daran liegen, dass viele Brasilianer sich nicht in Deutschland adaptieren können oder eine längere Adaptationszeit brauchen (andere Spielart bzw. anderes Wetter/Leben als in Brasilien), die Vereine jedoch diese Geduld nicht bereit sind aufzubringen. Etwas besser habe ich die direkten Statistiken aus dem „Fußball Studio“ sowie die Vereinsnamen-Attribute erwartet. Andererseits haben die Klassifizierer, die die Vereinsnamenattribute scheinbar nicht beachtet hatten, beim Aufbau der Tabellen Probleme bekommen. Auch haben Differenzen aus den Werten der letzten Saison (DiffPktLS uvm.) gezeigt, dass es sich durchaus lohnt einen Blick auf die Tabelle der Vergangenheit zu werfen und diese keinesfalls zu ignorieren. Dies hängt oft damit zusammen, dass starke Mannschaften aus der letzten Saison oft größeren Etat für die aktuelle Saison zur Verfügung haben und dadurch Ihre Leistungen aus der Vorsaison oft bestätigen können (ebenso wie die schwächeren Mannschaften einen kleineren Etat haben).

Zu den Gruppenattributen: Die Form- und die Avgplatz-Attribute habe ich von Anfang an zu den stärksten Attributen gezählt und diese auch ganz vorne erwartet, denn aus den Ergebnissen der letzten Spielen einer Mannschaft und den Durchschnittsplatzierungen der letzten Jahre kann eine wirklich sehr sinnvolle Basis für Aussagen gebildet werden. Was ich aber nicht erwartet habe, ist dass eine bestimmte Saisonzahl solch einen starken Leistungsanstieg wie bei DiffAvg1-Belegung (=3 Saisons) gegenüber den anderen Saisons anzeigt. Scheinbar habe ich hier ein Extrema gefunden, vor welchem die Aussagen noch zu ungenau (weil zu wenige Jahre) und nach welchem die Aussagen schon zu ungenau (weil zu viele Daten) waren. Abstiegsattribute habe ich in etwa so erwartet. Es sind starke Attribute, die aber nicht herausragen. Es gibt nur 3 Absteiger jede Saison und es gibt ebenfalls einen festen Kreis von Mannschaften die selten bis nie absteigen und die in diesem Attribut immer dieselben Werte vorweisen. Versusattribute hätte ich etwas stärker eingeschätzt. Mir war klar, dass Versus nicht an Form und AvgPlatz rankommt, jedoch erwartete ich, dass es in die Dimensionen von „Abstieg“-Attributen vordringen kann. Es scheint also doch nicht so sehr vielsagend zu sein, wie die Mannschaften letzte Saison gegeneinander gespielt haben.

Jetzt gehe ich noch auf die schwächere Attribute ein. Dass die zeitlichen Attribute dort landen würden, könnte man schon vermuten. Zu wenig aussagend waren die einzelnen Monate und Jahreszeiten. Das Attribut „VieleTore“ war, wie bereits im 2.Kapitel angedeutet, ein Testattribut für die Effektivität des Wrappers. Dass es bei manchen Wrappern nicht letzter wurde, spricht eher gegen diese Wrapper, als für das Attribut. Das „VieleTore“ aber Gesamtletztes wurde, bestätigt das meiner Meinung nach gute und vielaussagende Gesamtergebnis. Die Attribute rund um die österreichischen (National-)Spieler waren zu schwach, da es einfach zu wenige Spieler aus Österreich in die Bundesliga sowohl in den Kader als auch in die Anfangself einer Bundesligamannschaft geschafft haben, um

generell eine gute Aussage machen zu können. Dies wurde bereits bei Eintragungen der manuellen Daten schon bemerkt. Schließlich erwähne ich auch die kombinierten Attribute. Einige von diesen wurden nur zu Testzwecken gemacht und daher ist es nicht überraschend, dass diese im Testfeld auch hinten zu finden waren.

Ganz am Ende gehe ich auf die stärkeren Gastausprägungen ein. Ich kann mir dieses Ergebnis nur so erklären, dass die Auswärtsschwäche vieler Vereine, die Heimstärke anderer Vereine „übertrumpft“. Ich vermute und dies wird auch mit dem Blick auf viele Tabellen deutlich, dass es viel mehr Vereine gibt, denen Auswärts ein Negativtrend angekreidet werden kann, als Heimvereine, die einen positiven Heitmehr haben. Ausnahmen von den stärkeren Gastausprägungen sind die „TFHeim“ und „TFGast“-Attribute (Attribute die anzeigen ob der Tabellenführer zuhause bzw. auswärts spielt). Auch diese Ausnahme kann jedoch erklärt werden, denn ein Tabellenführer ist sehr viel mehr für Heimserien auszeichnend als für positive oder negative Gastserien. Ebenso wird ein Team mit einer negativen Gastserie nur in Ausnahmefällen Tabellenführer, in diesen seltenen Fällen müsste das Team dann gleichzeitig eine herausragende Heimbilanz vorweisen. Natürlich gibt es auch Ausnahmen aber im Generellen überwiegt diese Tendenz.

5.5.3 Betrachtung von Datensätzen und Tippstrategien

Kommen wir nun noch abschließend zu den Datensätzen.

Der beste Datensatz ist der „InfoGain40“-Datensatz vor dem „BestOf“-Datensatz. Ehrlich gesagt habe ich den „BestOf“-Datensatz, aufgrund der Mühsamkeit und der Optimalität der Zusammenstellung im 4.Kapitel, auf Rang1 erwartet, habe aber die beiden InfoGain-Datensätze ebenso sehr hoch eingeschätzt, da InfoGain-Ranking Ergebnisse bereits im 4.Kapitel sehr interessant und vielaussagend waren.

Als eher mittelmäßig haben sich Datensätze ohne Vereinsnamen erwiesen. Vor allem im Test der „Top40hg“ vs „Top40“-Datensätze, also 2 Datensätzen einmal mit und einmal ohne Vereinsnamen-Attribute, hat sich gezeigt, dass die Nichtberücksichtigung von Vereinsnamen-Attributen bei den meisten Klassifizierern mit Verlusten in Verbindung gebracht werden kann.

Eher positiv überrascht hat der kleinste „FStudio“-Datensatz. Bei vielen Klassifizierern erreichte er die vorderen Plätze und bei einigen „starren“ Klassifizierern erreichte er ein „Pendeln“ sowohl in positiver als auch in negativer Richtung. Im Zusammenhang mit Tabellenvorhersagen von Naive Bayes schnitt „FStudio“ sehr gut ab.

Im Vergleich der beiden Rankings schnitten Top40- und Top85-Datensätze etwas besser als die Fold40- und Fold85-Datensätze ab, was die Prioritätsverteilung zwischen den beiden Rankings im Kapitel 4 bestätigt.

Die größeren 85-Datensätze oder gar der Komplett-Datensatz haben meistens etwas schlechtere Genauigkeitswerte erreicht als die 40-Datensätze, was daran liegen könnte, dass viele Klassifizierer mit weniger Daten einfach besser zureckkommen. Es sollte jedoch nicht zu wenige Daten sein, denn es wurde deutlich, dass die 40er-Datensätze doch im Durchschnitt etwas bessere Ergebnisse zeigten, als der kleinste Datensatz FStudio.

Bei Betrachtung der Saisonlängen von Datensätzen fielen mehrere Aussagerichtungen auf, daher kann hier auf keine genaue Aussage geschlossen werden. Einerseits folgen die Saisonlängen der Datensätze der Formel „je größer die Trainingssets desto besser“, andererseits zeigen sich in den Tests mit optimaler Grenze einmal bei 3, einmal bei 4 und einmal bei 6 Saisons keine klaren Aussagen. Ebenso zeigt der große „1963“-Datensatz auch, dass ein zu großer Trainingsset einer genauen Aussage im Wege stehen und in einer Missklassifikation der Datensätze durch die

Bevorzugung der Traditionsviere, sowie durch die immense teils kontroverse Datenmenge resultieren kann.

Was bleibt sind die Differenzen- und der manuelle Kicker-Datensatz. Während sich letzterer nirgends hervortun könnte und im Mittelmaß versank, war der Differenzen-Datensatz bei SVM solange erfolgreich bis es damit an die Tabellenbildung ging. Dort versagte der Differenz-Datensatz ohne die Vereinsnamen-Attribute.

Bei den Tippstrategien schneidet „TSTeams“ vor „TS3LS“ und „TSLS“ ab. Es macht also durchaus Sinn sich ein paar starke und schwache Teams auszuwählen und immer für bzw. gegen diese zu tippen. „TS3LS“ berechnet die Tabelle über 3 Saisons und erzielt damit bessere Werte als die Werte einer Saison. Diese wiederum erzielen bessere Werte als die aktuellen Tabellenplätze. Es macht also ebenso Sinn, länger in der Vergangenheit zu gucken anstatt nur in der Gegenwart zu verweilen. Die schwächsten Strategien waren Versus und natürlich die Zufallsstrategie.

Somit bin ich mit der qualitativen Diskussion in diesem Kapitel fertig. Abschließen möchte ich das Kapitel mit einer Vorhersage für die aktuelle Saison 2009/2010.

5.6. Vorhersage der aktuellen Saison (2009/2010)

Diese Vorhersage wird nach meinen Rückschlüssen aus der Qualitativen Diskussion mit einem „BestOf40-3S_Naive Bayes“-Datensatz getätigt.

Als erste Entscheidung stand für mich fest, dass der Naive Bayes-Klassifizierer diese Daten am Besten, denn „am Menschlichsten“ vorhersagt. Bei den Datensätzen musste ich zwischen dem „BestOf“-Datensatz und dem Gesamtbesten „InfoGain40“-Datensatz entscheiden. Die Wahl fiel auf den „BestOf“-Datensatz, da es im Kapitel 4 die Naive Bayes Werte am Besten und auch in den meisten Fällen verbessert habe. Schwerer war es bei Auswahl der Trainingssetgrößen, da hier kein klarer Trend festgestellt werden konnte. Hier habe ich mich für eine Trainingssetgröße von 3 Saisons entschieden, da Naive Bayes mit dieser Trainingssetgröße bei Benchmark1 im 5.Kapitel die besten Ergebnisse vorzeigte und den Komplettdatensatz hier in 3 von 4 Fällen noch verbessert habe.

Die komplette Vorhersage selbst kann im Anhang I betrachtet werden, die Endtabelle drucke ich mit den jeweils markierten Bereichen hier nachfolgend ab (auf die Punktzahlen verzichte jedoch, da diese Naive Bayes-typisch, schlichtweg falsch angesetzt sind). In den Markierungen unterscheide ich diesmal auch zwischen „Champions League“ und „Europa League“-Teilnehmern sowie zwischen den Absteigern und dem „Relegationsspiel-Teilnehmer“.

Endtabelle für die Saison 2009/2010

Platz	Vereinsname
1.	FC Bayern München
2.	Hamburger SV
-.	Werder Bremen (P)
4.	VFL Wolfsburg (M)
5.	VFB Stuttgart
6.	Schalke 04
7.	Bayer Leverkusen
8.	Borussia Dortmund
9.	Hertha BSC Berlin
10.	1899 Hoffenheim
11.	Hannover 96
12.	1. FC Köln
13.	Mönchengladbach
14.	Eintracht Frankfurt
15.	VFL Bochum
16.	1.FC Nürnberg (N)
17.	FSV Mainz 05 (N)
18.	SC Freiburg (N)

Tabelle 5.59: Vorhersagetabelle für die Saison2009/2010 (erstellt mit Naive Bayes_3S_BestOf-Datensatz)

Erläuterung die Farbunterlegungen und Markierungen

Dunkelgrün-Meister

Olivgrün-Int.Wettbewerbe (Champions League)

Hellgrün-Int.Wettbewerbe (Europa League)

Weinrot-Relegationsspiel-Teilnehmer

Rot-Absteiger

(M)-der aktuelle Meister

(P)-der aktuelle Pokalsieger

(N)-der aktuelle Aufsteiger

(Hamburg und Bremen werden als punktgleich berechnet, daher keine Unterscheidung zwischen dem 2. und 3.-Tabellenplatz)

Damit bin ich mit dem praktischen Kapitel fertig und komme zum abschließenden Fazitkapitel. In diesem werde ich Fazit aus dem erarbeiteten ziehen und eventuell Vorschläge für die möglichen nachfolgenden Arbeiten machen, die auf Basis von dieser Arbeit aufbauen sollten.

Fazit

Abschließend versuche ich ein Fazit zu ziehen, sowie für eventuelle weitere Diplomarbeiten, die auf das gleiche Thema eingehen und aufbauen sollten, einige Empfehlungen zu geben. Da die Ergebnisse dieser Diplomarbeit bereits im Kapitel 5.5 diskutiert wurden, gehe ich hier nicht mehr detailliert auf die Resultate meiner Diplomarbeit ein.

Mit dieser Diplomarbeit habe ich vieles erreicht, was ich mir vorgenommen habe. Natürlich war es nicht möglich, die Tendenzen genaustens oder sogar nur für zwei Drittel der Fälle richtig vorherzusagen, dieses habe ich auch nicht erwartet. Was mir aber weitaus wichtiger ist, es wurde erreicht, dass viele Daten gesammelt und analysiert wurden, so dass auf Basis dieser Analysen akzeptable Aussagen getroffen werden könnten. Auch sehe ich die Ergebnisse der Klassifiziererdatensätze, die im Vergleich zu den einfachen Tippstrategien deutlich besser abgeschnitten haben, für gegebene Möglichkeiten, als zufriedenstellend an.

Doch nun ein kurzer Blick darauf, was in dieser Diplomarbeit erreicht wurde:

- Dem Leser wurde eine umfassende Information über mögliche Tipp- und Wettsysteme gegeben. 2 Beispiele der Tippligen wurden modelliert. Ebenso wurde der Leser über die verwendeten Algorithmen und Werkzeuge informiert
- Es wurden 170 Attribute und Attributsausprägungen in einer mühevollen Arbeit zusammengestellt und eingetragen. Es wurde in der Diplomarbeit sowohl der Datengewinnungsprozess, als auch eine Beschreibung der einzelner Attribute und Attributsausprägungen erläutert
- Aus diesen 170 Attributen wurden in einem großen Test die Besten ausgewählt und mehrere Rankings erstellt, sowie einige Statistiken verglichen. Dabei wurde genaustens auf die Ergebnisse von jedem Klassifizierer eingegangen
- Auf Basis der Analyse wurden Datensätze gebildet, unter anderem ein „BestOf“-Datensatz, der das Beste aus den Analyserückschlüssen in sich zusammenfasste. Auch die anderen Analyseinformationen stellten verschiedene interessante Vergleichsparadigmen da, so dass am Ende 12 Datensätze für die weiteren Tests erstellt wurden
- Es wurde auf Basis der Datensätze ein Vorher-Nachher Vergleich getätig, in dem die Genauigkeiten der Datensätze auf den verschiedenen Klassifizierern vor und nach der Analyse gemessen, verglichen und kurz diskutiert wurden
- Den 12 Datensätzen wurden ebenfalls 12 verschiedenartige Tippstrategien hinzugefügt. Auf Basis der Datensätze und Tippstrategien wurden mehrere Rankings gebildet, in welchen die Datensätze einerseits untereinander und andererseits gegen die Tippstrategien verglichen wurden
- Auf Basis dieser Tests wurde mit den besten Datensätzen für alle verwendeten Klassifizierer für 4 verschiedene Saisons eine Saisontabelle vorhergesagt. In dieser Vorhersage wurde ein eigens konzipiertes Ranking verwendet, der die Klassifizierer nach der Wahrheit Ihrer Vorhersagen zur Realität verglichen und sortiert hat
- Ferner wurde noch eine Analyse mit einem sehr großen Trainingsset (ab 1963, ca 13000 Attribute) durchgeführt und die Ergebnisse verteilt auf Klassifizierer analysiert. Hier wurde ebenso mit 2 Datensätzen getestet, einem „Komplett“-Datensatz mit 110 Attributen und einem eigens hierfür konzipierten BestOf-Datensatz mit 21 Attributen.
- Schließlich wurden die Ergebnisse in einer qualitativen Diskussion sehr detailliert zusammengefasst, analysiert und diskutiert. Als Rückschluss dieser qualitativen Diskussion wurde mit einem Klassifizierer die Vorhersage für die aktuelle, nicht vollendete Saison (2009/2010) vollzogen

Der Leser mag sich nun fragen wieso mit einer derart größeren Datenmenge nicht mehr als 50% Genauigkeit erreicht wurde. Nun, es war aufgrund der Datenbeschränkungen, trotz des enormen Zeitaufwands und der Größe der Datensammlung, nicht sehr viel mehr mit den gegebenen Daten möglich. Viele der Daten sind nur sehr schwer zu finden bzw. zu sammeln und das Auffinden dieser

stellt einen enormen zeitlichen und teilweise sogar finanziellen Aufwand dar (siehe „Fußball Studio“). Vielmehr war diese Diplomarbeit als Basis für spätere Arbeiten gedacht, denn auf den finalen Erkenntnissen lässt sich durchaus aufbauen und die Genauigkeit mit konsequenter Ergänzung der Attribute und Weiterführung von Saisonbetrachtungen noch weiter gesteigert werden.

Daher möchte ich auch nachfolgend noch einige Ratschläge geben, wie man auf dieser Diplomarbeit weiter aufbauen und diese noch sinnvoll ergänzen kann:

Daten möglichst früh anfangen zu sammeln

Es ist sehr zu empfehlen einige Daten schon im Vorfeld der Diplomarbeit anfangen zu sammeln. Ein Attribut „verletzte oder gesperrte Spieler“ beispielsweise, wäre vor jedem Spieltag mit Sicherheit sehr aussagekräftig und könnte die Genauigkeit durchaus steigern. Das Problem ist aber diese Daten in den letzten Saisons festzustellen, wenn man keine Notizen gemacht hat oder eben keine Quelle vorliegt.

Sich eine klare Deadline für die Sammlung der Attribute setzen und viel Zeit für die Analyse einplanen

Attribute zu sammeln und zu planen kann eine Art „Sucht“ auslösen, da es unglaublich viel Spass machen kann, sich immer weitere Attribute zu überlegen und die Daten weiter zu ergänzen. Hier sollte man sich eine sehr klare Deadline setzen ab der man nur noch mit der Analyse und praktischen Tests weitermacht, denn sowohl die Analyse der Attribute, als auch auch die praktischen Tests nehmen sehr viel Zeit im Anspruch. Man sollte dies nicht unterschätzen.

Bei den Daten mehr auf Kader und auf Mannschaftsaufstellungen sowie auf einzelne Trainer und Spieler versuchen einzugehen

Generell sollte man sich in mehrere Richtungen entwickeln, dieser Vorschlag ist aber eine Richtung, die bei mir aus zeitlichen und auch umfangstechnischen Gründen leider zu kurz gekommen ist. Es lassen sich mit den Mannschaftsaufstellungen sowie mit der Betrachtung von einzelnen Trainern und Spielern viel mehr interessante Attribute erstellen, die durchaus Platz zur Entfaltung bieten.

Versuchen Test- und Trainingsset zu verknüpfen und eventuell Ergebnisse von neueren Saisons zu betonen

In weiteren Tests und Vorhersagen mittels eines Tippers in meiner Tippliga, der auf den Aussagen aus meiner Diplomarbeit basiert, hat sich angedeutet, dass das Hinzufügen von bereits abgeschlossenen Daten eines Testsets in den Trainingsset, sowie Betonung von neueren Saisons im Trainingsset (mittels z.B. Verdopplung von neueren Daten) ein besseres Vorhersageverhalten der einzelnen Klassifizierer erzielen kann. Es kann bestimmt vom Interesse sein, diesen Sachverhalt weiter zu untersuchen.

Differenzen von möglichst vielen Ausprägungen eines Attributs bilden

Es wird aus meiner Diplomarbeit mehr als deutlich, dass Differenzen die besten Werte erzielt haben. Auch ist es einfacher bei Bildung von neuen Attributen mit Differenzen, als mit mehreren Ausprägungen eines Attributs, zu arbeiten.

Diese Ratschläge sind nur einige der vielen Ratschläge, die ich für weitere themenverwandte Diplomarbeiten geben kann. Weitere Tipps sind oft in der Diplomarbeit selbst verankert. Dies bedeutet, dass aus den Ergebnissen meiner Diplomarbeit man für sich noch deutlich mehr Hinweise auf sein weiteres Vorgehen ziehen kann. Ein letzter Ratschlag ist dennoch meiner Meinung nach notwendig, *man sollte sich auf jeden Fall in dem Thema auskennen, Fußball oder andere jeweils betrachtete Sportarten mögen und die Diplomarbeit nicht als ein Mittel zum Zweck sehen*. Sonst kann die Materie dieser oder ähnlicher Diplomarbeit sehr trocken und langatmig erscheinen.

Im letzten Absatz meiner Diplomarbeit möchte ich noch kurz auf den Namen von ebendieser eingehen und erläutern, wieso ich den Begriff „im Sport“ und nicht „im Fußball“ im Namen meiner Diplomarbeit verwendet habe. Zwar wurden alle Untersuchungen in der Diplomarbeit in nur einer Sportart „Fußball“ durchgeführt, jedoch lassen sich die Ergebnisse und die Vorgehensweisen (natürlich mit den sportarttypischen Anpassungen) auf die meisten der anderen Sportarten übertragen. Bei jeder Sportart gibt es Sieger und Verlierer und es ist ebenso in jeder Sportart interessant diese mittels einer größeren Datensammlung und Aufbau von Trainings- und Testsets versuchen vorherzusagen. Es wurden hierzu bereits einige Forschungen durchgeführt, ein markantes Beispiel hierzu ist sicherlich die „Data Mining in NBA (National Basketball Association)“-Forschung, die versucht habe anhand der gesammelten Daten interessante Zusammenhänge in der amerikanischen professionellen Basketball-Liga festzustellen. **[NBA97]**

Somit kann diese Diplomarbeit ebenso als Basis für die Vorhersagen in anderen Sportbereichen zur Verfügung stehen.

Literaturquellen

[ARFF] Erklärung des ARFF-Formats von Weka URL:

<http://www.cs.waikato.ac.nz/~ml/weka/arff.html> (besucht am 10.6.2009)

[Brockh97] Brockhaus-Die Enzyklopädie in 24 Bänden, Band 12 „KIR“-„LAGH“, Brockhaus Verlag, 1997

[FDat] Quellen für Fussballdaten URLs: www.fussballdaten.de , www.kicker.de , www.transfermarkt.de , www.weltfussball.de (besucht mehrmals im Verlauf von Juni, Juli, August, sowie zuletzt am 5.9.2009)

[FM09] Internetauftritt des Computerspiels „Fußball Manager 2009“ URL: www.fm09.de (besucht am 27.5.2009)

[FSt] Internetauftritt des Tools „Fußball Studio“ von Volker Mallmann URL:
<http://www.vmlogic.net/cms/index.php> (besucht am 5.6.2009)

[FStExp] Beschreibung der Exportfunktion von „Fußball Studio“ URL:
http://help.vmlogic.net/wiki/handbuch:grundsaetzliche_funktionsweise (besucht am 6.6.2009)

[FStVP] Variablen/Parameter der Exportfunktion des „Fußball Studio“ URL:
http://help.vmlogic.net/wiki/handbuch:alle_variablen_parameter (besucht am 6.6.2009)

[GFHK09] M. Gütlein, E. Frank, M. Hall, A. Karwath *Large-scale attribute selection using Wrappers*. In Proc IEEE Symposium on Computational Intelligence and Data Mining, IEEE 2009
URL: http://wiki.pentaho.com/download/attachments/3801465/guetlein_et_al.pdf

[HanK06] J. Han, M. Kamber. *Data Mining Concepts and Techniques*, 2nd Edition, Morgan Kaufmann Publishers, 2006

[HATF08] T. Hastie, R. Tibshirani, J. Friedman. *The Elements of Statistical Learning: Data Mining, Inference and Prediction*, 2nd Edition, Springer Verlag, 2008

[Ing] Prof. G. Ingargiola, Temple University, College of Science and Technology, *Building Classification Models: ID3 and C4.5* URL: <http://www.cis.temple.edu/~ingargio/cis587/readings/id3-c45.html> (besucht am 29.6.2009)

[Kicker] Kicker-Saisonhefte 2002-2008, Olympia Verlag, 2002-2008

[Loy08] Roland Loy, *Das Lexikon der Fußballirrtümer*, C.Bertelsmann Berlag, 2008

[Mit97] Tom M. Mitchell, *Machine Learning*, McGraw-Hill, 1997

[NBA97] Inderpal Bhandari, Edward Colet, Jennifer Parker, Zachary Pines, Rajiv Pratap, Krishnakumar Ramanujam , *Data Mining and Knowledge Discovery in NBA* (Zeitschriftenbeitrag), Springer-Netherlands, 1997 , URL: <http://www.springerlink.com/content/r32q737858160r97>

[Oddset] Offizielle Seite von „Oddset-die Sportwette von Lotto“ URL: www.oddset.de (besucht am 26.5.2009)

[Perl] Einführung in Perl,2005 URL: http://perl-seiten.homepage.t-online.de/html/perl_einf.html (besucht am 26.5.2009)

[SCPDF05] R.L.Schwartz, T.Phoenix, b. d foy. *Einführung in Perl*, O'Reilly Verlag, 2005

[Tipp] Tippliga des Diplomauthors URL: www.awerbuch.de/tippliga (zuletzt besucht am 5.9.2009). Kl-Datensatz heißt „Tver713“ und ist momentan in der 2.Liga anzutreffen.

[UEFA] Quelle zur Berechnung der UEFA-Koeffizienten, URL:
<http://www.xs4all.nl/~kassiesa/bert/uefa/data/> (besucht am 11.7.2009)

[Weka] Weka, offizielle Homepage URL: <http://www.cs.waikato.ac.nz/ml/weka/> (besucht am 26.5.2009)

[WIFR05] I.H. Witten, E.Frank. *Data Mining Practical Machine Learning Tools and Techniques, 2nd Edition*, Morgan Kaufmann Publishers, 2005

[WIFRD01] I.H. Witten, E.Frank. *Data Mining Praktische Werkzeuge und Techniken für das maschinelle Lernen*, Carl Hanser Verlag, 2001

Anhang

Inhaltsverzeichnis des Anhangs

Anhang A: Verwendete Werkzeuge	156
Anhang B: Attributliste	158
Anhang C: Belegungstests im Kapitel 4.2	163
Anhang D: Ergebnisse der Hauptanalyse im Kapitel 4.4	178
Anhang E: Ergebnisse aus den „Nachher“-Tests im Kapitel 4.7.....	270
Anhang F: Ergebnisse der Test Datensätze vs Tippstrategien (Kapitel 5.2)	275
Anhang G: Komplette Tabellenergebnisse aus den Tabellenvorhersagen im Kap.5.3.	289
Anhang H: Komplette Tabellenergebnisse aus den Tabellenvorhersagen im Kap.5.4.	300
Anhang I: Kapitel 5.6-Vorhersage der Saison 2009/2010	302

Anhang A: Verwendete Werkzeuge

In diesem Unterkapitel beschreibe ich 2 von mir für die Diplomarbeit benutzte Werkzeuge „Perl“ und „Weka“.

Perl

(als Basis für dieses Unterkapitel dienten einerseits die Internetquelle „Einführung in Perl“ [Perl] sowie die ersten Seiten des Buchs „Einführung in Perl“ [SCPDF05])

Perl steht für "Practical Extraction and Report Language". Die Sprache wurde 1987 von Larry Wall entwickelt und diente ursprünglich als ein Werkzeug zur Verarbeitung und Änderung von Textdateien. Im Folgenden zähle ich vor allem die Vorteile von Perl auf die erklären sollen wieso ich mich gerade für diese Programmiersprache entschieden habe.

Das größte Argument für diese Programmiersprache ist sicherlich ihr Umgang mit Texten. Perl ist dafür bekannt, dass es einerseits die wichtigsten Programmierbefehle wie z.B. Verzweigungen und Schleifen enthält, andererseits auch die Möglichkeit bietet sehr leicht Such- und Ersetzungsoptionen in einem Editor durchzuführen. Die Suchen- und Ersetzen Funktion zeichnet sich ebenso durch hohe Geschwindigkeit aus. Die regulären Ausdrücke von Perl machen es möglich mit einem bereits sehr kurzen Befehl den notwendigen Text zu finden oder zu ersetzen. Beispielsweise könnte ich mit dem Befehl „\$line=~s/ö/oe/g;“ alle Ö-Umlaute in einem Text finden und durch „Oe“ ersetzen. Aufgrund dieser Vorteile und da ich bei meiner Diplomarbeit sehr viel mit Datenextraktion aus größeren Texten konfrontiert wurde und riesige Ergebnis-Textdateien schnell und effektiv in eine für mich notwendige Form (ARFF-Format für Weka) umwandeln musste, habe ich mich für Perl entschieden.

Weitere wenn auch nicht so relevante Argumente die für Perl sprachen waren: Die Einfachheit der Befehle, leichte Übertragbarkeit und Editierbarkeit der Perl-Programme auf verschiedenen Computern und das Vorhandensein der vorherigen Erfahrungen mit Perl in der Universität.

Sicherlich gibt es auch einige Nachteile von Perl, zum größten zählt für mich die Unübersichtlichkeit des Quelltextes nach einigen Wochen Arbeit. Aufgrund des fehlenden Prozedurwissens und der Vielzahl von regulären Ausdrücken, hat sich nach einer Zeit, trotz meiner fleißigen Kommentierung ein sehr großes Gewirr entwickelt, bei dem man bereits nach einer kurzen Programmierpause, einige Zeit wieder brauchte um sich in den Quelltext wieder reinzuarbeiten. Diesen Nachteil habe ich aber aufgrund der schnellen Suchfähigkeit von Perl gerne in Kauf genommen.

Weka

(als Basis für dieses Unterkapitel diente der offizielle Internetauftritt des Programms „Weka“ an der Waikato University [Weka])

Das Programm Weka wurde in der University of Waikato entwickelt. Der Name Weka steht für "Waikato Environment for Knowledge Analysis".

Weka ist eine Algorithmensammlung aus dem Bereich des Maschinellen Lernens. Es verfügt über eine Benutzeroberfläche, kann aber auch über die Kommandozeile benutzt werden. Das Programm ist so ausgelegt, dass man auf eine schnelle und einfache Weise vorhandene Methoden und Algorithmen des Maschinellen Lernens (ML) auf eigenen oder neuerstellten Datensätzen ausprobieren kann. Somit kann das Programm den Benutzer bei dem umfangreichen Data Mining Prozess unterstützen. Weka kann also unter Anderem: Bearbeitung von Daten vor Input, statistische Evaluierung von Lernmodellen, Visualisierung von Input-Daten und vom Gelernten. Außer der

Algorithmensammlung enthält Weka auch eine größere Anzahl von Preprocessing-Werkzeugen. Schließlich unterstützt Weka fast jede mögliche Plattform und ist somit universell einsetzbar. Wekas Start-Benutzeroberfläche wird in 4 Hauptbereiche unterteilt. In „Explorer“, „Experimenter“, „Knowledge Flow“ und „Simple CLI“.

1. *Explorer*: In Explorer kann man unter anderem Datensätze evaluieren, signifikante Attribute des Datensatzes herausfinden, Assoziationen zwischen den Attributen des Datensatzes finden, also, wie der Name bereits sagt, den Datensatz „erforschen“. Bei überwiegendem Teil der Diplomarbeit stand der Weka Explorer im Mittelpunkt (siehe Bild 1.1).
2. *Experimenter*: In diesem Bereich können Experimente des Maschinellen Lernens mit dem Datensatz durchgeführt werden. Man kann hier unter Anderem: eine Liste von ML-Algorithmen auf ein oder sogar mehrere Datensätze anwenden.

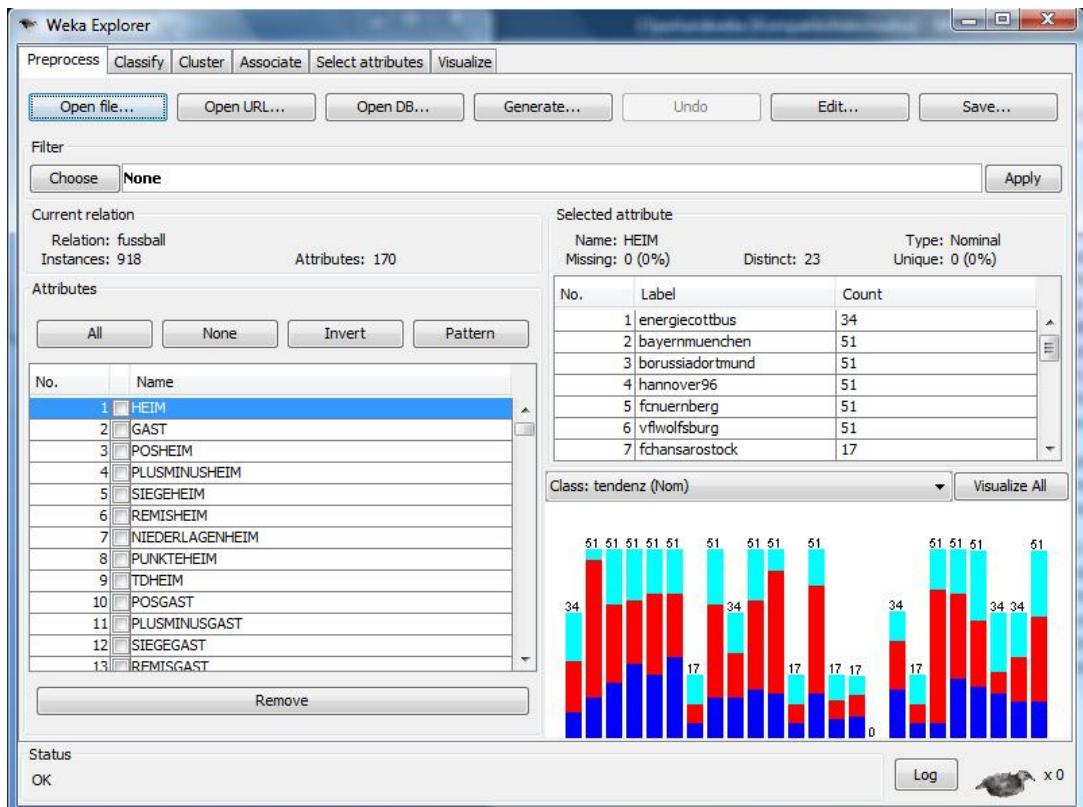


Bild A.1: Weka-Explorer-Interface

3. *Knowledge Flow*: Hier können Datenflüsse zwischen den einzelnen Komponenten modelliert werden. Somit kann mit „Knowledge Flow“ ein ML-Prozess sehr gut visualisiert werden.
4. *Simple CLI*: In Simple CLI kann man den vollen Funktionsumfang von Weka über eine Kommandozeile benutzen. Oft ist es von Vorteil, da man durch einen Befehl Zeit gegenüber mehreren Klicks in der Benutzeroberfläche sparen kann.

Wekas Inputdateien benutzen ein ARFF-Format auf das im Kapitel 2.2.1 eingegangen wurde.

Anhang B: Attributliste

Linke Spalte stellt die Attributnummer dar, die mittlere Spalte den jeweiligen Attributnamen und rechte Spalte-eine Kurzerklärung zum jeweiligen Attribut.

1/2	Heim/Gast	Heim(Gast)mannschaft des zu betrachtenden Spiels
3/10	PosHeim/-Gast	Position der Heim(Gast)mannschaft
4/11	PlusminusHeim/-Gast	Veränderung der Platzierung der Heim(Gast)mannschaft nach dem letzten Spieltag gegenüber der Platzierung vor dem letzten Spieltag
5/12	SiegeHeim/-Gast	Anzahl Siege der Heim(Gast)mannschaft im bisherigen Verlauf der Saison
6/13	RemisHeim/-Gast	Anzahl Unentschieden der Heim(Gast)mannschaft im bisherigen Verlauf der Saison
7/14	NiederlagenHeim/-Gast	Anzahl Niederlagen der Heim(Gast)mannschaft im bisherigen Verlauf der Saison
8/15	PunkteHeim/-Gast	Aktueller Punktestand der Heim(Gast)mannschaft
9/16	TDHeim/-Gast	Aktuelle Tordifferenz der Heim(Gast)mannschaft
17	Monat	Monat in dem das Spiel stattgefunden hat
18	Jahreszeit	Hier wurde das Attribut „Monat“ in 4 Jahreszeiten unterteilt: 4. „kalte Monate“ (Nov., Dez., Jan., Febr.) 5. „Monate mit durchschnittlicher Temperatur“ (März, April, Okt.) 6. „warme Monate“ (Rest)
19	Tabdiff	Differenz der aktuellen Tabellenplätze des Heim- und Gastteams
20	Tabtendenz	<pre>if (\$tabplatzdiff<=-6) {\$tabtendenz=1;} elseif (\$tabplatzdiff<=-1) {\$tabtendenz=2;} elseif (\$tabplatzdiff<=5) {\$tabtendenz=3;} else {\$tabtendenz=4;}</pre> <p>Das Attribut versucht also große Tabellenplatzunterschiede zu erkennen</p>
21	TabDiffForm	<p>Das Attribut versucht Tendenzen aus Plusminusattribut (#4 bzw. #11) zu erkennen. Dabei sind für dieses Attribut einfach nur die Fälle interessant, wenn ein Team sich am letzten Spieltag verschlechtert und das andere Team sich gleichzeitig verbessert hat.</p> <p>d.h. also: if (\$tabdiffheim>=0 && \$tabdiffgast<0) {\$tabdiffform=1;} elseif (\$tabdiffheim<0 && \$tabdiffgast>=0) {\$tabdiffform=2;} , else {\$tabdiffform=0;}</p> <p>Dabei ist ist Tabdiffheim="Plusminusheim"(Attr. #4) und Tabdiffgast=Plusminusgast (Attr. #11)</p>
22	TabDiffFormVerg	<p>Ähnlich zu Tabdiffform, jedoch werden hier die Vereine direkt miteinander verglichen</p> <pre>if (\$tabdiffheim>\$tabdiffgast) {\$tabdiffformverg=1;} elseif (\$tabdiffheim==\$tabdiffgast) {\$tabdiffformverg=0;} elseif (\$tabdiffheim<\$tabdiffgast) {\$tabdiffformverg=2;}</pre>
23	Siegdiff	Differenz der aktuellen Anzahl der Siege des Heim- und Gastteams
24	Remisdiff	Differenz der aktuellen Anzahl der Remis des Heim- und Gastteams

25	Niederlagendiff	Differenz der aktuellen Anzahl der Niederlagen des Heim- und Gastteams
26	Punktediff	Differenz der aktuellen Anzahl der Punkte des Heim- und Gastteams
27	TDdiff	Differenz der aktuellen Tordifferenz des Heim- und Gastteams
28	TorDiffForm	Ähnlich wie Tabdiffform (#21) nur mit Tordifferenz. Wenn Team A also positive Tordifferenz hat und gleichzeitig Team B negative Tordifferenz hat usw. if (\$tordiffheim>=0 && \$tordiffgast<0) {\$tordiffform=1;} elsif (\$tordiffheim<0 && \$tordiffgast>=0) {\$tordiffform=2;} else {\$tordiffform=0;}
29/30	TFHeim/-Gast	Wenn die Heim(Gast)mannschaft Tabellenführer ist, dann TFHeim(Gast)=1, sonst 0
31/32	Heimstark/Gaststart	Traditionell heimstarke (auswärtsschwache) Teams (manuell ausgewählt) erhalten eine 1, andere Teams 0
33	SiegeHeimTest	Testattribut, sobald die Anzahl der Heimsiege eine bestimmte Zahl nicht uebersteigt, SiegeHeimTest=0, else SiegeHeimTest=1 Sinn dieses Attributs ist es sehr heimschwache Teams zu kennzeichnen
34/35	Versus1Heim/-Gast	Ergebnisse der Aufeinandertreffen der beiden Teams in den letzten X1-Spielen von Team A (Heimteam) (bzw. Team B (Gastteam)) (<i>Belegung: 19</i>)
36	Versus1	Differenz von Versus1Heim und Versus1Gast
37/38	Versus2Heim/-Gast	Ergebnisse der Aufeinandertreffen der beiden Teams in den letzten X2-Spielen von Team A (Heimteam) (bzw. Team B (Gastteam)) (<i>Belegung: 18</i>)
39	Versus2	Differenz von Versus2Heim und Versus2Gast
40/41	Versus3Heim/-Gast	Ergebnisse der Aufeinandertreffen der beiden Teams in den letzten X3-Spielen von Team A (Heimteam) (bzw. Team B (Gastteam)) (<i>Belegung: 15</i>)
42	Versus3	Differenz von Versus3Heim und Versus3Gast
43/44	Versus4Heim/-Gast	Ergebnisse der Aufeinandertreffen der beiden Teams in den letzten X4-Spielen von Team A (Heimteam) (bzw. Team B (Gastteam)) (<i>Belegung: 12</i>)
45	Versus4	Differenz von Versus4Heim und Versus4Gast
46/47	Versus5Heim/-Gast	Ergebnisse der Aufeinandertreffen der beiden Teams in den letzten X5-Spielen von Team A (Heimteam) (bzw. Team B (Gastteam)) (<i>Belegung: 14</i>)
48	Versus5	Differenz von Versus5Heim und Versus5Gast
49	TabGast	Versucht einen "sichereren Heimsieg" herauszufinden, ab einem bestimmten Tabellenplatz des Gastteams ist Tabgast=1, sonst 0
50	Tabvergleich	Vergleicht den Tabellenplatz beider Mannschaften, wobei bei gleichem Tabellenplatz (soll es oft nach 1.Spieltag usw. geben) Heimteam den Vorzug hat if (\$tabelleheim<\$tabellegast) {\$tabvergleich=1;} elsif (\$tabelleheim>\$tabellegast) {\$tabvergleich=2;}

		else {\$stabvergleich=1;}
51/52	Form1HeimHeim/-GastAusw	Form des Heim(Gast)teams in den letzten X1-Heim(Auswärts)spielen) (<i>Belegung: 17</i>)
53/54	Form1Heim/-Gast	Form des Heim/(Gast)teams in den letzten X1-Spielen
55	DiffForm1	Differenz von Form1Heim und Form1Gast
56/57	Form2HeimHeim/-GastAusw	Form des Heim(Gast)teams in den letzten X2-Heim(Auswärts)spielen (<i>Belegung: 19</i>)
58/59	Form2Heim/-Gast	Form des Heim/(Gast)teams in den letzten X2-Spielen
60	DiffForm2	Differenz von Form2Heim und Form2Gast
61/62	Form3HeimHeim/-GastAusw	Form des Heim(Gast)teams in den letzten X3-Heim(Auswärts)spielen (<i>Belegung: 13</i>)
63/64	Form3Heim/-Gast	Form des Heim/(Gast)teams in den letzten X3-Spielen
65	DiffForm3	Differenz von Form3Heim und Form3Gast
66/67	Form4HeimHeim/-GastAusw	Form des Heim(Gast)teams in den letzten X4-Heim(Auswärts)spielen (<i>Belegung: 12</i>)
68/69	Form4Heim/-Gast	Form des Heim/(Gast)teams in den letzten X4-Spielen
70	DiffForm4	Differenz von Form4Heim und Form4Gast
71/72	Form5HeimHeim/-GastAusw	Form des Heim(Gast)teams in den letzten X5-Heim(Auswärts)spielen (<i>Belegung: 18</i>)
73/74	Form5Heim/-Gast	Form des Heim/(Gast)teams in den letzten X5-Spielen
75	DiffForm5	Differenz von Form5Heim und Form5Gast
76	VieleTore	<i>Testattribut</i> , Hier habe ich die Differenz: "Tordifferenz-Punkte" berechnet. Wenn die Tordifferenz die Punkte deutlich übersteigt gewinnt dieses Team oft sehr hoch (gute Moral)
77/78	Avg1PlatzHeim/-Gast	Durchschnittsplatzierung der Heim(Gast)mannschaft in den letzten X1 Saisons (<i>Belegung: 3</i>)
79	DiffAvg1	Differenz von Avg1PlatzHeim und Avg1PlatzGast
80/81	Avg2PlatzHeim/-Gast	Durchschnittsplatzierung der Heim(Gast)mannschaft in den letzten X2 Saisons (<i>Belegung: 7</i>)
82	DiffAvg2	Differenz von Avg2PlatzHeim und Avg2PlatzGast
83/84	Avg3PlatzHeim/-Gast	Durchschnittsplatzierung der Heim(Gast)mannschaft in den letzten X3 Saisons (<i>Belegung: 2</i>)
85	DiffAvg3	Differenz von Avg3PlatzHeim und Avg3PlatzGast
86/87	Avg4PlatzHeim/-Gast	Durchschnittsplatzierung der Heim(Gast)mannschaft in den letzten X4 Saisons (<i>Belegung: 4</i>)
88	DiffAvg4	Differenz von Avg4PlatzHeim und Avg4PlatzGast
89/90	Avg5PlatzHeim/-Gast	Durchschnittsplatzierung der Heim(Gast)mannschaft in den letzten X5 Saisons (<i>Belegung: 6</i>)
91	DiffAvg5	Differenz von Avg5PlatzHeim und Avg5PlatzGast
92/93	HaeufigstePlazierungHeim/-Gast	Dieses Attribut zählt auf welchem Tabellenplatz sich das Heim(Gast)team am häufigsten in der letzten Saison aufgehalten hat. Bei Gleichheit der Häufigkeit zweier Tabellenplätze, wird der am spätesten belegte genommen
94/95	ADDNationalTop9Heim/-Gast	Anzahl der Nationalspieler aus Top9-Ländern beim Heim(Gast)team
96/97	ADDNationalländerspieleTop9 Heim/-Gast	Anzahl der Länderspiele der Nationalspieler aus Top9-Ländern bei Heim(Gast)team

98/99	ADDNationalTop9PosHeim/-Gast	Anzahl der Nationalspieler aus Top9-Ländern des Heim(Gast)teams geordnet nach Position
100/101	ADDNationalLänderspielerTop9PosHeim/-Gast	Anzahl der Länderspiele der Nationalspieler aus Top9-Ländern bei Heim(Gast)team geordnet nach Position
102/103	ADDMittelNatlspHeim/-Gast	Durchschnitt aus dem Positionsattribut #98 und #100 (bei Gast Durchschnitt zwischen #99 und #101)
104/105	ADDLsPktHeim/-Gast	Erzielte Punkte des Heim(Gast)teams in der letzten Saison
106/107	ADDGeschosseneToreHeim/-Gast	Geschossene Tore des Heim(Gast)teams in der letzten Saison
108/109	ADD DurchgelasseneToreHeim/-Gast	Durchgelassene Tore des Heim(Gast)teams in der letzten Saison
110/111	ADDPositionHeim/-Gast	Position des Heim(Gast)teams in der letzten Saison
112/113	ADDStadiongroesseHeim/-Gast	Stadiongröße des Heim(Gast)teams
114/115	ADD10AbstiegHeim/-Gast	Anzahl Abstiege des Heim(Gast)teams in den letzten 10 Saisons
116/117	ADDBrazilHeim/-Gast	Anzahl der Brasilianer (nicht unbedingt Nationalspieler) im Kader des Heim(Gast)teams
118/119	ADDBrazilNatHeim/-Gast	Anzahl brasilianischer Nationalspieler im Kader des Heim(Gast)teams
120/121	ADDGermanyNatHeim/-Gast	Anzahl deutscher Nationalspieler im Kader des Heim(Gast)teams
122/123	ADDAustriaHeim/-Gast	Anzahl der Österreicher im Kader des Heim(Gast)teams (nicht unbedingt Nationalspieler)
124/125	ADDAustriaNatHeim/-Gast	Anzahl der österreichischen Nationalspieler im Kader des Heim(Gast)teams
126/127	ADDEwigeTabelleHeim/-Gast	Position des Heim(Gast)teams in der „Ewigen Tabelle“ (stand vor Saison 2008)
128/129	ADDUefa1JahrHeim/-Gast	Position des Heim(Gast)teams in der Uefa-Tabelle des jeweils letzten Jahres [UEFA]
130/131	ADDUefa5JahreHeim/-Gast	Position des Heim(Gast)teams in der Uefa-Tabelle der jeweils letzten 5 Jahre [UEFA]
132/133	ADDPunkteHeimHeim/-GastGast	Punkte des Heim(Gast)teams in der Heim(Gast)tabelle der letzten Saison
134/135	Abstieg1Heim/-Gast	Anzahl der Bundesliga-Saisons des Heim(Gast)teams in den letzten X1-Saisons (Belegung: 10)
136	DiffAbstieg1	Differenz von Abstieg1Heim und Abstieg1Gast
137/138	Abstieg2Heim/-Gast	Anzahl der Bundesliga-Saisons des Heim(Gast)teams in den letzten X2-Saisons (Belegung: 11)
139	DiffAbstieg2	Differenz von Abstieg2Heim und Abstieg2Gast
140/141	Abstieg3Heim/-Gast	Anzahl der Bundesliga-Saisons des Heim(Gast)teams in den letzten X3-Saisons (Belegung: 12)
142	DiffAbstieg3	Differenz von Abstieg3Heim und Abstieg3Gast
143/144	Abstieg4Heim/-Gast	Anzahl der Bundesliga-Saisons des Heim(Gast)teams in den letzten X4-Saisons (Belegung: 13)
145	DiffAbstieg4	Differenz von Abstieg4Heim und Abstieg4Gast
146/147	Abstieg5Heim/-Gast	Anzahl der Bundesliga-Saisons des Heim(Gast)teams in den letzten X5-Saisons (Belegung: 14)
148	DiffAbstieg5	Differenz von Abstieg5Heim und Abstieg5Gast
149	DiffPunkteHeimGast	Differenz der Attribute AddPunkteHeimHeim und AddPunkteGastGast

150	DiffPktLs	Differenz von AddLsPktHeim und AddLsPktGast
151	DiffGstLs	Differenz von ADDGeschosseneToreHeim und ADDGeschosseneToreGast
152	DiffKassTls	Differenz von ADDDurchgelasseneToreHeim und ADDDurchgelasseneToreGast
153	DiffPosLs	Differenz von ADDPositionHeim und ADDPositionGast
154	DiffEwigeTabelle	Differenz von ADDEwigeTabelleHeim und ADDEwigeTabelleGast
155	DiffNat	Differenz von ADDNationalTop9Heim und ADDNationalTop9Gast
156	DiffNat2	Differenz von ADDNationalalländerspieleTop9Heim und ADDNationalalländerspieleTop9Gast
157	DiffNatPos	Differenz von ADDNationalTop9PosHeim und ADDNationalTop9PosGast
158	DiffNat2Pos	Differenz von ADDNationalalländerspieleTop9PosHeim und ADDNationalalländerspieleTop9PosGast
159	DiffNatPosMit	Differenz von ADDMittelNatlspHeim und ADDMittelNatlspGast
160	DiffStadGr	Differenz von ADDStadiongroesseHeim und ADDStadiongroesseGast
161	Diff10Abstieg	Differenz von ADD10AbstiegHeim und ADD10AbstiegGast
162	DiffBrazil	Differenz von ADDBrazilHeim und ADDBrazilGast
163	DiffBrazilN	Differenz von ADDBrazilNatHeim und ADDBrazilNatGast
164	DiffGermanyN	Differenz von ADDGermanyNatHeim und ADDGermanyNatGast
165	DiffAustria	Differenz von ADDAustriaHeim und ADDAustriaGast
166	DiffAustriaN	Differenz von ADDAustriaNatHeim und ADDAustriaNatGast
167	DiffUefa1J	Differenz von ADDUefa1JahrHeim und ADDUefa1JahrGast
168	DiffUefa5J	Differenz von ADDUefa5JahreHeim und ADDUefa5JahreGast
169	DiffHaeufPlaz	Differenz von HaeufigsteplatzierungHeim und HaeufigstePlatzierungGast
170	Tendenz (Klassenattribut)	Die Ergebnistendenz des Spiels. Dabei hat diese Attribut 3 Ausgangswerte: "0", "1" und "2". Eine „0“ steht für ein Unentschieden „1“ steht für einen Sieg der Heimmannschaft „2“ steht für einen Sieg der Gastmannschaft

Tabelle A1: Aufzählung und Erklärung der verwendeten Attribute

Anhang C: Belegungstests im Kapitel 4.2

Versus

Trainingsset: **2002-2004** Legende: 1.Zeile-Belegung, 2.Zeile-Ausprägung, 3.Zeile-Merit (*100)

		10		11		12		13		14	
Heim	Gast	Gesamt									
0,7	2	2,3	1,1	2	2,3	0,2	2,1	2,9	1,2	2,3	2,6
		15		16		17		18		19	
Heim	Gast	Gesamt									
0,2	2	2,3	0,2	2,1	1,7	0	2,1	1,8	0	2,2	1,7

Trainingsset: **2003-2005** Legende: 1.Zeile-Belegung, 2.Zeile-Ausprägung, 3.Zeile-Merit (*100)

		10		11		12		13		14	
Heim	Gast	Gesamt									
0	2,3	2,1	0	2,3	2,8	0	2	3,4	0	1,9	3,4
		15		16		17		18		19	
Heim	Gast	Gesamt									
0	2,6	3,2	0	2,2	3,2	0	2,5	3,2	0,4	2,5	3,7

Trainingsset: **2004-2006** Legende: 1.Zeile-Belegung, 2.Zeile-Ausprägung, 3.Zeile-Merit (*100)

		10		11		12		13		14	
Heim	Gast	Gesamt									
0	1,8	2,2	0	1,4	2,2	0	1,5	2,3	0	1,3	2,3
		15		16		17		18		19	
Heim	Gast	Gesamt									
0,2	1,9	3	0,3	1,4	2,9	0,3	1,7	3	0,2	2	3,4

Trainingsset: **2005-2007** Legende: 1.Zeile-Belegung, 2.Zeile-Ausprägung, 3.Zeile-Merit (*100)

		10		11		12		13		14	
Heim	Gast	Gesamt									
0	2,1	2,7	0	2,1	2,4	0	2,3	2,7	0,2	1,1	2,3
		15		16		17		18		19	
Heim	Gast	Gesamt									
0	2,2	2,7	0,2	2,1	2,8	0,2	2,2	2,9	0,2	2,6	3,2

Trainingsset: 2002-2005 Legende: 1.Zeile-Belegung, 2.Zeile-Ausprägung, 3.Zeile-Merit (*100)

10		11		12		13		14						
Heim	Gast	Gesamt												
0,9	2,1	2,2	1,3	2,4	2,5	0,8	2,1	2,9	1,4	2,3	2,9	0,9	2	2,6
15		16		17		18		19						
Heim	Gast	Gesamt												
0,7	2,1	2,6	0,7	2,1	2,5	0,4	1,9	2,5	0,9	2	2,8	2,1	2	3

Trainingsset: **2003-2006** Legende: 1.Zeile-Belegung, 2.Zeile-Ausprägung, 3.Zeile-Merit (*100)

Trainingsset: 2004-2007 Legende: 1.Zeile-Belegung, 2.Zeile-Ausprägung, 3.Zeile-Merit (*100)

Wertungskriterien (2007-2008) - Logistische Lösung Berücksichtigung, Lineare Berücksichtigung, Lineare Werte (+ 2009)														
10			11			12			13			14		
Heim	Gast	Gesamt	Heim	Gast	Gesamt	Heim	Gast	Gesamt	Heim	Gast	Gesamt	Heim	Gast	Gesamt
0,1	2,3	2,4	0	2,2	2,3	0,4	2	2,5	0,4	1,7	2,2	0,4	1,9	2,6
15														
Heim	Gast	Gesamt	Heim	Gast	Gesamt	Heim	Gast	Gesamt	Heim	Gast	Gesamt	Heim	Gast	Gesamt
1	2	2,7	1	1,9	2,7	1,2	1,9	2,8	1	2,1	3,1	1,4	2,2	3,3

Trainingsset: 2002-2006 Legende: 1.Zeile-Belegung, 2.Zeile-Ausprägung, 3.Zeile-Merit (*100)

Trainingssatz 2002-2003 - Legende: 1.Zeile Belegung, 2.Zeile Ausprägung, 3.Zeile Wert (-100)																	
10				11				12				13				14	
Heim	Gast	Gesamt	Heim	Gast	Gesamt	Heim	Gast	Gesamt	Heim	Gast	Gesamt	Heim	Gast	Gesamt	Heim	Gast	Gesamt
0,3	1,8	2,1	0,6	1,7	2,2	0,8	1,7	2,5	1,1	1,8	2,2	0,9	1,7	2,5			
15																19	
Heim	Gast	Gesamt	Heim	Gast	Gesamt	Heim	Gast	Gesamt	Heim	Gast	Gesamt	Heim	Gast	Gesamt	Heim	Gast	Gesamt
0,8	1,7	2,6	0,9	1,7	2,3	0,7	1,6	2,4	1	1,7	2,5	1,1	1,6	2,4			

Trainingsset: 2003-2007 Legende: 1.Zeile-Belegung, 2.Zeile-Ausprägung, 3.Zeile-Merit (*100)

Winnungsgeset. 2000-2001 - Tabelle der Beurteilung, 1. Runde: Beurteilung, 2. Runde: Auswertung, 3. Runde: Wertung (+ 2001)															
10				11				12				13			
Heim	Gast	Gesamt	Heim	Gast	Gesamt	Heim	Gast	Gesamt	Heim	Gast	Gesamt	Heim	Gast	Gesamt	
0	2,2	2,3	0	2,4	2,5	0,8	2	2,5	0,7	1,9	2,6	0,5	2,1	2,8	
15															
Heim	Gast	Gesamt	Heim	Gast	Gesamt	Heim	Gast	Gesamt	Heim	Gast	Gesamt	Heim	Gast	Gesamt	
0,6	2,1	2,8	0,5	2	2,6	0,9	2,1	2,6	0,8	2,1	2,9	1	2,1	3,1	

Trainingsset: **2002-2007** Legende: 1.Zeile-Belegung, 2.Zeile-Ausprägung, 3.Zeile-Merit (*100)

10				11				12				13				14			
Heim	Gast	Gesamt																	
0,8	2,1	2,3	1	2,1	2,2	1,1	1,8	2,5	1,2	1,7	2,4	1,1	1,7	2,4					
15				16				17				18				19			
Heim	Gast	Gesamt																	
1,1	1,8	2,5	1,1	1,7	2,2	1,1	1,7	2,3	1	1,8	2,4	1,2	1,8	2,6					

Form

Ergebnisse der saisonübergreifenden Berechnung

Trainingsset: 2002-2004 Legende: 1.Zeile-Belegung, 2.Zeile-Ausprägung(HH-HeimHeim, GA-GastAusw),
3.Zeile-Merit (*100)

10					11					12				
HH	GA	Heim	Gast	Diff	HH	GA	Heim	Gast	Diff	HH	GA	Heim	Gast	Diff
0,9	3,6	2,1	2,5	3,6	1,6	3,6	2	2,4	3,6	0,4	3,4	2,8	3	5,3
13					14					15				
HH	GA	Heim	Gast	Diff	HH	GA	Heim	Gast	Diff	HH	GA	Heim	Gast	Diff
0,9	3,2	2,3	2,6	4,4	0,5	3,4	2,7	2,6	4,1	0,2	3,4	2,4	3,1	3,8
16					17					18				
HH	GA.	Heim	Gast	Diff	HH	GA.	Heim	Gast	Diff	HH	GA.	Heim	Gast	Diff
0,2	3,9	2,5	3,3	5,1	0,9	3,8	2,6	3,4	6,5	1,3	3,5	2,6	3,7	5,7
19														
HH	GA.	Heim	Gast	Diff										
1,6	3,4	2,6	3,4	5,7										

Trainingsset: 2003-2005 Legende: 1.Zeile-Belegung, 2.Zeile-Ausprägung(HH-HeimHeim, GA-GastAusw),
3.Zeile-Merit (*100)

10					11					12				
HH	GA	Heim	Gast	Diff	HH	GA	Heim	Gast	Diff	HH	GA	Heim	Gast	Diff
1,2	5,1	2,9	4,3	6,4	1	4,8	2,5	4,1	6,4	0,7	4,7	2,6	5,1	7,5
13					14					15				
HH	GA	Heim	Gast	Diff	HH	GA	Heim	Gast	Diff	HH	GA	Heim	Gast	Diff
0,5	4,6	2,5	4,5	8	0,4	4,7	2,5	4,5	7,9	0,6	4,8	2,6	5,7	6,9
16					17					18				
HH	GA.	Heim	Gast	Diff	HH	GA.	Heim	Gast	Diff	HH	GA.	Heim	Gast	Diff
1,2	4,9	2,7	4,7	7,1	1,1	4,9	3,1	4,8	7,3	1,2	5	3,2	5,4	7,7
19														
HH	GA.	Heim	Gast	Diff										
1,8	4,7	3	5,6	7,1										

Trainingsset: 2004-2006 Legende: 1.Zeile-Belegung, 2.Zeile-Ausprägung(HH-HeimHeim, GA-GastAusw),
3.Zeile-Merit (*100)

10					11					12				
HH	GA	Heim	Gast	Diff	HH	GA	Heim	Gast	Diff	HH	GA	Heim	Gast	Diff
0	4,5	2,6	3,7	5,9	0,4	4,5	2,7	3,6	6,3	0	4,3	2,7	3,8	6,3
13					14					15				
HH	GA	Heim	Gast	Diff	HH	GA	Heim	Gast	Diff	HH	GA	Heim	Gast	Diff
0,2	3,6	2,8	3,3	6,2	0,5	3,5	2,8	3,2	6	1,4	3,3	3,1	3,8	5,7
16					17					18				
HH	GA.	Heim	Gast	Diff	HH	GA.	Heim	Gast	Diff	HH	GA.	Heim	Gast	Diff
1,7	3,3	3,3	3,2	5,8	2	3,1	3,8	3,2	6,4	2,3	3,8	3,9	3,8	6,1
19														
HH	GA.	Heim	Gast	Diff										
2,5	3,6	4	3,4	6,1										

Trainingsset: 2005-2007 Legende: 1.Zeile-Belegung, 2.Zeile-Ausprägung(HH-HeimHeim, GA-GastAusw),
3.Zeile-Merit (*100)

10					11					12				
HH	GA	Heim	Gast	Diff	HH	GA	Heim	Gast	Diff	HH	GA	Heim	Gast	Diff
1,5	4,5	3	3,5	5,6	1,9	4,3	3,1	3,6	6,1	1,7	4,4	3,1	3,7	6
13					14					15				
HH	GA	Heim	Gast	Diff	HH	GA	Heim	Gast	Diff	HH	GA	Heim	Gast	Diff
1,7	3,4	3,1	3,7	6,9	2,1	3,1	3	3,3	6,6	2	3,4	3,5	3,4	7,6
16					17					18				
HH	GA.	Heim	Gast	Diff	HH	GA.	Heim	Gast	Diff	HH	GA.	Heim	Gast	Diff
2	3	3,7	3,1	6,8	2,3	3,2	3,6	3,1	6,1	2,5	3,5	3,6	3,2	5,1
19														
HH	GA.	Heim	Gast	Diff										
2,7	3,5	3,7	3,1	5,8										

Trainingsset: 2002-2005 Legende: 1.Zeile-Belegung, 2.Zeile-Ausprägung(HH-HeimHeim, GA-GastAusw),
3.Zeile-Merit (*100)

10					11					12				
HH	GA	Heim	Gast	Diff	HH	GA	Heim	Gast	Diff	HH	GA	Heim	Gast	Diff
1,3	4,4	2	3,4	5,6	1,1	4,2	1,9	3,3	5	1,2	4,3	2,4	3,4	6,8
13					14					15				
HH	GA	Heim	Gast	Diff	HH	GA	Heim	Gast	Diff	HH	GA	Heim	Gast	Diff
1,1	3,8	2	3,4	6,7	0,4	3,9	2,2	3,5	6,3	0,9	4	2,2	4	6,3
16					17					18				
HH	GA.	Heim	Gast	Diff	HH	GA.	Heim	Gast	Diff	HH	GA.	Heim	Gast	Diff
1,2	4,1	2,4	3,9	6,8	1,7	4,1	2,7	4,1	7,2	1,9	4,2	2,6	4,4	7,2
19														
HH	GA.	Heim	Gast	Diff										
2	4	2,4	4,2	7										

Trainingsset: 2003-2006 Legende: 1.Zeile-Belegung, 2.Zeile-Ausprägung(HH-HeimHeim, GA-GastAusw),
3.Zeile-Merit (*100)

10					11					12				
HH	GA	Heim	Gast	Diff	HH	GA	Heim	Gast	Diff	HH	GA	Heim	Gast	Diff
1,4	4,6	2,5	3,7	6,3	1,7	4,7	2,6	3,6	6,2	1,3	4,8	2,5	4,6	7,4
13					14					15				
HH	GA	Heim	Gast	Diff	HH	GA	Heim	Gast	Diff	HH	GA	Heim	Gast	Diff
1,6	4,6	2,5	3,7	7,5	1,6	5,1	2,4	3,7	5,7	1,9	4,2	2,7	4,7	6,5
16					17					18				
HH	GA.	Heim	Gast	Diff	HH	GA.	Heim	Gast	Diff	HH	GA.	Heim	Gast	Diff
2	4,2	2,9	3,5	6,4	1,9	4,3	3,2	3,6	7,2	2,2	4,5	3,3	3,9	6,2
19														
HH	GA.	Heim	Gast	Diff										
2,3	4,3	3,1	3,8	7,1										

Trainingsset: 2004-2007 Legende: 1.Zeile-Belegung, 2.Zeile-Ausprägung(HH-HeimHeim, GA-GastAusw),
3.Zeile-Merit (*100)

10					11					12				
HH	GA	Heim	Gast	Diff	HH	GA	Heim	Gast	Diff	HH	GA	Heim	Gast	Diff
1,2	3,8	2,8	3	5,3	1,3	3,5	2,7	3,1	5,9	1	3,5	2,8	3,2	5,6
13					14					15				
HH	GA	Heim	Gast	Diff	HH	GA	Heim	Gast	Diff	HH	GA	Heim	Gast	Diff
1,4	2,8	2,7	2,9	5,9	1,8	2,7	2,8	2,8	5,2	1,9	2,8	3,2	3,5	6,1
16					17					18				
HH	GA.	Heim	Gast	Diff	HH	GA.	Heim	Gast	Diff	HH	GA.	Heim	Gast	Diff
1,9	2,7	3,3	2,8	6,1	2	2,8	3,5	2,8	6,4	2,2	3	3,5	3,2	6,2
19														
HH	GA.	Heim	Gast	Diff										
2,4	3	3,4	3	6,3										

Trainingsset: 2002-2006 Legende: 1.Zeile-Belegung, 2.Zeile-Ausprägung(HH-HeimHeim, GA-GastAusw),
3.Zeile-Merit (*100)

10					11					12				
HH	GA	Heim	Gast	Diff	HH	GA	Heim	Gast	Diff	HH	GA	Heim	Gast	Diff
1,4	4,2	2,2	3	5,6	1,5	4,1	2	2,9	5,4	1,4	4,5	2,4	3,6	6,4
13					14					15				
HH	GA	Heim	Gast	Diff	HH	GA	Heim	Gast	Diff	HH	GA	Heim	Gast	Diff
1,5	3,7	2,1	3,1	6,3	1,4	3,9	2,1	3	5,8	1,6	3,9	2,2	3,5	5,6
16					17					18				
HH	GA.	Heim	Gast	Diff	HH	GA.	Heim	Gast	Diff	HH	GA.	Heim	Gast	Diff
1,6	3,8	2,4	3,3	6,2	1,8	3,8	2,8	3,3	6,7	2	4	2,9	3,4	6,3
19														
HH	GA.	Heim	Gast	Diff										
2,2	3,9	2,7	3,4	6,6										

Trainingsset: 2003-2007 Legende: 1.Zeile-Belegung, 2.Zeile-Ausprägung(HH-HeimHeim, GA-GastAusw),
3.Zeile-Merit (*100)

10					11					12				
HH	GA	Heim	Gast	Diff	HH	GA	Heim	Gast	Diff	HH	GA	Heim	Gast	Diff
1,7	4,1	2,7	3,1	6,3	1,7	4	2,6	3,3	6,1	1,5	4,1	2,5	4,1	6,7
13					14					15				
HH	GA	Heim	Gast	Diff	HH	GA	Heim	Gast	Diff	HH	GA	Heim	Gast	Diff
1,7	3,9	2,4	3,3	6,6	1,8	4	2,4	3,5	5,9	2	3,6	2,8	4,1	6,8
16					17					18				
HH	GA.	Heim	Gast	Diff	HH	GA.	Heim	Gast	Diff	HH	GA.	Heim	Gast	Diff
1,9	3,5	3	3,4	6,5	2	3,6	3,1	3,2	6,6	2,1	3,7	3,2	3,5	6,3
19														
HH	GA.	Heim	Gast	Diff										
2,2	3,6	3,1	3,3	7										

Trainingsset: 2002-2007 Legende: 1.Zeile-Belegung, 2.Zeile-Ausprägung(HH-HeimHeim, GA-GastAusw),
3.Zeile-Merit (*100)

10					11					12				
HH	GA	Heim	Gast	Diff	HH	GA	Heim	Gast	Diff	HH	GA	Heim	Gast	Diff
1,4	3,8	2,2	2,7	5,3	1,4	3,8	2,1	2,7	5,4	1,4	3,9	2,3	3	5,9
13					14					15				
HH	GA	Heim	Gast	Diff	HH	GA	Heim	Gast	Diff	HH	GA	Heim	Gast	Diff
1,6	3,2	2,1	2,8	5,7	1,5	3,3	2,1	2,7	5,4	1,7	3,4	2,3	3,1	5,8
16					17					18				
HH	GA.	Heim	Gast	Diff	HH	GA.	Heim	Gast	Diff	HH	GA.	Heim	Gast	Diff
1,7	3,4	2,5	3,1	6,1	1,8	3,5	2,7	3,1	6,4	2	3,5	2,8	3,2	6,1
19														
HH	GA.	Heim	Gast	Diff										
2,1	3,5	2,7	3,1	6,4										

Ergebnisse der saisoninternen Berechnung (Vergleichstests für 2 Saisons)

Trainingsset: 2002-2004 Legende: 1.Zeile-Belegung, 2.Zeile-Ausprägung(HH-HeimHeim, GA-GastAusw),
3.Zeile-Merit (*100)

10					11					12				
HH	GA	Heim	Gast	Diff	HH	GA	Heim	Gast	Diff	HH	GA	Heim	Gast	Diff
0	0	0	0	2,5	0	0	0	0,2	2,7	0	0	0	0	2,5
13					14					15				
HH	GA	Heim	Gast	Diff	HH	GA	Heim	Gast	Diff	HH	GA	Heim	Gast	Diff
0	0	0	0	2,6	0	0	0	0	2,4	0	0	0	0	2,9
16					17					18				
HH	GA.	Heim	Gast	Diff	HH	GA.	Heim	Gast	Diff	HH	GA.	Heim	Gast	Diff
0	0	0	0	2,6	0	0	0	0	<u>2,6</u>	0	0	0	0	2,6
19														
HH	GA.	Heim	Gast	Diff										
0	0	0	0	2,6										

Trainingsset: 2003-2007 Legende: 1.Zeile-Belegung, 2.Zeile-Ausprägung(HH-HeimHeim, GA-GastAusw),
3.Zeile-Merit (*100)

10					11					12				
HH	GA	Heim	Gast	Diff	HH	GA	Heim	Gast	Diff	HH	GA	Heim	Gast	Diff
0,7	0,8	1,1	1,5	3,6	0,3	0,5	0,9	1,4	3,6	0	1	1,3	1,5	3,6
13					14					15				
HH	GA	Heim	Gast	Diff	HH	GA	Heim	Gast	Diff	HH	GA	Heim	Gast	Diff
0	0,9	0,9	1,2	3,4	0	0,3	0,9	0,9	3,3	0	0,5	0,7	0,5	3,3
16					17					18				
HH	GA.	Heim	Gast	Diff	HH	GA.	Heim	Gast	Diff	HH	GA.	Heim	Gast	Diff
0	0,4	0,7	0,5	3,3	0	0,4	0,5	0,5	3,9	0	0,4	0,7	0,5	3,9
19														
HH	GA.	Heim	Gast	Diff										
0	0,4	0,7	0,5	3,9										

AvgPlatz

Trainingsset: 2002-2004 Legende: 1.Zeile-Belegung, 2.Zeile-Ausprägung, 3.Zeile-Merit (*100)

	1	2	3	4	5									
Heim	Gast	Gesamt												
2,4	2,4	3,3	2,4	3,4	4,2	2,9	3,4	5,4	2,5	2,3	4	2,5	1,9	4
	6	7	8	9	10									
Heim	Gast	Gesamt												
2,2	3,5	4,9	2,7	4,4	5,4	2,8	2,6	4,1	2,4	2,1	2,9	2	0,2	2,2

Trainingsset: 2003-2005 Legende: 1.Zeile-Belegung, 2.Zeile-Ausprägung, 3.Zeile-Merit (*100)

	1	2	3	4	5									
Heim	Gast	Gesamt												
3,1	4	4,7	2,7	5,3	6	2,8	5,8	8	2,6	5	5,2	2,8	3,7	4,1
	6	7	8	9	10									
Heim	Gast	Gesamt												
2,6	4,6	4,9	2,7	5,3	6,8	2,4	5,1	4,3	2,5	3,5	3,6	2,5	2,8	3

Trainingsset: 2004-2006 Legende: 1.Zeile-Belegung, 2.Zeile-Ausprägung, 3.Zeile-Merit (*100)

	1	2	3	4	5									
Heim	Gast	Gesamt												
2,8	3,9	5	2,8	4,1	5,4	3,7	4,7	8	2,9	4,5	5	3,2	3,5	4,6
	6	7	8	9	10									
Heim	Gast	Gesamt												
2,6	3,4	4,5	2,7	3,4	4,4	2,4	2,8	3,9	2,5	3,2	3,8	2,7	2,7	3,5

Trainingsset: 2005-2007 Legende: 1.Zeile-Belegung, 2.Zeile-Ausprägung, 3.Zeile-Merit (*100)

	1	2	3	4	5									
Heim	Gast	Gesamt												
2,8	3,2	5	3,5	3,5	5,5	3,8	3,8	5,7	3,2	3,8	5,2	3,6	3	5
	6	7	8	9	10									
Heim	Gast	Gesamt												
3,1	3	4,2	2,7	3	3,9	2,3	3	4,1	2,8	3,4	4	3,4	3,6	4,6

Trainingsset: 2002-2005 Legende: 1.Zeile-Belegung, 2.Zeile-Ausprägung, 3.Zeile-Merit (*100)

	1			2			3			4			5		
Heim	Gast	Gesamt													
2,5	3,2	4,6	2,9	3,7	4,7	2,9	2,9	5,9	2,5	3,1	4,4	3,2	2,5	3,8	
	6			7			8			9			10		
Heim	Gast	Gesamt													
2,4	3,8	4,7	2,9	4	5,7	2,3	3,4	3,9	2,3	2,2	3,1	2,2	2,2	2,8	

Trainingsset: 2003-2006 Legende: 1.Zeile-Belegung, 2.Zeile-Ausprägung, 3.Zeile-Merit (*100)

	1			2			3			4			5		
Heim	Gast	Gesamt													
2,7	3,6	4,3	2,5	4,6	5,3	2,9	4,9	7,4	2,6	4,1	4,8	2,6	2,9	3,8	
	6			7			8			9			10		
Heim	Gast	Gesamt													
2,8	3,5	4,4	2,7	4,2	6,3	2,2	3,4	4	2,3	3,4	3,7	2,3	2,8	3,2	

Trainingsset: 2004-2007 Legende: 1.Zeile-Belegung, 2.Zeile-Ausprägung, 3.Zeile-Merit (*100)

	1			2			3			4			5		
Heim	Gast	Gesamt													
2,9	3	4,7	3,1	3,3	5,2	3,7	3,8	7,2	3,1	3,9	5,1	3,4	3,1	4,8	
	6			7			8			9			10		
Heim	Gast	Gesamt													
3,1	3,2	4,5	3,2	3,2	4,5	2,8	2,8	4	2,6	2,8	3,6	2,8	2,4	3,7	

Trainingsset: 2002-2006 Legende: 1.Zeile-Belegung, 2.Zeile-Ausprägung, 3.Zeile-Merit (*100)

	1			2			3			4			5		
Heim	Gast	Gesamt													
2,3	3,1	4,6	3,3	3,5	5,2	2,8	3,7	6	2,6	2,9	4,6	3,2	2,3	3,7	
	6			7			8			9			10		
Heim	Gast	Gesamt													
2,9	3,2	4,3	3	3,5	5,6	2,3	2,5	4,5	2,2	2,1	3,3	2,2	2	2,8	

Trainingsset: 2003-2007 Legende: 1.Zeile-Belegung, 2.Zeile-Ausprägung, 3.Zeile-Merit (*100)

	1			2			3			4			5		
Heim	Gast	Gesamt													
2,9	3	4,3	3,1	3,8	5,3	3,1	4,1	7,1	3,3	3,7	5,2	3,2	2,8	5	
	6			7			8			9			10		
Heim	Gast	Gesamt													
3,1	3,3	4,6	3,4	3,9	5,4	3,1	3,4	5	2,7	2,8	3,6	2,5	2,6	3,3	

Trainingsset: **2002-2007** Legende: 1.Zeile-Belegung, 2.Zeile-Ausprägung, 3.Zeile-Merit (*100)

1			2			3			4			5		
Heim	Gast	Gesamt												
2,5	2,6	4,5	3,5	3,2	5,3	2,9	3,4	6,2	2,9	2,7	5,4	3,6	2,5	4,4
6			7			8			9			10		
Heim	Gast	Gesamt												
3,3	3,1	4,4	3,4	3,3	5,2	3	2,6	4,9	2,5	1,9	3,2	2,3	2,1	2,9

Abstieg

Trainingsset: 2002-2004 Legende: 1.Zeile-Belegung, 2.Zeile-Ausprägung, 3.Zeile-Merit (*100)

6			7			8			9			10		
Heim	Gast	Gesamt												
2,2	2,5	3	2,4	3,5	3,3	1,9	3,2	3,3	1,7	3,2	3,2	3	3,9	3,4
11			12			13			14			15		
Heim	Gast	Gesamt												
3	3,9	3,4	3,1	3,6	3,6	3,4	3,4	3,3	3,1	3	3	3,1	3	2,9

Trainingsset: 2003-2005 Legende: 1.Zeile-Belegung, 2.Zeile-Ausprägung, 3.Zeile-Merit (*100)

6			7			8			9			10		
Heim	Gast	Gesamt												
1,8	3,4	3,2	1,7	4,2	3,9	1,6	4,3	4	2,1	5,2	3,5	3,2	6	4,1
11			12			13			14			15		
Heim	Gast	Gesamt												
3,2	6	4,3	3,2	6	4,6	3,3	5,8	4,2	3	5,4	3,8	2,8	4,9	3,7

Trainingsset: 2004-2006 Legende: 1.Zeile-Belegung, 2.Zeile-Ausprägung, 3.Zeile-Merit (*100)

6			7			8			9			10		
Heim	Gast	Gesamt												
2	3,3	3,1	2	3,1	3,3	2	3,1	3,1	2,3	3,7	3,2	3	4,6	4
11			12			13			14			15		
Heim	Gast	Gesamt												
3	4,6	4,3	3	4,6	4,4	3	4,6	3,9	2,7	4,2	3,6	2,5	3,9	3,3

Trainingsset: 2005-2007 Legende: 1.Zeile-Belegung, 2.Zeile-Ausprägung, 3.Zeile-Merit (*100)

6			7			8			9			10		
Heim	Gast	Gesamt												
2,3	3	3,7	2,3	3	3,6	2,3	3	3,4	2,8	3,7	3,8	3,1	4,2	4,2
11			12			13			14			15		
Heim	Gast	Gesamt												
3	4,1	4,1	3	4,1	4,2	3	4,1	3,8	3	4,1	3,7	2,7	3,8	3,5

Trainingsset: 2002-2005 Legende: 1.Zeile-Belegung, 2.Zeile-Ausprägung, 3.Zeile-Merit (*100)

	6	7	8	9	10									
Heim	Gast	Gesamt												
2,2	3	4,2	2,3	3,8	4,9	1,9	3,7	3,7	2,1	4,1	3,4	2,9	4,7	3,8
	11		12		13		14		15					
Heim	Gast	Gesamt												
2,9	4,7	4	3	4,5	4,2	3,1	4,4	3,7	2,9	4,1	3,3	2,8	3,8	3,2

Trainingsset: 2003-2006 Legende: 1.Zeile-Belegung, 2.Zeile-Ausprägung, 3.Zeile-Merit (*100)

	6	7	8	9	10									
Heim	Gast	Gesamt												
1,9	2,8	3,7	1,9	3,7	3,5	1,8	3,5	3,7	2	4,2	3,2	3	5,2	3,9
	11		12		13		14		15					
Heim	Gast	Gesamt												
3	5,2	4,1	3	5,2	4	3	5,1	4,3	2,8	4,7	3,5	2,6	4,4	3,3

Trainingsset: 2004-2007 Legende: 1.Zeile-Belegung, 2.Zeile-Ausprägung, 3.Zeile-Merit (*100)

	6	7	8	9	10									
Heim	Gast	Gesamt												
2,4	3	3,4	2,4	3	3,9	2,4	3	3,6	2,6	3,4	3,3	3,2	4	3,9
	11		12		13		14		15					
Heim	Gast	Gesamt												
3,2	3,9	3,8	3,2	3,9	3,9	3,2	3,9	3,5	3	3,6	3,3	2,7	3,4	3,2

Trainingsset: 2002-2006 Legende: 1.Zeile-Belegung, 2.Zeile-Ausprägung, 3.Zeile-Merit (*100)

	6	7	8	9	10									
Heim	Gast	Gesamt												
1,9	2,7	4	2,2	3,4	4,5	1,8	3,2	3,9	2	3,7	3,5	2,8	4,4	3,9
	11		12		13		14		15					
Heim	Gast	Gesamt												
2,8	4,4	3,9	2,9	4,3	4,1	2,9	4,1	3,6	2,8	3,9	3,5	2,6	3,7	3

Trainingsset: 2003-2007 Legende: 1.Zeile-Belegung, 2.Zeile-Ausprägung, 3.Zeile-Merit (*100)

	6	7	8	9	10									
Heim	Gast	Gesamt												
2,2	2,8	4	2,2	3,5	4,3	2,1	3,4	4,4	2,3	3,9	3,6	3,1	4,6	3,9
	11		12		13		14		15					
Heim	Gast	Gesamt												
3,1	4,5	3,8	3,1	4,5	3,9	3,2	4,4	3,6	3	4,2	3,4	2,8	3,9	3,3

Trainingsset: 2002-2007 Legende: 1.Zeile-Belegung, 2.Zeile-Ausprägung, 3.Zeile-Merit (*100)

			6			7			8			9			10		
Heim	Gast	Gesamt															
2	2,7	4	2,1	3,2	4,5	1,9	3,1	4,3	2,2	3,5	4	2,9	4	3,8			
			11			12			13			14			15		
Heim	Gast	Gesamt															
2,9	4	3,7	3	3,9	3,8	3	3,8	3,5	2,9	3,6	3,3	2,8	3,4	3,4			

Anhang D: Ergebnisse der Hauptanalyse im Kapitel 4.4

Legende für den gesamten Anhang D

NR – Attributnummer

Platz – Gesamtplatzierung im entsprechenden Ranking

Name – Attributname

24 – Anzahl Folds ausgewählt bei dem entsprechenden Verfahren am Trainingsset mit Saisons 2002-2004 (entsprechend 35-am Trainingsset mit Saisons 2003-2005 usw.)

3SD – Durchschnittswert von Trainingssets mit der Länge von 3 Saisons (entsprechend 4SD-Durchschnitt 4 Saison-Trainingssets usw.)

Ges – Gesamtdurchschnitt entsprechendes Verfahren

Gesamt – Gesamtdurchschnitt über alle 3 Verfahren (nur im jeweiligen Gesamtunterkapitel)

Prozent – Prozentwert der Foldauswahl über alle 3 Verfahren (nur im jeweiligen Gesamtunterkapitel)

Anmerkung: Leere Felder in einigen Methoden sind gleichzusetzen mit 0.

Wrapper mit Klassifizierer: „Naive Bayes“ und Suchmethode: „RankSearch“

NR	Platz	Name	24	35	46	57	25	36	47	26	37	27	3SD	4SD	5SD	Ges
1	121	HEIM	0	0	0	0	0	0	0	0	0	0	0,00	0,00	0,00	0,00
2	121	GAST	0	0	0	0	0	0	0	0	0	0	0,00	0,00	0,00	0,00
3	105	POSHEIM	0	0	1	1	0	0	0	0	0	0	0,50	0,00	0,00	0,20
4	121	PLUSMINUSHEIM	0	0	0	0	0	0	0	0	0	0	0,00	0,00	0,00	0,00
5	121	SIEGEHEIM	0	0	0	0	0	0	0	0	0	0	0,00	0,00	0,00	0,00
6	121	REMISHEIM	0	0	0	0	0	0	0	0	0	0	0,00	0,00	0,00	0,00
7	116	NIEDERLAGENHEIM	0	0	0	0	0	0	1	0	0	0	0,00	0,33	0,00	0,10
8	121	PUNKTEHEIM	0	0	0	0	0	0	0	0	0	0	0,00	0,00	0,00	0,00
9	92	TDHEIM	0	0	3	2	0	0	0	0	0	0	1,25	0,00	0,00	0,50
10	77	POSGAST	0	7	0	1	0	0	0	0	0	1	2,00	0,00	0,00	0,90
11	121	PLUSMINUSGAST	0	0	0	0	0	0	0	0	0	0	0,00	0,00	0,00	0,00
12	73	SIEGEGAST	0	7	0	1	1	1	1	0	0	0	2,00	1,00	0,00	1,10
13	121	REMISGAST	0	0	0	0	0	0	0	0	0	0	0,00	0,00	0,00	0,00
14	121	NIEDERLAGENGAST	0	0	0	0	0	0	0	0	0	0	0,00	0,00	0,00	0,00
15	92	PUNKTEGAST	0	3	0	0	0	0	2	0	0	0	0,75	0,67	0,00	0,50
16	83	TDGAST	0	6	1	1	0	0	0	0	0	0	2,00	0,00	0,00	0,80
17	121	MONAT	0	0	0	0	0	0	0	0	0	0	0,00	0,00	0,00	0,00
18	121	JAHRESZEIT	0	0	0	0	0	0	0	0	0	0	0,00	0,00	0,00	0,00
19	77	TABDIFF	1	7	0	1	0	0	0	0	0	0	2,25	0,00	0,00	0,90
20	121	TABTENDENZ	0	0	0	0	0	0	0	0	0	0	0,00	0,00	0,00	0,00
21	121	TABDIFFFORM	0	0	0	0	0	0	0	0	0	0	0,00	0,00	0,00	0,00
22	121	TABDIFFFORMVERG	0	0	0	0	0	0	0	0	0	0	0,00	0,00	0,00	0,00
23	77	SIEGDIFF	0	6	1	0	0	2	0	0	0	0	1,75	0,67	0,00	0,90
24	121	REMISDIFF	0	0	0	0	0	0	0	0	0	0	0,00	0,00	0,00	0,00
25	86	NIEDDIFF	0	5	0	0	0	0	0	0	1	0	1,25	0,00	0,50	0,60
26	61	PUNKTEDIFF	0	10	2	1	0	2	0	0	0	0	3,25	0,67	0,00	1,50
27	15	TDDIFF	7	10	2	1	4	2	1	4	0	4	5,00	2,33	2,00	3,50
28	121	TORDIFFFORM	0	0	0	0	0	0	0	0	0	0	0,00	0,00	0,00	0,00
29	31	TFHEIM	0	8	5	3	3	1	2	1	0	3	4,00	2,00	0,50	2,60
30	105	TFGAST	0	0	0	1	0	0	0	0	0	1	0,25	0,00	0,00	0,20

31	105	HEIMSTARK	0	0	0	1	0	1	0	0	0	0	0,25	0,33	0,00	0,20
32	121	GASTSTARK	0	0	0	0	0	0	0	0	0	0	0,00	0,00	0,00	0,00
33	121	SIEGEHEIM	0	0	0	0	0	0	0	0	0	0	0,00	0,00	0,00	0,00
34	70	VERSUS1HEIM	1	8	0	0	3	0	0	0	0	0	2,25	1,00	0,00	1,20
35	64	VERSUS1GAST	10	0	0	0	3	0	0	1	0	0	2,50	1,00	0,50	1,40
36	73	VERSUS1	0	6	1	1	0	2	0	0	0	1	2,00	0,67	0,00	1,10
37	96	VERSUS2HEIM	0	3	0	0	1	0	0	0	0	0	0,75	0,33	0,00	0,40
38	70	VERSUS2GAST	10	0	0	0	2	0	0	0	0	0	2,50	0,67	0,00	1,20
39	77	VERSUS2	0	4	1	1	0	2	0	0	0	1	1,50	0,67	0,00	0,90
40	121	VERSUS3HEIM	0	0	0	0	0	0	0	0	0	0	0,00	0,00	0,00	0,00
41	116	VERSUS3GAST	1	0	0	0	0	0	0	0	0	0	0,25	0,00	0,00	0,10
42	86	VERSUS3	0	2	0	1	0	2	0	0	0	1	0,75	0,67	0,00	0,60
43	121	VERSUS4HEIM	0	0	0	0	0	0	0	0	0	0	0,00	0,00	0,00	0,00
44	100	VERSUS4GAST	0	0	0	0	2	0	0	0	0	1	0,00	0,67	0,00	0,30
45	85	VERSUS4	0	3	0	0	1	2	0	0	0	1	0,75	1,00	0,00	0,70
46	121	VERSUS5HEIM	0	0	0	0	0	0	0	0	0	0	0,00	0,00	0,00	0,00
47	105	VERSUS5GAST	0	0	0	0	2	0	0	0	0	0	0,00	0,67	0,00	0,20
48	86	VERSUS5	0	3	0	0	0	2	0	0	0	1	0,75	0,67	0,00	0,60
49	121	TABGAST	0	0	0	0	0	0	0	0	0	0	0,00	0,00	0,00	0,00
50	96	TABVERGLEICH	0	4	0	0	0	0	0	0	0	0	1,00	0,00	0,00	0,40
51	121	FORM1HEIMHEIM	0	0	0	0	0	0	0	0	0	0	0,00	0,00	0,00	0,00
52	35	FORM1GASTAUSW	0	9	0	1	4	3	0	4	0	3	2,50	2,33	2,00	2,40
53	77	FORM1HEIM	0	2	0	2	1	2	0	0	0	2	1,00	1,00	0,00	0,90
54	40	FORM1GAST	0	10	0	1	4	2	0	3	0	3	2,75	2,00	1,50	2,30
55	21	DIFFFORM1	1	10	0	2	4	3	1	4	1	6	3,25	2,67	2,50	3,20
56	121	FORM2HEIMHEIM	0	0	0	0	0	0	0	0	0	0	0,00	0,00	0,00	0,00
57	35	FORM2GASTAUSW	1	8	1	1	4	3	0	3	0	3	2,75	2,33	1,50	2,40
58	77	FORM2HEIM	0	4	2	2	0	1	0	0	0	0	2,00	0,33	0,00	0,90
59	41	FORM2GAST	0	10	0	1	4	3	0	1	0	3	2,75	2,33	0,50	2,20
60	4	DIFFFORM2	2	10	5	3	4	5	2	4	2	6	5,00	3,67	3,00	4,30
61	121	FORM3HEIMHEIM	0	0	0	0	0	0	0	0	0	0	0,00	0,00	0,00	0,00
62	46	FORM3GASTAUSW	0	8	2	1	1	2	0	2	0	3	2,75	1,00	1,00	1,90
63	96	FORM3HEIM	0	2	0	2	0	0	0	0	0	0	1,00	0,00	0,00	0,40
64	52	FORM3GAST	0	9	0	2	1	2	0	1	0	2	2,75	1,00	0,50	1,70
65	4	DIFFFORM3	0	10	9	2	4	4	4	5	1	4	5,25	4,00	3,00	4,30
66	121	FORM4HEIMHEIM	0	0	0	0	0	0	0	0	0	0	0,00	0,00	0,00	0,00
67	21	FORM4GASTAUSW	0	10	3	2	4	4	0	4	0	5	3,75	2,67	2,00	3,20
68	105	FORM4HEIM	0	0	0	1	1	0	0	0	0	0	0,25	0,33	0,00	0,20
69	49	FORM4GAST	0	8	2	2	2	3	0	0	0	1	3,00	1,67	0,00	1,80
70	6	DIFFFORM4	0	10	5	3	4	4	4	4	2	5	4,50	4,00	3,00	4,10
71	121	FORM5HEIMHEIM	0	0	0	0	0	0	0	0	0	0	0,00	0,00	0,00	0,00
72	33	FORM5GASTAUSW	0	9	1	1	4	3	0	4	0	3	2,75	2,33	2,00	2,50
73	67	FORM5HEIM	0	7	1	2	1	2	0	0	0	0	2,50	1,00	0,00	1,30
74	41	FORM5GAST	0	10	0	1	4	2	0	2	0	3	2,75	2,00	1,00	2,20
75	3	DIFFFORM5	0	10	5	4	4	8	1	4	3	5	4,75	4,33	3,50	4,40
76	121	VIELETORE	0	0	0	0	0	0	0	0	0	0	0,00	0,00	0,00	0,00
77	105	AVG1PLATZHEIM	0	0	0	2	0	0	0	0	0	0	0,50	0,00	0,00	0,20

78	25	AVG1PLATZGAST	0	10	5	2	3	4	0	2	0	4	4,25	2,33	1,00	3,00
79	2	DIFFAVG1	3	10	5	5	8	4	1	4	1	5	5,75	4,33	2,50	4,60
80	64	AVG2PLATZHEIM	1	7	1	1	3	1	0	0	0	0	2,50	1,33	0,00	1,40
81	58	AVG2PLATZGAST	0	10	0	1	2	1	0	0	0	2	2,75	1,00	0,00	1,60
82	6	DIFFAVG2	4	10	5	1	5	2	1	6	1	6	5,00	2,67	3,50	4,10
83	67	AVG3PLATZHEIM	0	6	0	2	3	1	0	0	0	1	2,00	1,33	0,00	1,30
84	41	AVG3PLATZGAST	0	10	2	1	3	3	0	1	0	2	3,25	2,00	0,50	2,20
85	16	DIFFAVG3	0	10	5	2	4	3	1	1	4	4	4,25	2,67	2,50	3,40
86	100	AVG4PLATZHEIM	0	2	0	1	0	0	0	0	0	0	0,75	0,00	0,00	0,30
87	52	AVG4PLATZGAST	0	10	3	1	0	2	1	0	0	0	3,50	1,00	0,00	1,70
88	19	DIFFAVG4	3	10	4	2	4	2	1	3	1	3	4,75	2,33	2,00	3,30
89	64	AVG5PLATZHEIM	1	7	1	1	3	1	0	0	0	0	2,50	1,33	0,00	1,40
90	58	AVG5PLATZGAST	0	10	0	1	2	1	0	0	0	2	2,75	1,00	0,00	1,60
91	8	DIFFAVG5	4	10	5	1	4	2	1	6	1	6	5,00	2,33	3,50	4,00
92	116	HAEUFIGSTEPLAZIERUNGHEIM	0	0	0	1	0	0	0	0	0	0	0,25	0,00	0,00	0,10
93	105	HAEUFIGSTEPLAZIERUNGGAST	0	2	0	0	0	0	0	0	0	0	0,50	0,00	0,00	0,20
94	100	ADDNATIONALTOP9HEIM	0	1	0	2	0	0	0	0	0	0	0,75	0,00	0,00	0,30
95	46	ADDNATIONALTOP9GAST	0	10	0	0	2	4	0	1	0	2	2,50	2,00	0,50	1,90
96	35	ADDNATIONALLAENDERSP.TOP9H.	0	9	0	1	4	1	0	4	0	5	2,50	1,67	2,00	2,40
97	100	ADDNATIONALLAENDERSP.TOP9G.	0	2	0	0	0	0	0	0	0	1	0,50	0,00	0,00	0,30
98	105	ADDNATIONALTOP9POSHEIM	0	1	0	1	0	0	0	0	0	0	0,50	0,00	0,00	0,20
99	49	ADDNATIONALTOP9POSGAST	0	10	0	0	2	4	0	1	0	1	2,50	2,00	0,50	1,80
100	96	ADDNAT.LAENDERSP.TOPPOS9H.	0	0	0	1	1	1	0	1	0	0	0,25	0,67	0,50	0,40
101	121	ADDNAT.LAENDERSP.TOPPOS9G.	0	0	0	0	0	0	0	0	0	0	0,00	0,00	0,00	0,00
102	100	ADDMITTELNATLSPHEIM	0	1	0	1	1	0	0	0	0	0	0,50	0,33	0,00	0,30
103	105	ADDMITTELNATLSPGAST	0	2	0	0	0	0	0	0	0	0	0,50	0,00	0,00	0,20
104	73	ADDLSPKTHEIM	0	9	1	1	0	0	0	0	0	0	2,75	0,00	0,00	1,10
105	61	ADDLSPKTGAST	0	8	0	1	3	2	0	0	0	1	2,25	1,67	0,00	1,50
106	52	ADDGESCHOSSENTOREHEIM	0	9	4	1	0	2	1	0	0	0	3,50	1,00	0,00	1,70
107	86	ADDGESCHOSSENTOREGAST	0	6	0	0	0	0	0	0	0	0	1,50	0,00	0,00	0,60
108	121	ADDDURCHGELASSENTOREHEIM	0	0	0	0	0	0	0	0	0	0	0,00	0,00	0,00	0,00
109	121	ADDDURCHGELASSENTOREGAST	0	0	0	0	0	0	0	0	0	0	0,00	0,00	0,00	0,00
110	52	ADDPOSITIONHEIM	0	8	2	2	2	2	0	0	0	1	3,00	1,33	0,00	1,70
111	92	ADDPOSITIONGAST	0	5	0	0	0	0	0	0	0	0	1,25	0,00	0,00	0,50
112	105	ADDSTADIONGROESSEHEIM	0	0	1	1	0	0	0	0	0	0	0,50	0,00	0,00	0,20
113	121	ADDSTADIONGROESSEGAST	0	0	0	0	0	0	0	0	0	0	0,00	0,00	0,00	0,00
114	121	ADD10ABSTIEGHEIM	0	0	0	0	0	0	0	0	0	0	0,00	0,00	0,00	0,00
115	86	ADD10ABSTIEGGAST	0	5	0	0	1	0	0	0	0	0	1,25	0,33	0,00	0,60
116	121	ADDBRAZILHEIM	0	0	0	0	0	0	0	0	0	0	0,00	0,00	0,00	0,00
117	121	ADDBRAZILGAST	0	0	0	0	0	0	0	0	0	0	0,00	0,00	0,00	0,00
118	121	ADDBRAZILNATHEIM	0	0	0	0	0	0	0	0	0	0	0,00	0,00	0,00	0,00
119	121	ADDBRAZILNATGAST	0	0	0	0	0	0	0	0	0	0	0,00	0,00	0,00	0,00
120	46	ADDGERMANYNATHEIM	0	9	4	0	2	4	0	0	0	0	3,25	2,00	0,00	1,90
121	35	ADDGERMANYNATGAST	0	10	1	0	4	3	0	2	0	4	2,75	2,33	1,00	2,40
122	121	ADDAUSTRIAHEIM	0	0	0	0	0	0	0	0	0	0	0,00	0,00	0,00	0,00
123	121	ADDAUSTRIAGAST	0	0	0	0	0	0	0	0	0	0	0,00	0,00	0,00	0,00
124	121	ADDAUSTRIANATHEIM	0	0	0	0	0	0	0	0	0	0	0,00	0,00	0,00	0,00

125	121	ADDAUSTRIANATGAST	0	0	0	0	0	0	0	0	0,00	0,00	0,00	0,00		
126	52	ADDEWIGETABELLEHEIM	0	4	2	1	2	3	1	1	2	1	1,75	2,00	1,50	1,70
127	29	ADDEWIGETABELLEGAST	0	9	2	3	3	3	1	1	1	4	3,50	2,33	1,00	2,70
128	86	ADDUEFA1JAHRHEIM	0	0	0	6	0	0	0	0	0	0	1,50	0,00	0,00	0,60
129	25	ADDUEFA1JAHRGAST	0	10	1	1	4	3	1	2	3	5	3,00	2,67	2,50	3,00
130	92	ADDUEFA5JAHREHEIM	0	2	0	1	0	0	1	0	0	1	0,75	0,33	0,00	0,50
131	49	ADDUEFA5JAHREGAST	0	9	1	1	1	2	0	2	0	2	2,75	1,00	1,00	1,80
132	67	ADDPUNKTEHEIMHEIM	0	9	1	2	0	1	0	0	0	0	3,00	0,33	0,00	1,30
133	58	ADDPUNKTEGASTGAST	0	9	1	1	0	2	0	0	0	3	2,75	0,67	0,00	1,60
134	121	ABSTIEG1HEIM	0	0	0	0	0	0	0	0	0	0	0,00	0,00	0,00	0,00
135	16	ABSTIEG1GAST	0	10	4	1	4	4	0	4	2	5	3,75	2,67	3,00	3,40
136	44	DIFFABSTIEG1	0	8	2	2	2	2	0	1	0	4	3,00	1,33	0,50	2,10
137	121	ABSTIEG2HEIM	0	0	0	0	0	0	0	0	0	0	0,00	0,00	0,00	0,00
138	12	ABSTIEG2GAST	0	10	4	2	4	4	1	4	2	5	4,00	3,00	3,00	3,60
139	35	DIFFABSTIEG2	0	8	2	2	3	3	0	2	0	4	3,00	2,00	1,00	2,40
140	116	ABSTIEG3HEIM	0	0	0	0	1	0	0	0	0	0	0,00	0,33	0,00	0,10
141	12	ABSTIEG3GAST	0	10	4	2	4	4	1	3	3	5	4,00	3,00	3,00	3,60
142	29	DIFFABSTIEG3	0	8	2	2	3	4	0	3	0	5	3,00	2,33	1,50	2,70
143	105	ABSTIEG4HEIM	0	0	0	0	1	1	0	0	0	0	0,00	0,67	0,00	0,20
144	12	ABSTIEG4GAST	0	10	4	2	4	5	1	4	1	5	4,00	3,33	2,50	3,60
145	61	DIFFABSTIEG4	0	7	1	1	1	2	0	1	0	2	2,25	1,00	0,50	1,50
146	121	ABSTIEG5HEIM	0	0	0	0	0	0	0	0	0	0	0,00	0,00	0,00	0,00
147	23	ABSTIEG5GAST	0	10	3	2	4	4	0	3	1	4	3,75	2,67	2,00	3,10
148	83	DIFFABSTIEG5	0	2	0	1	1	2	0	1	0	1	0,75	1,00	0,50	0,80
149	8	DIFFPUNKTEHEIMGAST	0	10	4	6	3	4	1	4	3	5	5,00	2,67	3,50	4,00
150	31	DIFFPKTLS	0	10	5	3	2	2	1	1	0	2	4,50	1,67	0,50	2,60
151	44	DIFFGSTLS	0	10	1	1	4	2	0	0	0	3	3,00	2,00	0,00	2,10
152	52	DIFFKASSTLS	0	7	4	1	2	2	0	1	0	0	3,00	1,33	0,50	1,70
153	27	DIFFPOSLS	0	9	5	4	2	3	1	2	0	2	4,50	2,00	1,00	2,80
154	73	DIFFEWIGETABELLE	0	6	1	1	1	1	0	0	0	1	2,00	0,67	0,00	1,10
155	10	DIFFNAT	6	10	5	2	3	3	1	5	1	3	5,75	2,33	3,00	3,90
156	16	DIFNAT2	0	9	4	4	3	2	2	4	1	5	4,25	2,33	2,50	3,40
157	11	DIFNATPOS	2	10	4	2	3	4	5	3	1	4	4,50	4,00	2,00	3,80
158	33	DIFNATPOS2	4	8	3	2	2	2	1	1	0	2	4,25	1,67	0,50	2,50
159	19	DIFNATPOSMIT	0	10	5	2	3	3	1	4	1	4	4,25	2,33	2,50	3,30
160	121	DIFFSTADGR	0	0	0	0	0	0	0	0	0	0	0,00	0,00	0,00	0,00
161	116	DIFF10ABSTIEG	0	1	0	0	0	0	0	0	0	0	0,25	0,00	0,00	0,10
162	121	DIFFBRAZIL	0	0	0	0	0	0	0	0	0	0	0,00	0,00	0,00	0,00
163	121	DIFBRAZILN	0	0	0	0	0	0	0	0	0	0	0,00	0,00	0,00	0,00
164	23	DIFFGERMANYN	0	10	3	2	3	4	0	4	0	5	3,75	2,33	2,00	3,10
165	121	DIFFAUSTRIA	0	0	0	0	0	0	0	0	0	0	0,00	0,00	0,00	0,00
166	121	DIFFAUSTRIAN	0	0	0	0	0	0	0	0	0	0	0,00	0,00	0,00	0,00
167	27	DIFFUEFA1J	0	10	2	5	4	2	0	2	0	3	4,25	2,00	1,00	2,80
168	1	DIFFUEFA5J	10	10	5	3	6	4	1	5	0	6	7,00	3,67	2,50	5,00
169	70	DIFFHAEUFPLAZ	0	6	2	1	2	1	0	0	0	0	2,25	1,00	0,00	1,20

Wrapper mit Klassifizierer: „Naive Bayes“ und Suchmethode: „LinearForwardSelection“

NR	Platz	Name	24	35	46	57	25	36	47	26	37	27	3SD	4SD	5SD	Ges
1	59	HEIM	0	0	0	0	0	0	0	0	0	0	0,00	0,00	0,00	0,00
2	59	GAST	0	0	0	0	0	0	0	0	0	0	0,00	0,00	0,00	0,00
3	59	POSHEIM	0	0	0	0	0	0	0	0	0	0	0,00	0,00	0,00	0,00
4	59	PLUSMINUSHEIM	0	0	0	0	0	0	0	0	0	0	0,00	0,00	0,00	0,00
5	59	SIEGEHEIM	0	0	0	0	0	0	0	0	0	0	0,00	0,00	0,00	0,00
6	59	REMISHEIM	0	0	0	0	0	0	0	0	0	0	0,00	0,00	0,00	0,00
7	59	NIEDERLAGENHEIM	0	0	0	0	0	0	0	0	0	0	0,00	0,00	0,00	0,00
8	59	PUNKTEHEIM	0	0	0	0	0	0	0	0	0	0	0,00	0,00	0,00	0,00
9	59	TDHEIM	0	0	0	0	0	0	0	0	0	0	0,00	0,00	0,00	0,00
10	43	POSGAST	1	0	0	0	0	0	0	0	0	0	0,25	0,00	0,00	0,10
11	59	PLUSMINUSGAST	0	0	0	0	0	0	0	0	0	0	0,00	0,00	0,00	0,00
12	36	SIEGEGAST	0	0	0	0	0	1	0	0	1	0	0,00	0,33	0,50	0,20
13	59	REMISGAST	0	0	0	0	0	0	0	0	0	0	0,00	0,00	0,00	0,00
14	59	NIEDERLAGENGAST	0	0	0	0	0	0	0	0	0	0	0,00	0,00	0,00	0,00
15	59	PUNKTEGAST	0	0	0	0	0	0	0	0	0	0	0,00	0,00	0,00	0,00
16	15	TDGAST	2	2	0	1	0	0	0	0	1	1	1,25	0,00	0,50	0,70
17	59	MONAT	0	0	0	0	0	0	0	0	0	0	0,00	0,00	0,00	0,00
18	59	JAHRESZEIT	0	0	0	0	0	0	0	0	0	0	0,00	0,00	0,00	0,00
19	43	TABDIFF	0	1	0	0	0	0	0	0	0	0	0,25	0,00	0,00	0,10
20	59	TABTENDENZ	0	0	0	0	0	0	0	0	0	0	0,00	0,00	0,00	0,00
21	59	TABDIFFFORM	0	0	0	0	0	0	0	0	0	0	0,00	0,00	0,00	0,00
22	59	TABDIFFFORMVERG	0	0	0	0	0	0	0	0	0	0	0,00	0,00	0,00	0,00
23	43	SIEGDIFF	0	1	0	0	0	0	0	0	0	0	0,25	0,00	0,00	0,10
24	59	REMISDIFF	0	0	0	0	0	0	0	0	0	0	0,00	0,00	0,00	0,00
25	59	NIEDDIFF	0	0	0	0	0	0	0	0	0	0	0,00	0,00	0,00	0,00
26	59	PUNKTEDIFF	0	0	0	0	0	0	0	0	0	0	0,00	0,00	0,00	0,00
27	36	TDDIFF	0	0	0	0	0	1	1	0	0	0	0,00	0,67	0,00	0,20
28	59	TORDIFFFORM	0	0	0	0	0	0	0	0	0	0	0,00	0,00	0,00	0,00
29	59	TFHEIM	0	0	0	0	0	0	0	0	0	0	0,00	0,00	0,00	0,00
30	36	TFGAST	1	0	0	0	1	0	0	0	0	0	0,25	0,33	0,00	0,20
31	59	HEIMSTARK	0	0	0	0	0	0	0	0	0	0	0,00	0,00	0,00	0,00
32	59	GASTSTARK	0	0	0	0	0	0	0	0	0	0	0,00	0,00	0,00	0,00
33	59	SIEGEHEIM	0	0	0	0	0	0	0	0	0	0	0,00	0,00	0,00	0,00
34	59	VERSUS1HEIM	0	0	0	0	0	0	0	0	0	0	0,00	0,00	0,00	0,00
35	59	VERSUS1GAST	0	0	0	0	0	0	0	0	0	0	0,00	0,00	0,00	0,00
36	13	VERSUS1	0	1	3	2	0	1	1	0	0	0	1,50	0,67	0,00	0,80
37	59	VERSUS2HEIM	0	0	0	0	0	0	0	0	0	0	0,00	0,00	0,00	0,00
38	59	VERSUS2GAST	0	0	0	0	0	0	0	0	0	0	0,00	0,00	0,00	0,00
39	22	VERSUS2	0	0	1	0	2	0	0	0	0	1	0,25	0,67	0,00	0,40
40	59	VERSUS3HEIM	0	0	0	0	0	0	0	0	0	0	0,00	0,00	0,00	0,00
41	59	VERSUS3GAST	0	0	0	0	0	0	0	0	0	0	0,00	0,00	0,00	0,00
42	59	VERSUS3	0	0	0	0	0	0	0	0	0	0	0,00	0,00	0,00	0,00
43	59	VERSUS4HEIM	0	0	0	0	0	0	0	0	0	0	0,00	0,00	0,00	0,00
44	59	VERSUS4GAST	0	0	0	0	0	0	0	0	0	0	0,00	0,00	0,00	0,00
45	27	VERSUS4	0	0	0	2	0	0	1	0	0	0	0,50	0,33	0,00	0,30

46	59	VERSUS5HEIM	0	0	0	0	0	0	0	0	0	0	0,00	0,00	0,00	0,00
47	59	VERSUS5GAST	0	0	0	0	0	0	0	0	0	0	0,00	0,00	0,00	0,00
48	43	VERSUS5	0	0	0	0	1	0	0	0	0	0	0,00	0,33	0,00	0,10
49	59	TABGAST	0	0	0	0	0	0	0	0	0	0	0,00	0,00	0,00	0,00
50	59	TABVERGLEICH	0	0	0	0	0	0	0	0	0	0	0,00	0,00	0,00	0,00
51	59	FORM1HEIMHEIM	0	0	0	0	0	0	0	0	0	0	0,00	0,00	0,00	0,00
52	36	FORM1GASTAUSW	0	0	0	0	0	1	0	0	0	1	0,00	0,33	0,00	0,20
53	59	FORM1HEIM	0	0	0	0	0	0	0	0	0	0	0,00	0,00	0,00	0,00
54	27	FORM1GAST	0	0	0	2	0	0	1	0	0	0	0,50	0,33	0,00	0,30
55	43	DIFFFORM1	0	0	0	0	0	0	0	0	0	1	0,00	0,00	0,00	0,10
56	59	FORM2HEIMHEIM	0	0	0	0	0	0	0	0	0	0	0,00	0,00	0,00	0,00
57	10	FORM2GASTAUSW	0	4	0	0	2	0	2	0	1	0	1,00	1,33	0,50	0,90
58	59	FORM2HEIM	0	0	0	0	0	0	0	0	0	0	0,00	0,00	0,00	0,00
59	43	FORM2GAST	0	1	0	0	0	0	0	0	0	0	0,25	0,00	0,00	0,10
60	36	DIFFFORM2	1	0	0	0	0	0	0	1	0	0	0,25	0,00	0,50	0,20
61	59	FORM3HEIMHEIM	0	0	0	0	0	0	0	0	0	0	0,00	0,00	0,00	0,00
62	19	FORM3GASTAUSW	1	2	0	0	1	0	0	0	0	1	0,75	0,33	0,00	0,50
63	59	FORM3HEIM	0	0	0	0	0	0	0	0	0	0	0,00	0,00	0,00	0,00
64	22	FORM3GAST	0	2	0	2	0	0	0	0	0	0	1,00	0,00	0,00	0,40
65	27	DIFFFORM3	0	0	0	2	0	0	0	0	1	0	0,50	0,00	0,50	0,30
66	59	FORM4HEIMHEIM	0	0	0	0	0	0	0	0	0	0	0,00	0,00	0,00	0,00
67	4	FORM4GASTAUSW	0	0	0	0	2	0	2	8	0	9	0,00	1,33	4,00	2,10
68	59	FORM4HEIM	0	0	0	0	0	0	0	0	0	0	0,00	0,00	0,00	0,00
69	22	FORM4GAST	0	1	0	1	0	1	1	0	0	0	0,50	0,67	0,00	0,40
70	27	DIFFFORM4	0	0	0	1	0	1	1	0	0	0	0,25	0,67	0,00	0,30
71	59	FORM5HEIMHEIM	0	0	0	0	0	0	0	0	0	0	0,00	0,00	0,00	0,00
72	15	FORM5GASTAUSW	1	2	0	0	2	0	0	0	2	0	0,75	0,67	1,00	0,70
73	59	FORM5HEIM	0	0	0	0	0	0	0	0	0	0	0,00	0,00	0,00	0,00
74	27	FORM5GAST	0	1	0	2	0	0	0	0	0	0	0,75	0,00	0,00	0,30
75	27	DIFFFORM5	0	0	0	3	0	0	0	0	0	0	0,75	0,00	0,00	0,30
76	59	VIELETORE	0	0	0	0	0	0	0	0	0	0	0,00	0,00	0,00	0,00
77	59	AVG1PLATZHEIM	0	0	0	0	0	0	0	0	0	0	0,00	0,00	0,00	0,00
78	15	AVG1PLATZGAST	0	0	0	0	0	3	2	0	2	0	0,00	1,67	1,00	0,70
79	3	DIFFAVG1	0	1	9	1	3	0	5	8	0	8	2,75	2,67	4,00	3,50
80	59	AVG2PLATZHEIM	0	0	0	0	0	0	0	0	0	0	0,00	0,00	0,00	0,00
81	59	AVG2PLATZGAST	0	0	0	0	0	0	0	0	0	0	0,00	0,00	0,00	0,00
82	10	DIFFAVG2	3	1	0	1	2	0	1	0	0	1	1,25	1,00	0,00	0,90
83	59	AVG3PLATZHEIM	0	0	0	0	0	0	0	0	0	0	0,00	0,00	0,00	0,00
84	19	AVG3PLATZGAST	0	1	1	0	1	1	1	0	0	0	0,50	1,00	0,00	0,50
85	43	DIFFAVG3	0	1	0	0	0	0	0	0	0	0	0,25	0,00	0,00	0,10
86	59	AVG4PLATZHEIM	0	0	0	0	0	0	0	0	0	0	0,00	0,00	0,00	0,00
87	43	AVG4PLATZGAST	0	0	0	1	0	0	0	0	0	0	0,25	0,00	0,00	0,10
88	43	DIFFAVG4	0	1	0	0	0	0	0	0	0	0	0,25	0,00	0,00	0,10
89	59	AVG5PLATZHEIM	0	0	0	0	0	0	0	0	0	0	0,00	0,00	0,00	0,00
90	59	AVG5PLATZGAST	0	0	0	0	0	0	0	0	0	0	0,00	0,00	0,00	0,00
91	8	DIFFAVG5	5	1	0	1	3	0	0	0	0	0	1,75	1,00	0,00	1,00
92	59	HAEUFIGSTEPLAZIERUNGHEIM	0	0	0	0	0	0	0	0	0	0	0,00	0,00	0,00	0,00

93	59	HAEUFIGSTEPLAZIERUNGSGAST	0	0	0	0	0	0	0	0	0	0	0,00	0,00	0,00
94	59	ADDNATIONALTOP9HEIM	0	0	0	0	0	0	0	0	0	0	0,00	0,00	0,00
95	22	ADDNATIONALTOP9GAST	0	0	1	1	0	0	1	0	1	0	0,50	0,33	0,50
96	59	ADDNATIONALLAENDERSP.TOP9H.	0	0	0	0	0	0	0	0	0	0	0,00	0,00	0,00
97	27	ADDNATIONALLAENDERSP.TOP9G.	0	0	0	1	0	0	1	0	0	1	0,25	0,33	0,00
98	59	ADDNATIONALTOP9POSHEIM	0	0	0	0	0	0	0	0	0	0	0,00	0,00	0,00
99	36	ADDNATIONALTOP9POSGAST	0	1	0	0	0	1	0	0	0	0	0,25	0,33	0,00
100	59	ADDNAT.LAENDERSP.TOPPOS9H.	0	0	0	0	0	0	0	0	0	0	0,00	0,00	0,00
101	59	ADDNAT.LAENDERSP.TOPPOS9G.	0	0	0	0	0	0	0	0	0	0	0,00	0,00	0,00
102	59	ADDMITTELNATLSPHEIM	0	0	0	0	0	0	0	0	0	0	0,00	0,00	0,00
103	59	ADDMITTELNATLSPGAST	0	0	0	0	0	0	0	0	0	0	0,00	0,00	0,00
104	59	ADDLSPKTHEIM	0	0	0	0	0	0	0	0	0	0	0,00	0,00	0,00
105	59	ADDLSPKTGAST	0	0	0	0	0	0	0	0	0	0	0,00	0,00	0,00
106	59	ADDGESCHOSSENETOREHEIM	0	0	0	0	0	0	0	0	0	0	0,00	0,00	0,00
107	59	ADDGESCHOSSENETOREGAST	0	0	0	0	0	0	0	0	0	0	0,00	0,00	0,00
108	59	ADDDURCHGELASSENETOREHEIM	0	0	0	0	0	0	0	0	0	0	0,00	0,00	0,00
109	59	ADDDURCHGELASSENETOREGAST	0	0	0	0	0	0	0	0	0	0	0,00	0,00	0,00
110	59	ADDPOSITIONHEIM	0	0	0	0	0	0	0	0	0	0	0,00	0,00	0,00
111	59	ADDPOSITIONGAST	0	0	0	0	0	0	0	0	0	0	0,00	0,00	0,00
112	59	ADDSTADIONGROESSEHEIM	0	0	0	0	0	0	0	0	0	0	0,00	0,00	0,00
113	59	ADDSTADIONGROESSEGAST	0	0	0	0	0	0	0	0	0	0	0,00	0,00	0,00
114	59	ADD10ABSTIEGHEIM	0	0	0	0	0	0	0	0	0	0	0,00	0,00	0,00
115	59	ADD10ABSTIEGGAST	0	0	0	0	0	0	0	0	0	0	0,00	0,00	0,00
116	59	ADDBRAZILHEIM	0	0	0	0	0	0	0	0	0	0	0,00	0,00	0,00
117	59	ADDBRAZILGAST	0	0	0	0	0	0	0	0	0	0	0,00	0,00	0,00
118	59	ADDBRAZILNATHEIM	0	0	0	0	0	0	0	0	0	0	0,00	0,00	0,00
119	59	ADDBRAZILNATGAST	0	0	0	0	0	0	0	0	0	0	0,00	0,00	0,00
120	59	ADDGERMANYNATHEIM	0	0	0	0	0	0	0	0	0	0	0,00	0,00	0,00
121	5	ADDGERMANYNATGAST	2	8	3	0	3	2	0	0	0	1	3,25	1,67	0,00
122	59	ADDAUSTRIAHEIM	0	0	0	0	0	0	0	0	0	0	0,00	0,00	0,00
123	59	ADDAUSTRIAGAST	0	0	0	0	0	0	0	0	0	0	0,00	0,00	0,00
124	59	ADDAUSTRIANATHEIM	0	0	0	0	0	0	0	0	0	0	0,00	0,00	0,00
125	59	ADDAUSTRIANATGAST	0	0	0	0	0	0	0	0	0	0	0,00	0,00	0,00
126	59	ADDEWIGETABELLEHEIM	0	0	0	0	0	0	0	0	0	0	0,00	0,00	0,00
127	59	ADDEWIGETABELLEGAST	0	0	0	0	0	0	0	0	0	0	0,00	0,00	0,00
128	59	ADDUEFA1JAHRHEIM	0	0	0	0	0	0	0	0	0	0	0,00	0,00	0,00
129	43	ADDUEFA1JAHRGAST	0	0	0	0	0	0	1	0	0	0	0,00	0,33	0,00
130	59	ADDUEFA5JAHREHEIM	0	0	0	0	0	0	0	0	0	0	0,00	0,00	0,00
131	43	ADDUEFA5JAHREGAST	0	0	1	0	0	0	0	0	0	0	0,25	0,00	0,00
132	59	ADDPUNKTEHEIMHEIM	0	0	0	0	0	0	0	0	0	0	0,00	0,00	0,00
133	22	ADDPUNKTEGASTGAST	0	0	3	1	0	0	0	0	0	0	1,00	0,00	0,40
134	59	ABSTIEG1HEIM	0	0	0	0	0	0	0	0	0	0	0,00	0,00	0,00
135	59	ABSTIEG1GAST	0	0	0	0	0	0	0	0	0	0	0,00	0,00	0,00
136	43	DIFFABSTIEG1	0	0	0	0	0	0	0	0	1	0	0,00	0,00	0,50
137	59	ABSTIEG2HEIM	0	0	0	0	0	0	0	0	0	0	0,00	0,00	0,00
138	59	ABSTIEG2GAST	0	0	0	0	0	0	0	0	0	0	0,00	0,00	0,00
139	59	DIFFABSTIEG2	0	0	0	0	0	0	0	0	0	0	0,00	0,00	0,00

140	59	ABSTIEG3HEIM	0	0	0	0	0	0	0	0	0	0	0,00	0,00	0,00
141	59	ABSTIEG3GAST	0	0	0	0	0	0	0	0	0	0	0,00	0,00	0,00
142	59	DIFFABSTIEG3	0	0	0	0	0	0	0	0	0	0	0,00	0,00	0,00
143	59	ABSTIEG4HEIM	0	0	0	0	0	0	0	0	0	0	0,00	0,00	0,00
144	59	ABSTIEG4GAST	0	0	0	0	0	0	0	0	0	0	0,00	0,00	0,00
145	43	DIFFABSTIEG4	0	0	0	1	0	0	0	0	0	0	0,25	0,00	0,00
146	59	ABSTIEG5HEIM	0	0	0	0	0	0	0	0	0	0	0,00	0,00	0,00
147	59	ABSTIEG5GAST	0	0	0	0	0	0	0	0	0	0	0,00	0,00	0,00
148	43	DIFFABSTIEG5	0	0	0	0	0	0	0	1	0	0	0,00	0,00	0,50
149	8	DIFFPUNKTEHEIMGAST	0	1	0	4	2	0	0	1	2	0	1,25	0,67	1,50
150	36	DIFFPKTLS	0	1	0	1	0	0	0	0	0	0	0,50	0,00	0,00
151	19	DIFFGSTLS	1	1	0	2	0	1	0	0	0	0	1,00	0,33	0,00
152	43	DIFFKASSTLS	0	1	0	0	0	0	0	0	0	0	0,25	0,00	0,00
153	59	DIFFPOSLS	0	0	0	0	0	0	0	0	0	0	0,00	0,00	0,00
154	59	DIFFEWIGETABELLE	0	0	0	0	0	0	0	0	0	0	0,00	0,00	0,00
155	15	DIFFNAT	0	0	0	2	0	0	1	0	0	4	0,50	0,33	0,00
156	13	DIFNAT2	0	3	2	1	1	0	1	0	0	0	1,50	0,67	0,00
157	6	DIFFNATPOS	6	4	0	0	7	0	0	0	0	0	2,50	2,33	0,00
158	43	DIFNATPOS2	0	0	0	0	0	0	1	0	0	0	0,00	0,33	0,00
159	59	DIFFNATPOSMIT	0	0	0	0	0	0	0	0	0	0	0,00	0,00	0,00
160	59	DIFFSTADGR	0	0	0	0	0	0	0	0	0	0	0,00	0,00	0,00
161	10	DIFF10ABSTIEG	1	5	0	0	1	1	0	0	1	0	1,50	0,67	0,50
162	27	DIFFBRAZIL	0	0	3	0	0	0	0	0	0	0	0,75	0,00	0,00
163	27	DIFBRAZILN	0	0	0	0	0	0	1	1	0	1	0,00	0,33	0,50
164	2	DIFFGERMANYN	2	7	0	1	2	10	4	1	10	0	2,50	5,33	5,50
165	59	DIFFAUSTRIA	0	0	0	0	0	0	0	0	0	0	0,00	0,00	0,00
166	59	DIFFAUSTRIAN	0	0	0	0	0	0	0	0	0	0	0,00	0,00	0,00
167	1	DIFFUEFA1J	0	6	8	8	1	6	7	2	9	2	5,50	4,67	5,50
168	6	DIFFUEFA5J	5	3	2	2	2	1	1	0	0	1	3,00	1,33	0,00
169	59	DIFFHAEUFPLAZ	0	0	0	0	0	0	0	0	0	0	0,00	0,00	0,00

Wrapper mit Klassifiz.: „Naive Bayes“ und Suchmethode: „BestFirst-BackwardElimination“

NR	Platz	Name	24	35	46	57	25	36	47	26	37	27	3SD	4SD	5SD	Ges
1	118	HEIM	9	9	7	9	8	9	7	8	9	8	8,50	8,00	8,50	8,30
2	40	GAST	10	8	8	9	10	10	9	10	10	10	8,75	9,67	10,00	9,40
3	143	POSHEIM	7	8	9	8	5	6	9	7	5	9	8,00	6,67	6,00	7,30
4	167	PLUSMINUSHEIM	4	1	3	6	7	2	4	7	6	5	3,50	4,33	6,50	4,50
5	114	SIEGEHEIM	10	9	7	9	9	9	8	9	9	5	8,75	8,67	9,00	8,40
6	162	REMISHEIM	5	1	3	7	6	4	7	5	9	5	4,00	5,67	7,00	5,20
7	154	NIEDERLAGENHEIM	5	3	7	9	5	6	8	6	5	7	6,00	6,33	5,50	6,10
8	68	PUNKTEHEIM	8	9	10	9	8	9	9	9	10	9	9,00	8,67	9,50	9,00
9	137	TDHEIM	8	7	8	10	6	7	7	7	7	9	8,25	6,67	7,00	7,60
10	40	POSGAST	8	10	9	10	10	9	9	10	10	9	9,25	9,33	10,00	9,40
11	160	PLUSMINUSGAST	9	3	6	7	6	6	5	4	5	3	6,25	5,67	4,50	5,40
12	149	SIEGEGAST	7	6	8	10	10	4	5	8	6	4	7,75	6,33	7,00	6,80
13	164	REMISGAST	4	4	7	10	3	7	2	5	2	2	6,25	4,00	3,50	4,60
14	98	NIEDERLAGENGAST	10	9	7	7	10	10	7	9	8	9	8,25	9,00	8,50	8,60
15	147	PUNKTEGAST	6	6	7	8	7	7	7	8	7	8	6,75	7,00	7,50	7,10
16	63	TDGAST	10	9	8	10	10	10	8	9	9	8	9,25	9,33	9,00	9,10
17	152	MONAT	5	5	8	6	5	8	5	8	10	6	6,00	6,00	9,00	6,60
18	164	JAHRESZEIT	5	4	6	8	2	6	2	6	3	4	5,75	3,33	4,50	4,60
19	46	TABDIFF	10	9	8	9	10	9	10	9	10	9	9,00	9,67	9,50	9,30
20	46	TABTENDENZ	10	8	9	10	10	10	10	7	10	9	9,25	10,00	8,50	9,30
21	150	TABDIFFFORM	8	4	8	4	10	4	7	8	6	8	6,00	7,00	7,00	6,70
22	155	TABDIFFFORMVERG	8	7	5	5	6	6	3	6	5	9	6,25	5,00	5,50	6,00
23	84	SIEGDIFF	9	7	10	9	8	10	8	8	9	10	8,75	8,67	8,50	8,80
24	130	REMISDIFF	8	5	8	7	8	8	8	9	9	9	7,00	8,00	9,00	7,90
25	74	NIEDDIFF	9	9	10	9	8	9	10	9	7	9	9,25	9,00	8,00	8,90
26	108	PUNKTEDIFF	8	5	7	10	8	9	10	9	9	10	7,50	9,00	9,00	8,50
27	68	TDDIFF	10	7	9	9	10	10	8	9	9	9	8,75	9,33	9,00	9,00
28	140	TORDIFFFORM	8	9	8	4	9	6	8	8	7	7	7,25	7,67	7,50	7,40
29	168	TFHEIM	0	0	3	2	1	2	1	2	1	0	1,25	1,33	1,50	1,20
30	160	TFGAST	3	6	7	5	6	5	5	7	5	5	5,25	5,33	6,00	5,40
31	40	HEIMSTARK	10	10	10	9	9	9	9	9	9	10	9,75	9,00	9,00	9,40
32	153	GASTSTARK	1	5	9	6	7	9	9	9	6	4	5,25	8,33	7,50	6,50
33	140	SIEGEHEIM	7	10	10	4	9	9	4	8	6	7	7,75	7,33	7,00	7,40
34	140	VERSUS1HEIM	3	8	10	7	8	8	9	7	7	7	7,00	8,33	7,00	7,40
35	128	VERSUS1GAST	8	7	9	7	7	10	8	9	9	6	7,75	8,33	9,00	8,00
36	130	VERSUS1	8	7	9	8	7	9	6	8	8	9	8,00	7,33	8,00	7,90
37	108	VERSUS2HEIM	7	8	9	9	9	9	10	9	9	6	8,25	9,33	9,00	8,50
38	74	VERSUS2GAST	8	8	10	10	8	10	8	9	9	9	9,00	8,67	9,00	8,90
39	68	VERSUS2	7	10	10	9	8	10	7	10	9	10	9,00	8,33	9,50	9,00
40	118	VERSUS3HEIM	8	9	9	8	9	8	8	8	9	7	8,50	8,33	8,50	8,30
41	22	VERSUS3GAST	10	9	9	10	10	10	10	10	9	10	9,50	10,00	9,50	9,70
42	118	VERSUS3	8	8	8	8	8	8	9	7	9	10	8,00	8,33	8,00	8,30
43	148	VERSUS4HEIM	4	7	7	10	8	7	6	7	7	7	7,00	7,00	7,00	7,00
44	53	VERSUS4GAST	10	8	9	10	10	7	10	10	9	9	9,25	9,00	9,50	9,20
45	136	VERSUS4	10	6	9	9	9	4	6	4	10	10	8,50	6,33	7,00	7,70

46	125	VERSUS5HEIM	7	8	9	10	6	8	8	8	10	7	8,50	7,33	9,00	8,10
47	40	VERSUS5GAST	10	9	9	9	10	9	9	9	10	10	9,25	9,33	9,50	9,40
48	98	VERSUS5	9	8	10	9	8	10	8	9	8	7	9,00	8,67	8,50	8,60
49	84	TABGAST	8	8	9	9	9	8	9	10	9	9	8,50	8,67	9,50	8,80
50	28	TABVERGLEICH	10	9	9	10	9	9	9	10	10	10	9,50	9,33	10,00	9,60
51	145	FORM1HEIMHEIM	7	5	10	9	6	6	9	7	8	5	7,75	7,00	7,50	7,20
52	6	FORM1GASTAUSW	10	9	10	10	10	10	10	10	10	10	9,75	10,00	10,00	9,90
53	130	FORM1HEIM	9	7	9	9	8	8	9	6	7	7	8,50	8,33	6,50	7,90
54	28	FORM1GAST	10	10	10	9	10	9	9	10	10	9	9,75	9,33	10,00	9,60
55	74	DIFFFORM1	9	9	8	10	8	10	8	10	8	9	9,00	8,67	9,00	8,90
56	130	FORM2HEIMHEIM	7	6	7	10	7	9	9	8	8	8	7,50	8,33	8,00	7,90
57	6	FORM2GASTAUSW	10	10	10	10	10	9	10	10	10	10	10,00	9,67	10,00	9,90
58	114	FORM2HEIM	7	6	10	8	6	9	9	10	10	9	7,75	8,00	10,00	8,40
59	1	FORM2GAST	10	10	10	10	10	10	10	10	10	10	10,00	10,00	10,00	10,00
60	53	DIFFFORM2	9	9	9	10	9	10	9	8	9	10	9,25	9,33	8,50	9,20
61	130	FORM3HEIMHEIM	6	8	8	9	9	7	9	9	8	6	7,75	8,33	8,50	7,90
62	1	FORM3GASTAUSW	10	10	10	10	10	10	10	10	10	10	10,00	10,00	10,00	10,00
63	35	FORM3HEIM	10	9	10	10	8	10	9	9	10	10	9,75	9,00	9,50	9,50
64	6	FORM3GAST	9	10	10	10	10	10	10	10	10	10	9,75	10,00	10,00	9,90
65	14	DIFFFORM3	10	10	10	8	10	10	10	10	10	10	9,50	10,00	10,00	9,80
66	98	FORM4HEIMHEIM	8	6	10	9	7	9	8	10	10	9	8,25	8,00	10,00	8,60
67	6	FORM4GASTAUSW	10	10	9	10	10	10	10	10	10	10	9,75	10,00	10,00	9,90
68	74	FORM4HEIM	9	7	10	10	8	7	9	9	10	10	9,00	8,00	9,50	8,90
69	14	FORM4GAST	9	10	10	10	10	10	9	10	10	10	9,75	9,67	10,00	9,80
70	46	DIFFFORM4	10	10	9	10	9	7	9	10	10	9	9,75	8,33	10,00	9,30
71	108	FORM5HEIMHEIM	8	7	10	10	9	6	9	8	10	8	8,75	8,00	9,00	8,50
72	22	FORM5GASTAUSW	10	8	9	10	10	10	10	10	10	10	9,25	10,00	10,00	9,70
73	53	FORM5HEIM	8	10	9	10	10	9	10	10	8	8	9,25	9,67	9,00	9,20
74	1	FORM5GAST	10	10	10	10	10	10	10	10	10	10	10,00	10,00	10,00	10,00
75	68	DIFFFORM5	10	10	7	8	10	10	7	10	9	9	8,75	9,00	9,50	9,00
76	169	VIELETORE	1	1	0	1	1	1	0	0	1	0	0,75	0,67	0,50	0,60
77	123	AVG1PLATZHEIM	8	8	9	10	8	9	9	7	6	8	8,75	8,67	6,50	8,20
78	14	AVG1PLATZGAST	9	10	10	9	10	10	10	10	10	10	9,50	10,00	10,00	9,80
79	98	DIFFAVG1	8	8	10	9	9	8	8	9	9	8	8,75	8,33	9,00	8,60
80	118	AVG2PLATZHEIM	9	7	8	9	8	8	9	8	7	10	8,25	8,33	7,50	8,30
81	6	AVG2PLATZGAST	10	10	10	10	10	10	10	10	9	10	10,00	10,00	9,50	9,90
82	35	DIFFAVG2	8	10	10	9	9	10	10	10	10	9	9,25	9,67	10,00	9,50
83	74	AVG3PLATZHEIM	10	8	10	9	9	8	10	9	9	7	9,25	9,00	9,00	8,90
84	28	AVG3PLATZGAST	10	10	10	9	9	9	10	10	9	10	9,75	9,33	9,50	9,60
85	108	DIFFAVG3	10	8	10	7	8	9	8	8	9	8	8,75	8,33	8,50	8,50
86	63	AVG4PLATZHEIM	10	9	10	9	9	9	8	9	9	9	9,50	8,67	9,00	9,10
87	35	AVG4PLATZGAST	8	9	10	8	10	10	10	10	10	10	8,75	10,00	10,00	9,50
88	128	DIFFAVG4	7	7	10	8	4	8	8	10	8	10	8,00	6,67	9,00	8,00
89	6	AVG5PLATZHEIM	10	10	10	10	10	9	10	10	10	10	10,00	9,67	10,00	9,90
90	1	AVG5PLATZGAST	10	10	10	10	10	10	10	10	10	10	10,00	10,00	10,00	10,00
91	6	DIFFAVG5	10	10	10	10	10	10	10	10	10	9	10,00	10,00	10,00	9,90
92	91	HAEUFIGSTEPLAZIERUNGHEIM	10	7	9	10	10	7	9	7	10	8	9,00	8,67	8,50	8,70

93	74	HAEUFIGSTEPLAZIERUNGGAST	10	10	9	9	7	8	9	10	9	8	9,50	8,00	9,50	8,90
94	91	ADDNATIONALTOP9HEIM	8	8	9	10	9	7	9	9	10	8	8,75	8,33	9,50	8,70
95	22	ADDNATIONALTOP9GAST	9	10	10	9	10	10	10	10	10	9	9,50	10,00	10,00	9,70
96	138	ADDNATIONALLAENDERSP.TOP9H.	8	8	10	9	6	5	9	6	8	6	8,75	6,67	7,00	7,50
97	1	ADDNATIONALLAENDERSP.TOP9G.	10	10	10	10	10	10	10	10	10	10	10,00	10,00	10,00	10,00
98	74	ADDNATIONALTOP9POSHEIM	8	9	9	10	7	10	10	7	9	10	9,00	9,00	8,00	8,90
99	22	ADDNATIONALTOP9POSGAST	10	10	9	9	10	10	10	10	10	9	9,50	10,00	10,00	9,70
100	91	ADDNAT.LAENDERSP.TOPPOS9H.	7	8	10	10	10	8	8	7	10	9	8,75	8,67	8,50	8,70
101	46	ADDNAT.LAENDERSP.TOPPOS9G.	10	9	10	9	8	9	9	10	9	10	9,50	8,67	9,50	9,30
102	53	ADDMITTELNATLSPHEIM	9	10	10	10	10	9	8	7	9	10	9,75	9,00	8,00	9,20
103	14	ADDMITTELNATLSPGAST	9	9	10	10	10	10	10	10	10	10	9,50	10,00	10,00	9,80
104	91	ADDLSPKTHEIM	10	8	8	10	10	5	9	9	8	10	9,00	8,00	8,50	8,70
105	22	ADDLSPKTGAST	9	10	9	9	10	10	10	10	10	10	9,25	10,00	10,00	9,70
106	114	ADDGESCHOSSENETOREHEIM	7	8	9	10	6	9	9	10	9	7	8,50	8,00	9,50	8,40
107	46	ADDGESCHOSSENETOREGAST	9	9	10	8	10	9	9	10	9	10	9,00	9,33	9,50	9,30
108	98	ADDDURCHGELASSENETOREHEIM	9	7	6	10	10	7	9	10	8	10	8,00	8,67	9,00	8,60
109	84	ADDDURCHGELASSENETOREGAST	10	7	10	9	8	8	10	9	9	8	9,00	8,67	9,00	8,80
110	35	ADDPOSITIONHEIM	10	9	9	8	9	10	10	10	10	10	9,00	9,67	10,00	9,50
111	46	ADDPOSITIONGAST	10	9	10	10	8	9	8	10	10	9	9,75	8,33	10,00	9,30
112	150	ADDSTADIONGROESSEHEIM	8	4	8	8	6	7	8	6	6	6	7,00	7,00	6,00	6,70
113	114	ADDSTADIONGROESSEGAST	9	5	10	8	10	7	9	10	9	7	8,00	8,67	9,50	8,40
114	145	ADD10ABSTIEGHEIM	7	8	9	9	7	7	9	7	5	4	8,25	7,67	6,00	7,20
115	53	ADD10ABSTIEGGAST	8	8	10	9	10	10	9	9	10	9	8,75	9,67	9,50	9,20
116	138	ADDBRAZILHEIM	5	7	9	10	5	7	8	10	7	7	7,75	6,67	8,50	7,50
117	84	ADDBRAZILGAST	9	7	10	9	10	9	9	9	9	7	8,75	9,33	9,00	8,80
118	123	ADDBRAZILNATHEIM	10	8	9	7	8	9	8	8	6	9	8,50	8,33	7,00	8,20
119	68	ADDBRAZILNATGAST	10	8	8	9	9	10	10	9	7	10	8,75	9,67	8,00	9,00
120	53	ADDGERMANYNATHEIM	10	9	10	9	9	8	9	9	10	9	9,50	8,67	9,50	9,20
121	6	ADDGERMANYNATGAST	10	10	10	10	10	10	9	10	10	10	10,00	9,67	10,00	9,90
122	158	ADDAUSTRIAHEIM	5	4	9	6	3	8	5	6	4	5	6,00	5,33	5,00	5,50
123	156	ADDAUSTRIAGAST	7	4	8	5	6	5	4	9	5	6	6,00	5,00	7,00	5,90
124	164	ADDAUSTRIANATHEIM	2	4	9	5	2	5	6	3	4	6	5,00	4,33	3,50	4,60
125	163	ADDAUSTRIANATGAST	3	2	5	7	5	5	5	5	6	5	4,25	5,00	5,50	4,80
126	143	ADDEWIGETABELLEHEIM	9	6	6	8	9	7	6	8	9	5	7,25	7,33	8,50	7,30
127	35	ADDEWIGETABELLEGAST	8	9	10	9	10	10	10	10	10	9	9,00	10,00	10,00	9,50
128	84	ADDUEFA1JAHRHEIM	8	9	10	9	8	8	8	9	9	10	9,00	8,00	9,00	8,80
129	22	ADDUEFA1JAHRGAST	9	10	10	10	10	9	9	10	10	10	9,75	9,33	10,00	9,70
130	91	ADDUEFA5JAIREHEIM	9	9	9	10	8	6	10	9	8	9	9,25	8,00	8,50	8,70
131	14	ADDUEFA5JAHREGAST	9	10	10	10	10	10	10	10	10	9	9,75	10,00	10,00	9,80
132	108	ADDPUNKTEHEIMHEIM	7	6	7	10	7	9	10	10	9	10	7,50	8,67	9,50	8,50
133	14	ADDPUNKTEGASTGAST	10	10	10	10	9	10	10	9	10	10	10,00	9,67	9,50	9,80
134	63	ABSTIEG1HEIM	9	10	10	10	10	7	8	9	9	9	9,75	8,33	9,00	9,10
135	98	ABSTIEG1GAST	9	8	8	8	9	9	8	10	10	7	8,25	8,67	10,00	8,60
136	74	DIFFABSTIEG1	9	9	8	10	8	9	9	9	9	9	9,00	8,67	9,00	8,90
137	46	ABSTIEG2HEIM	9	9	9	10	9	8	10	10	10	9	9,25	9,00	10,00	9,30
138	63	ABSTIEG2GAST	8	10	10	9	9	10	7	10	9	9	9,25	8,67	9,50	9,10
139	63	DIFFABSTIEG2	9	9	9	9	7	10	10	10	10	8	9,00	9,00	10,00	9,10

140	53	ABSTIEG3HEIM	9	9	10	10	10	9	7	8	10	10	9,50	8,67	9,00	9,20
141	53	ABSTIEG3GAST	9	10	9	10	9	8	10	9	9	9	9,50	9,00	9,00	9,20
142	28	DIFFABSTIEG3	10	10	8	10	10	9	10	10	10	9	9,50	9,67	10,00	9,60
143	53	ABSTIEG4HEIM	9	10	10	10	9	9	9	8	10	8	9,75	9,00	9,00	9,20
144	40	ABSTIEG4GAST	9	9	9	10	10	9	9	10	10	9	9,25	9,33	10,00	9,40
145	84	DIFFABSTIEG4	10	6	6	10	10	8	10	10	8	10	8,00	9,33	9,00	8,80
146	84	ABSTIEG5HEIM	7	10	10	10	9	8	9	9	8	8	9,25	8,67	8,50	8,80
147	14	ABSTIEG5GAST	9	10	10	10	10	10	10	10	9	10	9,75	10,00	9,50	9,80
148	91	DIFFABSTIEG5	8	8	6	8	10	9	9	9	10	10	7,50	9,33	9,50	8,70
149	98	DIFFPUNKTEHEIMGAST	8	7	9	9	9	6	10	9	10	9	8,25	8,33	9,50	8,60
150	98	DIFFPKTLS	10	9	8	6	9	9	8	9	8	10	8,25	8,67	8,50	8,60
151	98	DIFFGSTLS	9	8	9	5	9	9	9	8	10	10	7,75	9,00	9,00	8,60
152	40	DIFFKASSTLS	10	10	9	9	8	9	10	10	10	9	9,50	9,00	10,00	9,40
153	28	DIFFPOSLS	9	10	10	10	10	10	8	10	9	10	9,75	9,33	9,50	9,60
154	125	DIFFEWIGETABELLE	4	5	10	10	9	9	10	9	7	8	7,25	9,33	8,00	8,10
155	91	DIFFNAT	9	9	8	7	9	10	9	7	10	9	8,25	9,33	8,50	8,70
156	130	DIFNAT2	9	7	10	9	8	7	5	9	8	7	8,75	6,67	8,50	7,90
157	28	DIFFNATPOS	10	10	10	10	10	10	9	9	8	10	10,00	9,67	8,50	9,60
158	28	DIFNATPOS2	8	10	10	10	10	10	10	9	10	9	9,50	10,00	9,50	9,60
159	14	DIFFNATPOSMIT	10	10	10	10	10	8	10	10	10	10	10,00	9,33	10,00	9,80
160	74	DIFFSTADGR	10	7	10	9	10	7	10	9	9	8	9,00	9,00	9,00	8,90
161	74	DIFF10ABSTIEG	8	9	10	8	9	8	9	10	8	10	8,75	8,67	9,00	8,90
162	108	DIFFBRAZIL	9	9	10	7	10	10	7	8	9	6	8,75	9,00	8,50	8,50
163	125	DIFBRAZILN	9	8	9	8	7	9	9	7	9	6	8,50	8,33	8,00	8,10
164	53	DIFFGERMANYN	10	10	8	10	8	10	10	10	8	8	9,50	9,33	9,00	9,20
165	157	DIFFAUSTRIA	6	4	8	5	4	4	8	7	6	4	5,75	5,33	6,50	5,60
166	158	DIFFAUSTRIAN	6	4	9	7	4	6	6	4	5	4	6,50	5,33	4,50	5,50
167	118	DIFFUEFA1J	9	8	10	9	10	5	9	7	8	8	9,00	8,00	7,50	8,30
168	98	DIFFUEFA5J	9	9	9	10	9	7	10	10	7	6	9,25	8,67	8,50	8,60
169	68	DIFFHAEUFPLAZ	10	8	9	10	8	9	10	8	8	10	9,25	9,00	8,00	9,00

Gesamtergebnis Wrapper mit Klassifiz.: „Naive Bayes“ über alle 3 Suchmethoden

NR	Platz	Name	Gesamt	Prozent
1	133	HEIM	2,77	27,67
2	91	GAST	3,13	31,33
3	143	POSHEIM	2,50	25,00
4	167	PLUSMINUSHEIM	1,50	15,00
5	129	SIEGEHEIM	2,80	28,00
6	162	REMISHEIM	1,73	17,33
7	154	NIEDERLAGENHEIM	2,07	20,67
8	109	PUNKTEHEIM	3,00	30,00
9	136	TDHEIM	2,70	27,00
10	62	POSGAST	3,47	34,67
11	161	PLUSMINUSGAST	1,80	18,00
12	136	SIEGEGAST	2,70	27,00
13	164	REMISGAST	1,53	15,33
14	124	NIEDERLAGENGAST	2,87	28,67
15	142	PUNKTEGAST	2,53	25,33
16	59	TDGAST	3,53	35,33
17	152	MONAT	2,20	22,00
18	164	JAHRESZEIT	1,53	15,33
19	64	TABDIFF	3,43	34,33
20	96	TABTENDENZ	3,10	31,00
21	151	TABDIFFFORM	2,23	22,33
22	155	TABDIFFFORMVERG	2,00	20,00
23	77	SIEGDIFF	3,27	32,67
24	139	REMISDIFF	2,63	26,33
25	88	NIEDDIFF	3,17	31,67
26	70	PUNKTEDIFF	3,33	33,33
27	24	TDDIFF	4,23	42,33
28	145	TORDIFFFORM	2,47	24,67
29	168	TFHEIM	1,27	12,67
30	157	TFGAST	1,93	19,33
31	85	HEIMSTARK	3,20	32,00
32	153	GASTSTARK	2,17	21,67
33	145	SIEGEHEIM	2,47	24,67
34	124	VERSUS1HEIM	2,87	28,67
35	91	VERSUS1GAST	3,13	31,33
36	77	VERSUS1	3,27	32,67
37	113	VERSUS2HEIM	2,97	29,67
38	68	VERSUS2GAST	3,37	33,67
39	64	VERSUS2	3,43	34,33
40	133	VERSUS3HEIM	2,77	27,67
41	77	VERSUS3GAST	3,27	32,67
42	113	VERSUS3	2,97	29,67
43	149	VERSUS4HEIM	2,33	23,33
44	88	VERSUS4GAST	3,17	31,67
45	123	VERSUS4	2,90	29,00

46	136	VERSUS5HEIM	2,70	27,00
47	85	VERSUS5GAST	3,20	32,00
48	96	VERSUS5	3,10	31,00
49	116	TABGAST	2,93	29,33
50	70	TABVERGLEICH	3,33	33,33
51	147	FORM1HEIMHEIM	2,40	24,00
52	26	FORM1GASTAUSW	4,17	41,67
53	116	FORM1HEIM	2,93	29,33
54	33	FORM1GAST	4,07	40,67
55	33	DIFFFORM1	4,07	40,67
56	139	FORM2HEIMHEIM	2,63	26,33
57	17	FORM2GASTAUSW	4,40	44,00
58	96	FORM2HEIM	3,10	31,00
59	30	FORM2GAST	4,10	41,00
60	11	DIFFFORM2	4,57	45,67
61	139	FORM3HEIMHEIM	2,63	26,33
62	28	FORM3GASTAUSW	4,13	41,33
63	73	FORM3HEIM	3,30	33,00
64	38	FORM3GAST	4,00	40,00
65	9	DIFFFORM3	4,80	48,00
66	124	FORM4HEIMHEIM	2,87	28,67
67	5	FORM4GASTAUSW	5,07	50,67
68	104	FORM4HEIM	3,03	30,33
69	38	FORM4GAST	4,00	40,00
70	11	DIFFFORM4	4,57	45,67
71	128	FORM5HEIMHEIM	2,83	28,33
72	20	FORM5GASTAUSW	4,30	43,00
73	61	FORM5HEIM	3,50	35,00
74	26	FORM5GAST	4,17	41,67
75	11	DIFFFORM5	4,57	45,67
76	169	VIELETORE	0,20	2,00
77	129	AVG1PLATZHEIM	2,80	28,00
78	15	AVG1PLATZGAST	4,50	45,00
79	1	DIFFAVG1	5,57	55,67
80	84	AVG2PLATZHEIM	3,23	32,33
81	47	AVG2PLATZGAST	3,83	38,33
82	8	DIFFAVG2	4,83	48,33
83	66	AVG3PLATZHEIM	3,40	34,00
84	30	AVG3PLATZGAST	4,10	41,00
85	38	DIFFAVG3	4,00	40,00
86	91	AVG4PLATZHEIM	3,13	31,33
87	51	AVG4PLATZGAST	3,77	37,67
88	49	DIFFAVG4	3,80	38,00
89	51	AVG5PLATZHEIM	3,77	37,67
90	46	AVG5PLATZGAST	3,87	38,67
91	7	DIFFAVG5	4,97	49,67
92	116	HAEUFIGSTEPLAZIERUNGHEIM	2,93	29,33

93	104	HAEUFIGSTEPLAZIERUNGGAST	3,03	30,33
94	109	ADDNATIONALTOP9HEIM	3,00	30,00
95	38	ADDNATIONALTOP9GAST	4,00	40,00
96	73	ADDNATIONALLAENDERSP.TOP9H.	3,30	33,00
97	59	ADDNATIONALLAENDERSP.TOP9G.	3,53	35,33
98	104	ADDNATIONALTOP9POSHEIM	3,03	30,33
99	44	ADDNATIONALTOP9POSGAST	3,90	39,00
100	104	ADDNAT.LAENDERSP.TOPPOS9H.	3,03	30,33
101	96	ADDNAT.LAENDERSP.TOPPOS9G.	3,10	31,00
102	88	ADDMITTELNATLSPHEIM	3,17	31,67
103	70	ADDMITTELNATLSPGAST	3,33	33,33
104	77	ADDLSPKTHEIM	3,27	32,67
105	53	ADDLSPKTGAST	3,73	37,33
106	68	ADDGESCHOSSENETOREHEIM	3,37	33,67
107	73	ADDGESCHOSSENETOREGAST	3,30	33,00
108	124	ADDDURCHGELASSENETOREHEIM	2,87	28,67
109	116	ADDDURCHGELASSENETOREGAST	2,93	29,33
110	53	ADDPOSITIONHEIM	3,73	37,33
111	77	ADDPOSITIONGAST	3,27	32,67
112	150	ADDSTADIONGROESSEHEIM	2,30	23,00
113	129	ADDSTADIONGROESSEGAST	2,80	28,00
114	147	ADD10ABSTIEGHEIM	2,40	24,00
115	77	ADD10ABSTIEGGAST	3,27	32,67
116	143	ADDBRAZILHEIM	2,50	25,00
117	116	ADDBRAZILGAST	2,93	29,33
118	135	ADDBRAZILNATHEIM	2,73	27,33
119	109	ADDBRAZILNATGAST	3,00	30,00
120	57	ADDGERMANYNATHEIM	3,70	37,00
121	10	ADDGERMANYNATGAST	4,73	47,33
122	159	ADDAUSTRIAHEIM	1,83	18,33
123	156	ADDAUSTRIAGAST	1,97	19,67
124	164	ADDAUSTRIANATHEIM	1,53	15,33
125	163	ADDAUSTRIANATGAST	1,60	16,00
126	109	ADDEWIGETABELLEHEIM	3,00	30,00
127	33	ADDEWIGETABELLEGAST	4,07	40,67
128	91	ADDUEFA1JAHRHEIM	3,13	31,33
129	22	ADDUEFA1JAHRGAST	4,27	42,67
130	102	ADDUEFA5JAHREHEIM	3,07	30,67
131	44	ADDUEFA5JAHREGAST	3,90	39,00
132	77	ADDPUNKTEHEIMHEIM	3,27	32,67
133	43	ADDPUNKTEGASTGAST	3,93	39,33
134	104	ABSTIEG1HEIM	3,03	30,33
135	38	ABSTIEG1GAST	4,00	40,00
136	57	DIFFABSTIEG1	3,70	37,00
137	96	ABSTIEG2HEIM	3,10	31,00
138	24	ABSTIEG2GAST	4,23	42,33
139	47	DIFFABSTIEG2	3,83	38,33

140	96	ABSTIEG3HEIM	3,10	31,00
141	22	ABSTIEG3GAST	4,27	42,67
142	30	DIFFABSTIEG3	4,10	41,00
143	91	ABSTIEG4HEIM	3,13	31,33
144	19	ABSTIEG4GAST	4,33	43,33
145	62	DIFFABSTIEG4	3,47	34,67
146	116	ABSTIEG5HEIM	2,93	29,33
147	20	ABSTIEG5GAST	4,30	43,00
148	85	DIFFABSTIEG5	3,20	32,00
149	14	DIFFPUNKTEHEIMGAST	4,53	45,33
150	49	DIFFPKTLS	3,80	38,00
151	53	DIFFGSTLS	3,73	37,33
152	53	DIFFKASSTLS	3,73	37,33
153	28	DIFFPOSLS	4,13	41,33
154	102	DIFFEWIGETABELLE	3,07	30,67
155	16	DIFFNAT	4,43	44,33
156	37	DIFNAT2	4,03	40,33
157	6	DIFFNATPOS	5,03	50,33
158	33	DIFNATPOS2	4,07	40,67
159	18	DIFFNATPOSMIT	4,37	43,67
160	113	DIFFSTADGR	2,97	29,67
161	73	DIFF10ABSTIEG	3,30	33,00
162	116	DIFFBRAZIL	2,93	29,33
163	129	DIFFBRAZILN	2,80	28,00
164	2	DIFFGERMANYN	5,33	53,33
165	158	DIFFAUSTRIA	1,87	18,67
166	159	DIFFAUSTRIAN	1,83	18,33
167	2	DIFFUEFA1J	5,33	53,33
168	4	DIFFUEFA5J	5,10	51,00
169	66	DIFFHAEUFPLAZ	3,40	34,00

Wrapper mit Klassifizierer: „JRip“ und Suchmethode: „RankSearch“

NR	Platz	Name	24	35	46	57	25	36	47	26	37	27	3SD	4SD	5SD	Ges
1	139	HEIM	3	1	2		1			2			1,50	0,33	1,00	0,90
2	134	GAST	3	1	2		1	1		2			1,50	0,67	1,00	1,00
3	105	POSHEIM	3	1	5	4	1	3			2		3,25	1,33	1,00	1,90
4	149	PLUSMINUSHEIM	2	1	2					2			1,25	0,00	1,00	0,70
5	144	SIEGEHEIM	2	1	2			1		2			1,25	0,33	1,00	0,80
6	149	REMISHEIM	2	1	2					2			1,25	0,00	1,00	0,70
7	139	NIEDERLAGENHEIM	3	1	2				1	2			1,50	0,33	1,00	0,90
8	139	PUNKTEHEIM	3	1	2			1		2			1,50	0,33	1,00	0,90
9	69	TDHEIM	3	3	7	6	1	6	4		4		4,75	3,67	2,00	3,40
10	97	POSGAST	3	5	2	3	2	3	1	2			3,25	2,00	1,00	2,10
11	139	PLUSMINUSGAST	3	1	2		1			2			1,50	0,33	1,00	0,90
12	97	SIEGEGAST	3	4	2	1	3	3	2		3		2,50	2,67	1,50	2,10
13	139	REMISGAST	3	1	2		1			2			1,50	0,33	1,00	0,90
14	144	NIEDERLAGENGAST	2	1	2		1			2			1,25	0,33	1,00	0,80
15	117	PUNKTEGAST	2	4	2		1	2	2		3		2,00	1,67	1,50	1,60
16	93	TDGAST	3	3	3	5	1	3	1	4			3,50	1,67	2,00	2,30
17	153	MONAT	2	1	1					2			1,00	0,00	1,00	0,60
18	153	JAHRESZEIT	2	1	1					2			1,00	0,00	1,00	0,60
19	56	TABDIFF	6	7	6	5	3	5	2	1	4		6,00	3,33	2,50	3,90
20	130	TABTENDENZ	3	2	2		1	1		2			1,75	0,67	1,00	1,10
21	160	TABDIFFFORM	2	1						1			0,75	0,00	0,50	0,40
22	160	TABDIFFFORMVERG	3							1			0,75	0,00	0,50	0,40
23	86	SIEGDIFF	4	4	5	2	2	5		1	3		3,75	2,33	2,00	2,60
24	166	REMISDIFF	2							1			0,50	0,00	0,50	0,30
25	83	NIEDDIFF	4	5	4	3	1	4	2		4		4,00	2,33	2,00	2,70
26	50	PUNKTEDIFF	3	8	6	6	1	7	3	1	7		5,75	3,67	4,00	4,20
27	27	TDDIFF	10	8	8	5	5	6	5	8	5	1	7,75	5,33	6,50	6,10
28	130	TORDIFFFORM	2	1	2		1	3		2			1,25	1,33	1,00	1,10
29	32	TFHEIM	2	7	10	7	3	6	9	4	7	3	6,50	6,00	5,50	5,80
30	134	TFGAST	1	2	1	3	1			2			1,75	0,33	1,00	1,00
31	75	HEIMSTARK	2	3	5	5	1	3	5	1	5	1	3,75	3,00	3,00	3,10
32	166	GASTSTARK	1	1							1		0,50	0,00	0,50	0,30
33	160	SIEGEHEIM	1	1	1						1		0,75	0,00	0,50	0,40
34	93	VERSUS1HEIM	5	7	1		8			2			3,25	2,67	1,00	2,30
35	93	VERSUS1GAST	10	1	3	1	4	2		2			3,75	2,00	1,00	2,30
36	73	VERSUS1	3	5	4	6	1	5	3		5		4,50	3,00	2,50	3,20
37	144	VERSUS2HEIM	3	1			2			2			1,00	0,67	1,00	0,80
38	97	VERSUS2GAST	10	2	2		3	2		2			3,50	1,67	1,00	2,10
39	83	VERSUS2	3	4	5	4	1	5	1		4		4,00	2,33	2,00	2,70
40	166	VERSUS3HEIM			1					2			0,25	0,00	1,00	0,30
41	118	VERSUS3GAST	5	2	2		2	2		2			2,25	1,33	1,00	1,50
42	91	VERSUS3	3	4	4	1	1	5		2	3	1	3,00	2,00	2,50	2,40
43	166	VERSUS4HEIM			1					2			0,25	0,00	1,00	0,30
44	122	VERSUS4GAST	4	2	2		2	1		2			2,00	1,00	1,00	1,30
45	105	VERSUS4	3	4	2	1	1	4		1	2	1	2,50	1,67	1,50	1,90

46	158	VERSUS5HEIM	2	1			2	0,75	0,00	1,00	0,50					
47	120	VERSUS5GAST	4	2	2	2	2	2,00	1,33	1,00	1,40					
48	86	VERSUS5	3	4	4	2	1	5	2	3,25	2,00	3,00	2,60			
49	149	TABGAST	2	1	1	1			2	1,00	0,33	1,00	0,70			
50	111	TABVERGLEICH	3	4	2	1	1	3	3	2,50	1,33	1,50	1,70			
51	144	FORM1HEIMHEIM	2			2	1	1	2	1,00	0,67	1,00	0,80			
52	50	FORM1GASTAUSW	5	7	4	2	5	6	4	4,50	3,67	4,50	4,20			
53	50	FORM1HEIM	4	3	5	6	1	5	8	4,50	4,67	4,00	4,20			
54	48	FORM1GAST	4	8	4	2	5	5	1	5	4,50	3,67	5,00	4,30		
55	6	DIFFFORM1	10	10	10	6	8	7	8	9	9,00	7,67	7,50	8,20		
56	134	FORM2HEIMHEIM	2	1	2	1	1	1		2	1,50	0,67	1,00	1,00		
57	47	FORM2GASTAUSW	5	5	6	5	4	7		4	5	3,25	3,67	4,50	4,40	
58	59	FORM2HEIM	4	4	7	8	1	3	4	1	5	1	5,75	2,67	3,00	3,80
59	55	FORM2GAST	4	9	4	2	5	5	1	4	4	2	4,75	3,67	4,00	4,00
60	4	DIFFFORM2	9	10	10	7	7	9	8	9	7	8	9,00	8,00	8,00	8,40
61	127	FORM3HEIMHEIM	2	2	1		1	3	1		2		1,25	1,67	1,00	1,20
62	65	FORM3GASTAUSW	3	5	7	5	3	4	1	5	2		5,00	2,67	3,50	3,50
63	103	FORM3HEIM	4	1	2	5	1	3	2		2		3,00	2,00	1,00	2,00
64	59	FORM3GAST	3	5	5	5	4	5	2	3	4	2	4,50	3,67	3,50	3,80
65	1	DIFFFORM3	9	10	10	8	10	10	7	10	7	6	9,25	9,00	8,50	8,70
66	122	FORM4HEIMHEIM	2	3	1	1		2	1		3		1,75	1,00	1,50	1,30
67	27	FORM4GASTAUSW	4	5	8	7	7	8	3	8	6	5	6,00	6,00	7,00	6,10
68	83	FORM4HEIM	5	4	4	4	1	4	2	1	2		4,25	2,33	1,50	2,70
69	56	FORM4GAST	6	5	7	5	3	5	1	3	3	1	5,75	3,00	3,00	3,90
70	2	DIFFFORM4	6	10	10	7	8	10	10	10	10	4	8,25	9,33	10,00	8,50
71	134	FORM5HEIMHEIM	2	1	2	1	1	1			2		1,50	0,67	1,00	1,00
72	42	FORM5GASTAUSW	4	6	5	4	5	7	1	6	5	3	4,75	4,33	5,50	4,60
73	59	FORM5HEIM	3	4	5	6	1	5	4	1	7	2	4,50	3,33	4,00	3,80
74	39	FORM5GAST	5	9	6	5	7	5	1	4	6	2	6,25	4,33	5,00	5,00
75	8	DIFFFORM5	10	9	10	8	7	10	8	7	7	4	9,25	8,33	7,00	8,00
76	153	VIELETORE	2	1	1						2		1,00	0,00	1,00	0,60
77	69	AVG1PLATZHEIM	3	2	6	6	1	3	7	1	4	1	4,25	3,67	2,50	3,40
78	30	AVG1PLATZGAST	3	10	8	6	4	9	6	5	6	3	6,75	6,33	5,50	6,00
79	2	DIFFAVG1	10	10	10	8	9	8	8	8	9	5	9,50	8,33	8,50	8,50
80	71	AVG2PLATZHEIM	5	8	5	3	7	2	1		2		5,25	3,33	1,00	3,30
81	86	AVG2PLATZGAST	3	6	4	1	2	3	1	3	3		3,50	2,00	3,00	2,60
82	8	DIFFAVG2	10	6	8	7	10	7	6	10	7	9	7,75	7,67	8,50	8,00
83	41	AVG3PLATZHEIM	7	7	3	7	6	5	7		5		6,00	6,00	2,50	4,70
84	35	AVG3PLATZGAST	5	8	7	5	3	8	3	4	6	3	6,25	4,67	5,00	5,20
85	16	DIFFAVG3	8	10	9	7	7	7	9	5	8	5	8,50	7,67	6,50	7,50
86	105	AVG4PLATZHEIM	3	3	3	2	1	2	2	1	2		2,75	1,67	1,50	1,90
87	48	AVG4PLATZGAST	3	7	8	5	1	7	5	2	5		5,75	4,33	3,50	4,30
88	19	DIFFAVG4	10	7	8	7	8	7	9	6	7	3	8,00	8,00	6,50	7,20
89	71	AVG5PLATZHEIM	5	8	5	3	7	2	1		2		5,25	3,33	1,00	3,30
90	89	AVG5PLATZGAST	3	5	4	1	2	3	1	3	3		3,25	2,00	3,00	2,50
91	10	DIFFAVG5	10	6	8	7	10	7	6	10	6	9	7,75	7,67	8,00	7,90
92	111	HAEUFIGSTEPLAZIERUNGHEIM	2	2	3	3	1	2	2		2		2,50	1,67	1,00	1,70

93	111	HAEUFIGSTEPLAZIERUNGGAST	3	4	2	2	1	2	1	2	2,75	1,33	1,00	1,70		
94	103	ADDNATIONALTOP9HEIM	3	3	3	3	1	3	1	1	2	3,00	1,67	1,50	2,00	
95	63	ADDNATIONALTOP9GAST	5	7	3		3	8	1	4	5	1	3,75	4,00	4,50	3,70
96	35	ADDNATIONALLAENDERSP.TOP9H.	10	5	5	3	10	3	2	6	3	5	5,75	5,00	4,50	5,20
97	111	ADDNATIONALLAENDERSP.TOP9G.	3	4	2	3	1	2			2		3,00	1,00	1,00	1,70
98	105	ADDNATIONALTOP9POSHEIM	3	3	3	1	2	3	1	1	2		2,50	2,00	1,50	1,90
99	56	ADDNATIONALTOP9POSGAST	6	7	3		4	8	1	4	4	2	4,00	4,33	4,00	3,90
100	77	ADDNAT.LAENDERSP.TOPPOS9H.	10	3	4	2	4	3	1	1	2		4,75	2,67	1,50	3,00
101	111	ADDNAT.LAENDERSP.TOPPOS9G.	3	4	2	2	1	2	1		2		2,75	1,33	1,00	1,70
102	81	ADDMITTELNATLSPHEIM	6	5	5	3	2	2	3		2		4,75	2,33	1,00	2,80
103	105	ADDMITTELNATLSPGAST	3	4	2	3	2	2	1		2		3,00	1,67	1,00	1,90
104	73	ADDLSPKTHEIM	3	7	5	4	1	2	6		4		4,75	3,00	2,00	3,20
105	64	ADDLSPKTGAST	3	5	4	5	2	5	2	4	4	2	4,25	3,00	4,00	3,60
106	45	ADDGESCHOSSENTOREHEIM	2	8	9	4	1	8	8		5		5,75	5,67	2,50	4,50
107	118	ADDGESCHOSSENTOREGAST	3	4	1	2	1	2			2		2,50	1,00	1,00	1,50
108	122	ADD DURCHGELASSEN TOREHEIM	3	1	3		1	3			2		1,75	1,33	1,00	1,30
109	127	ADD DURCHGELASSEN TOREGAST	2	2	2	2	1	1			2		2,00	0,67	1,00	1,20
110	40	ADDPOSITIONHEIM	3	5	8	7	3	7	8	1	6	1	5,75	6,00	3,50	4,90
111	105	ADDPOSITIONGAST	3	4	3	2	1	3	1		2		3,00	1,67	1,00	1,90
112	130	ADDSTADIONGROESSEHEIM	2	1	2	2	1		1		2		1,75	0,67	1,00	1,10
113	130	ADDSTADIONGROESSEGAST	3	1	2		1	2			2		1,50	1,00	1,00	1,10
114	144	ADD10ABSTIEGHEIM	2	1	2		1				2		1,25	0,33	1,00	0,80
115	97	ADD10ABSTIEGGAST	3	4	3	1	3	3		1	2	1	2,75	2,00	1,50	2,10
116	153	ADDBRAZILHEIM	1	1	2						2		1,00	0,00	1,00	0,60
117	120	ADDBRAZILGAST	3	3	3		1	2			2		2,25	1,00	1,00	1,40
118	134	ADDBRAZILNATHEIM	3	2	2		1				2		1,75	0,33	1,00	1,00
119	127	ADDBRAZILNATGAST	3	2	2		1	2			2		1,75	1,00	1,00	1,20
120	59	ADDGERMANYNATHEIM	3	8	10	2	1	8	3		3		5,75	4,00	1,50	3,80
121	42	ADDGERMANYNATGAST	7	8	4	1	5	8	1	4	5	3	5,00	4,67	4,50	4,60
122	160	ADDAUSTRIAHEIM	1	1							2		0,50	0,00	1,00	0,40
123	160	ADDAUSTRIAGAST	1	1							2		0,50	0,00	1,00	0,40
124	149	ADDAUSTRIANATHEIM	2	1	1		1				2		1,00	0,33	1,00	0,70
125	153	ADDAUSTRIANATGAST	1	1	1		1				2		0,75	0,33	1,00	0,60
126	65	ADDEWIGETABELLEHEIM	4	5	5	3	3	8	1	2	4		4,25	4,00	3,00	3,50
127	32	ADDEWIGETABELLEGAST	3	9	8	7	4	6	7	2	9	3	6,75	5,67	5,50	5,80
128	96	ADDUEFA1JAHRHEIM	3	2	3	9	1	2			2		4,25	1,00	1,00	2,20
129	24	ADDUEFA1JAHRGAST	6	9	5	6	6	7	6	5	7	5	6,50	6,33	6,00	6,20
130	79	ADDUEFA5JAHREHEIM	3	5	4	6	1	3	4	1	2		4,50	2,67	1,50	2,90
131	65	ADDUEFA5JAHREGAST	3	6	6	4	2	6	1	3	4		4,75	3,00	3,50	3,50
132	91	ADDPUNKTEHEIMHEIM	2	7	4	4		4	1		2		4,25	1,67	1,00	2,40
133	53	ADDPUNKTEGASTGAST	3	7	8	5	2	7	2	4	3		5,75	3,67	3,50	4,10
134	89	ABSTIEG1HEIM	4	4	4	2	2	4	1	1	3		3,50	2,33	2,00	2,50
135	11	ABSTIEG1GAST	5	10	9	6	9	9	6	8	9	7	7,50	8,00	8,50	7,80
136	38	DIFFABSTIEG1	4	4	7	7	3	6	6	4	6	4	5,50	5,00	5,00	5,10
137	81	ABSTIEG2HEIM	4	4	4	2	2	4	2	1	4	1	3,50	2,67	2,50	2,80
138	11	ABSTIEG2GAST	5	10	9	7	9	9	6	8	8	7	7,75	8,00	8,00	7,80
139	35	DIFFABSTIEG2	4	5	7	7	4	7	5	5	6	2	5,75	5,33	5,50	5,20

140	79	ABSTIEG3HEIM
141	16	ABSTIEG3GAST
142	24	DIFFABSTIEG3
143	77	ABSTIEG4HEIM
144	20	ABSTIEG4GAST
145	45	DIFFABSTIEG4
146	97	ABSTIEG5HEIM
147	24	ABSTIEG5GAST
148	65	DIFFABSTIEG5
149	7	DIFFPUNKTEHEIMGAST
150	27	DIFFPKTLS
151	34	DIFFGSTLS
152	42	DIFFKASSTLS
153	23	DIFFPOSLS
154	75	DIFFEWIGETABELLE
155	13	DIFFNAT
156	18	DIFNAT2
157	14	DIFFNATPOS
158	30	DIFNATPOS2
159	14	DIFFNATPOSMIT
160	122	DIFFSTADGR
161	97	DIFF10ABSTIEG
162	122	DIFFBRAZIL
163	111	DIFBRAZILN
164	21	DIFFGERMANYN
165	158	DIFFAUSTRIA
166	160	DIFFAUSTRIAN
167	21	DIFFUEFA1J
168	5	DIFFUEFA5J
169	53	DIFFHAEUFPLAZ

4	4	4	2	2	4	2	1	4	2	3,50	2,67	2,50	2,90
5	10	9	7	8	9	7	6	9	5	7,75	8,00	7,50	7,50
5	6	9	7	5	8	6	6	6	4	6,75	6,33	6,00	6,20
4	4	4	2	3	4	2	1	4	2	3,50	3,00	2,50	3,00
3	10	9	7	7	9	6	7	7	5	7,25	7,33	7,00	7,00
4	5	7	6	3	6	4	4	5	1	5,50	4,33	4,50	4,50
3	2	3	2	2	3		1	3	2	2,50	1,67	2,00	2,10
3	9	8	7	5	8	5	6	7	4	6,75	6,00	6,50	6,20
4	4	4	5	3	5	2	2	5	1	4,25	3,33	3,50	3,50
5	9	9	10	8	9	8	9	9	5	8,25	8,33	9,00	8,10
3	9	10	7	5	7	8	4	5	3	7,25	6,67	4,50	6,10
6	9	4	6	7	7	4	3	5	3	6,25	6,00	4,00	5,40
7	8	9	5	3	6	3	2	3		7,25	4,00	2,50	4,60
3	8	9	8	3	8	9	5	7	3	7,00	6,67	6,00	6,30
5	6	5	4	1	3	1	2	3	1	5,00	1,67	2,50	3,10
10	8	10	7	7	9	8	6	6	6	8,75	8,00	6,00	7,70
6	7	10	7	6	7	9	5	10	6	7,50	7,33	7,50	7,30
9	9	9	7	6	9	10	5	7	5	8,50	8,33	6,00	7,60
8	7	8	7	4	5	9	3	6	3	7,50	6,00	4,50	6,00
9	10	9	7	9	7	7	7	7	4	8,75	7,67	7,00	7,60
4	1	4		1	1				2	2,25	0,67	1,00	1,30
3	4	3	2	2	2		1	3	1	3,00	1,33	2,00	2,10
3	3	2		1	2				2	2,00	1,00	1,00	1,30
5	4	3		1	2				2	3,00	1,00	1,00	1,70
9	8	9	6	6	8	4	6	8	5	8,00	6,00	7,00	6,90
1		2							2	0,75	0,00	1,00	0,50
1		1							2	0,50	0,00	1,00	0,40
6	9	7	9	10	7	6	6	7	2	7,75	7,67	6,50	6,90
10	9	10	7	9	8	8	9	6	7	9,00	8,33	7,50	8,30
3	6	6	6	3	6	5	2	3	1	5,25	4,67	2,50	4,10

Wrapper mit Klassifizierer: „JRip“ und Suchmethode: „LinearForwardSelection“

NR	Platz	Name	24	35	46	57	25	36	47	26	37	27	3SD	4SD	5SD	Ges	
1	59	HEIM		1	1								0,50	0,00	0,00	0,20	
2	8	GAST			1	1	1	2	2	1	2	3	0,75	1,67	2,50	1,30	
3	103	POSHEIM											0,00	0,00	0,00	0,00	
4	103	PLUSMINUSHEIM											0,00	0,00	0,00	0,00	
5	103	SIEGEHEIM											0,00	0,00	0,00	0,00	
6	74	REMISHEIM							1				0,25	0,00	0,00	0,10	
7	103	NIEDERLAGENHEIM											0,00	0,00	0,00	0,00	
8	103	PUNKTEHEIM											0,00	0,00	0,00	0,00	
9	103	TDHEIM											0,00	0,00	0,00	0,00	
10	36	POSGAST					1	1		1	1		0,50	0,33	0,50	0,40	
11	103	PLUSMINUSGAST											0,00	0,00	0,00	0,00	
12	27	SIEGEGAST					1	2		1		1	0,75	0,33	0,50	0,60	
13	103	REMISGAST											0,00	0,00	0,00	0,00	
14	103	NIEDERLAGENGAST											0,00	0,00	0,00	0,00	
15	16	PUNKTEGAST		1	1	1	3	1		1			1,50	0,67	0,00	0,80	
16	14	TDGAST		2		1	2		1	2	1		1,25	1,00	0,50	0,90	
17	74	MONAT					1						0,25	0,00	0,00	0,10	
18	103	JAHRESZEIT											0,00	0,00	0,00	0,00	
19	5	TABDIFF					2	6		1	2	1	5	2,00	1,00	3,00	1,70
20	103	TABTENDENZ											0,00	0,00	0,00	0,00	
21	103	TABDIFFFORM											0,00	0,00	0,00	0,00	
22	103	TABDIFFFORMVERG											0,00	0,00	0,00	0,00	
23	74	SIEGDIFF					1						0,25	0,00	0,00	0,10	
24	103	REMISDIFF											0,00	0,00	0,00	0,00	
25	103	NIEDDIFF											0,00	0,00	0,00	0,00	
26	19	PUNKTEDIFF					3		2		1	1	0,75	0,67	0,50	0,70	
27	10	TDDIFF					3		1	2	1	3	1	0,75	1,00	2,00	1,10
28	103	TORDIFFFORM											0,00	0,00	0,00	0,00	
29	103	TFHEIM											0,00	0,00	0,00	0,00	
30	103	TFGAST											0,00	0,00	0,00	0,00	
31	103	HEIMSTARK											0,00	0,00	0,00	0,00	
32	103	GASTSTARK											0,00	0,00	0,00	0,00	
33	103	SIEGEHEIM											0,00	0,00	0,00	0,00	
34	36	VERSUS1HEIM					1		2		1		0,25	0,67	0,50	0,40	
35	74	VERSUS1GAST					1						0,25	0,00	0,00	0,10	
36	36	VERSUS1					1	1	1	1			0,50	0,67	0,00	0,40	
37	74	VERSUS2HEIM					1						0,25	0,00	0,00	0,10	
38	74	VERSUS2GAST						1					0,25	0,00	0,00	0,10	
39	74	VERSUS2					1						0,25	0,00	0,00	0,10	
40	74	VERSUS3HEIM					1						0,25	0,00	0,00	0,10	
41	36	VERSUS3GAST						3			1		0,75	0,00	0,50	0,40	
42	103	VERSUS3											0,00	0,00	0,00	0,00	
43	103	VERSUS4HEIM											0,00	0,00	0,00	0,00	
44	59	VERSUS4GAST								1	1		0,00	0,67	0,00	0,20	
45	36	VERSUS4					1		1	1	1		0,50	0,33	0,50	0,40	

46	103	VERSUS5HEIM								0,00	0,00	0,00	0,00
47	74	VERSUS5GAST		1						0,25	0,00	0,00	0,10
48	53	VERSUS5	1		1				1	0,50	0,00	0,50	0,30
49	103	TABGAST								0,00	0,00	0,00	0,00
50	103	TABVERGLEICH								0,00	0,00	0,00	0,00
51	103	FORM1HEIMHEIM								0,00	0,00	0,00	0,00
52	74	FORM1GASTAUSW		1						0,25	0,00	0,00	0,10
53	74	FORM1HEIM			1					0,25	0,00	0,00	0,10
54	10	FORM1GAST	1		2	1	4	2	1	0,25	1,00	3,00	1,10
55	27	DIFFFORM1	1		1	1		1	2	0,25	0,67	0,50	0,60
56	103	FORM2HEIMHEIM								0,00	0,00	0,00	0,00
57	36	FORM2GASTAUSW		2		1			1	0,50	0,33	0,50	0,40
58	103	FORM2HEIM								0,00	0,00	0,00	0,00
59	59	FORM2GAST				1	1			0,00	0,67	0,00	0,20
60	16	DIFFFORM2	3		1		2		2	0,75	0,33	1,00	0,80
61	103	FORM3HEIMHEIM								0,00	0,00	0,00	0,00
62	74	FORM3GASTAUSW			1					0,25	0,00	0,00	0,10
63	103	FORM3HEIM								0,00	0,00	0,00	0,00
64	74	FORM3GAST							1	0,00	0,00	0,50	0,10
65	1	DIFFFORM3	2	1	6	5		7	3	0,75	3,67	5,00	2,70
66	74	FORM4HEIMHEIM			1					0,25	0,00	0,00	0,10
67	12	FORM4GASTAUSW		1	1		2	2	1	0,50	1,33	1,50	1,00
68	59	FORM4HEIM		1		1				0,25	0,33	0,00	0,20
69	74	FORM4GAST			1					0,25	0,00	0,00	0,10
70	3	DIFFFORM4			1		4	2	3	0,25	2,00	4,00	2,00
71	103	FORM5HEIMHEIM								0,00	0,00	0,00	0,00
72	27	FORM5GASTAUSW		1		3			1	0,25	1,00	1,00	0,60
73	103	FORM5HEIM								0,00	0,00	0,00	0,00
74	36	FORM5GAST		1		1			1	0,50	0,33	0,50	0,40
75	36	DIFFFORM5		1	1		1	1		0,50	0,67	0,00	0,40
76	103	VIELETORE								0,00	0,00	0,00	0,00
77	74	AVG1PLATZHEIM			1					0,25	0,00	0,00	0,10
78	36	AVG1PLATZGAST		1	1			2		0,50	0,67	0,00	0,40
79	2	DIFFAVG1		4	1	2	2	3	2	2	2	1,75	2,33
80	103	AVG2PLATZHEIM								0,00	0,00	0,00	0,00
81	16	AVG2PLATZGAST		2		1	1		4	0,50	0,67	2,00	0,80
82	53	DIFFAVG2				1			2	0,25	0,00	1,00	0,30
83	53	AVG3PLATZHEIM				1		1	1	0,25	0,33	0,50	0,30
84	59	AVG3PLATZGAST		1				1		0,25	0,33	0,00	0,20
85	12	DIFFAVG3		1		4	2		2	1	1,25	0,67	1,50
86	74	AVG4PLATZHEIM			1					0,25	0,00	0,00	0,10
87	59	AVG4PLATZGAST			1			1		0,25	0,33	0,00	0,20
88	59	DIFFAVG4		1				1		0,25	0,33	0,00	0,20
89	74	AVG5PLATZHEIM						1		0,00	0,33	0,00	0,10
90	19	AVG5PLATZGAST		2		2			1	0,50	1,33	0,00	0,70
91	59	DIFFAVG5							2	0,00	0,00	0,00	0,20
92	103	HAEUFIGSTEPLAZIERUNGHEIM								0,00	0,00	0,00	0,00

93	103	HAEUFIGSTEPLAZIERUNGGAST									0,00	0,00	0,00	0,00		
94	74	ADDNATIONALTOP9HEIM		1							0,25	0,00	0,00	0,10		
95	36	ADDNATIONALTOP9GAST	1	1	1		1				0,75	0,33	0,00	0,40		
96	74	ADDNATIONALLAENDERSP.TOP9H.	1								0,25	0,00	0,00	0,10		
97	36	ADDNATIONALLAENDERSP.TOP9G.			1	1			1	1	0,25	0,33	0,50	0,40		
98	59	ADDNATIONALTOP9POSHEIM					1	1			0,00	0,33	0,50	0,20		
99	32	ADDNATIONALTOP9POSGAST	2		1				2		0,75	0,00	1,00	0,50		
100	59	ADDNAT.LAENDERSP.TOPPOS9H.	1						1		0,25	0,00	0,50	0,20		
101	36	ADDNAT.LAENDERSP.TOPPOS9G.		1	1				1	1	0,50	0,00	1,00	0,40		
102	19	ADDMITTELNATLSPHEIM			3	1	2	1			0,75	1,00	0,50	0,70		
103	36	ADDMITTELNATLSPGAST	1				1	1	1		0,25	0,33	1,00	0,40		
104	74	ADDLSPKTHEIM			1						0,25	0,00	0,00	0,10		
105	27	ADDLSPKTGAST		2	1	1				1	1	0,75	0,33	0,50	0,60	
106	59	ADDGESCHOSSENTOREHEIM			1						1	0,25	0,00	0,00	0,20	
107	59	ADDGESCHOSSENTOREGAST		1	1							0,50	0,00	0,00	0,20	
108	74	ADD DURCHGELASSEN TOREHEIM		1								0,25	0,00	0,00	0,10	
109	19	ADD DURCHGELASSEN TOREGAST		4	2		1					1,50	0,33	0,00	0,70	
110	103	ADDPOSITIONHEIM										0,00	0,00	0,00	0,00	
111	59	ADDPOSITIONGAST			1	1						0,50	0,00	0,00	0,20	
112	103	ADDSTADIONGROESSEHEIM										0,00	0,00	0,00	0,00	
113	53	ADDSTADIONGROESSEGAST	1	1				1				0,50	0,33	0,00	0,30	
114	103	ADD10ABSTIEGHEIM										0,00	0,00	0,00	0,00	
115	103	ADD10ABSTIEGGAST										0,00	0,00	0,00	0,00	
116	103	ADDBRAZILHEIM										0,00	0,00	0,00	0,00	
117	74	ADDBRAZILGAST		1								0,25	0,00	0,00	0,10	
118	74	ADDBRAZILNATHEIM				1						0,00	0,33	0,00	0,10	
119	74	ADDBRAZILNATGAST			1							0,25	0,00	0,00	0,10	
120	103	ADDGERMANYNATHEIM										0,00	0,00	0,00	0,00	
121	32	ADDGERMANYNATGAST			1	1	1	1	1			0,75	0,67	0,00	0,50	
122	103	ADDAUSTRIAHEIM										0,00	0,00	0,00	0,00	
123	103	ADDAUSTRIAGAST										0,00	0,00	0,00	0,00	
124	103	ADDAUSTRIANATHEIM										0,00	0,00	0,00	0,00	
125	103	ADDAUSTRIANATGAST										0,00	0,00	0,00	0,00	
126	103	ADDEWIGETABELLEHEIM										0,00	0,00	0,00	0,00	
127	4	ADDEWIGETABELLEGAST		5	2	1	1	2	3	1	2	1	2,00	2,00	1,50	1,80
128	103	ADDUEFA1JAHRHEIM										0,00	0,00	0,00	0,00	
129	59	ADDUEFA1JAHRGAST		1		1						0,50	0,00	0,00	0,20	
130	103	ADDUEFA5JAHRHEIM										0,00	0,00	0,00	0,00	
131	19	ADDUEFA5JAHREGAST			1	1	1	1		2	1	0,50	0,67	1,00	0,70	
132	103	ADDPUNKTEHEIMHEIM										0,00	0,00	0,00	0,00	
133	36	ADDPUNKTEGASTGAST			1	1			1		1	0,50	0,33	0,00	0,40	
134	103	ABSTIEG1HEIM										0,00	0,00	0,00	0,00	
135	103	ABSTIEG1GAST										0,00	0,00	0,00	0,00	
136	103	DIFFABSTIEG1										0,00	0,00	0,00	0,00	
137	103	ABSTIEG2HEIM										0,00	0,00	0,00	0,00	
138	74	ABSTIEG2GAST					1					0,25	0,00	0,00	0,10	
139	103	DIFFABSTIEG2										0,00	0,00	0,00	0,00	

140	103	ABSTIEG3HEIM						0,00	0,00	0,00	0,00		
141	103	ABSTIEG3GAST						0,00	0,00	0,00	0,00		
142	53	DIFFABSTIEG3	2			1		0,50	0,33	0,00	0,30		
143	103	ABSTIEG4HEIM						0,00	0,00	0,00	0,00		
144	103	ABSTIEG4GAST						0,00	0,00	0,00	0,00		
145	103	DIFFABSTIEG4						0,00	0,00	0,00	0,00		
146	103	ABSTIEG5HEIM						0,00	0,00	0,00	0,00		
147	74	ABSTIEG5GAST		1				0,25	0,00	0,00	0,10		
148	59	DIFFABSTIEG5	1		1		1	0,25	0,33	0,00	0,20		
149	53	DIFFPUNKTEHEIMGAST			1	1		0,25	0,33	0,50	0,30		
150	19	DIFFPKTLS	1		2	1		1	0,75	0,33	0,50	0,70	
151	32	DIFFGSTLS		2		1			0,50	0,33	0,00	0,50	
152	36	DIFFKASSTLS	2	1		1			0,75	0,33	0,00	0,40	
153	19	DIFFPOSLS			1	1		2	2	1	0,25	0,33	
154	36	DIFFEWIGETABELLE						4			2,00	0,70	
155	7	DIFFNAT	5		1		1		1	6	1,50	0,33	
156	32	DIFNAT2	2	1	1		1				1,00	0,33	
157	6	DIFFNATPOS	2	1	1	3	1	2	1	3	2	1,75	1,00
158	74	DIFNATPOS2				1					0,25	0,00	
159	27	DIFFNATPOSMIT		2	1		1		2			0,75	1,00
160	103	DIFFSTADGR									0,00	0,00	
161	103	DIFF10ABSTIEG									0,00	0,00	
162	103	DIFFBRAZIL									0,00	0,00	
163	74	DIFBRAZILN			1						0,25	0,00	
164	14	DIFFGERMANYN		1		2		2	1	2	1	0,75	1,00
165	103	DIFFAUSTRIA									1,00	0,90	
166	103	DIFFAUSTRIAN									0,00	0,00	
167	19	DIFFUEFA1J			1	1		2		1	2	0,50	0,67
168	9	DIFFUEFA5J		1	1	3		3	2	2		1,25	1,67
169	36	DIFFHAEUFPLAZ				1		1		2		0,25	0,33

Wrapper mit Klassifizierer: „JRip“ und Suchmethode: „BestFirst-Backward Elimination“

NR	Platz	Name	24	35	46	57	25	36	47	26	37	27	3SD	4SD	5SD	Ges
1	75	HEIM	10	10	10	10	10	10	10	10	9	9	10,00	10,00	9,50	9,80
2	1	GAST	10	10	10	10	10	10	10	10	10	10	10,00	10,00	10,00	10,00
3	133	POSHEIM	9	10	9	9	10	10	10	10	10	9	9,25	10,00	10,00	9,60
4	75	PLUSMINUSHEIM	10	10	10	10	10	9	10	9	10	10	10,00	9,67	9,50	9,80
5	1	SIEGEHEIM	10	10	10	10	10	10	10	10	10	10	10,00	10,00	10,00	10,00
6	1	REMISHEIM	10	10	10	10	10	10	10	10	10	10	10,00	10,00	10,00	10,00
7	112	NIEDERLAGENHEIM	10	9	9	9	10	10	10	10	10	10	9,25	10,00	10,00	9,70
8	1	PUNKTEHEIM	10	10	10	10	10	10	10	10	10	10	10,00	10,00	10,00	10,00
9	75	TDHEIM	10	10	9	9	10	10	10	10	10	10	9,50	10,00	10,00	9,80
10	1	POSGAST	10	10	10	10	10	10	10	10	10	10	10,00	10,00	10,00	10,00
11	75	PLUSMINUSGAST	10	10	10	10	10	10	10	10	9	9	10,00	10,00	9,50	9,80
12	1	SIEGEGAST	10	10	10	10	10	10	10	10	10	10	10,00	10,00	10,00	10,00
13	112	REMSGAST	9	10	9	9	10	10	10	10	10	10	9,25	10,00	10,00	9,70
14	75	NIEDERLAGENGAST	9	10	10	10	10	10	10	10	9	10	9,75	10,00	9,50	9,80
15	1	PUNKTEGAST	10	10	10	10	10	10	10	10	10	10	10,00	10,00	10,00	10,00
16	1	TDGAST	10	10	10	10	10	10	10	10	10	10	10,00	10,00	10,00	10,00
17	35	MONAT	10	9	10	10	10	10	10	10	10	10	9,75	10,00	10,00	9,90
18	1	JAHRESZEIT	10	10	10	10	10	10	10	10	10	10	10,00	10,00	10,00	10,00
19	35	TABDIFF	10	9	10	10	10	10	10	10	10	10	9,75	10,00	10,00	9,90
20	112	TABTENDENZ	10	10	9	9	9	10	10	10	10	10	9,50	9,67	10,00	9,70
21	75	TABDIFFFORM	10	10	9	9	10	10	10	10	10	10	9,50	10,00	10,00	9,80
22	75	TABDIFFFORMVERG	10	9	10	10	10	10	10	10	9	10	9,75	10,00	9,50	9,80
23	75	SIEGDIFF	10	9	10	10	10	10	10	10	10	9	9,75	10,00	10,00	9,80
24	75	REMISDIFF	10	10	10	10	10	8	10	10	10	10	10,00	9,33	10,00	9,80
25	75	NIEDDIFF	10	10	10	10	10	9	10	9	10	10	10,00	9,67	9,50	9,80
26	35	PUNKTEDIFF	10	9	10	10	10	10	10	10	10	10	9,75	10,00	10,00	9,90
27	35	TDDIFF	10	10	10	10	10	10	10	10	10	9	10,00	10,00	10,00	9,90
28	133	TORDIFFFORM	10	10	9	9	10	9	10	10	10	9	9,50	9,67	10,00	9,60
29	1	TFHEIM	10	10	10	10	10	10	10	10	10	10	10,00	10,00	10,00	10,00
30	35	TFGAST	10	10	10	10	10	10	10	10	10	9	10,00	10,00	10,00	9,90
31	35	HEIMSTARK	10	10	10	10	10	9	10	10	10	10	10,00	9,67	10,00	9,90
32	1	GASTSTARK	10	10	10	10	10	10	10	10	10	10	10,00	10,00	10,00	10,00
33	35	SIEGEHEIM	10	9	10	10	10	10	10	10	10	10	9,75	10,00	10,00	9,90
34	153	VERSUS1HEIM	10	10	10	10	8	10	9	9	10	9	10,00	9,00	9,50	9,50
35	75	VERSUS1GAST	9	10	10	10	10	9	10	10	10	10	9,75	9,67	10,00	9,80
36	75	VERSUS1	10	10	10	10	10	10	9	9	10	10	10,00	9,67	9,50	9,80
37	112	VERSUS2HEIM	10	10	9	9	10	10	9	10	10	10	9,50	9,67	10,00	9,70
38	1	VERSUS2GAST	10	10	10	10	10	10	10	10	10	10	10,00	10,00	10,00	10,00
39	75	VERSUS2	10	10	9	9	10	10	10	10	10	10	9,50	10,00	10,00	9,80
40	133	VERSUS3HEIM	10	10	8	8	10	10	10	10	10	10	9,00	10,00	10,00	9,60
41	75	VERSUS3GAST	10	10	10	10	9	10	10	10	9	10	10,00	9,67	9,50	9,80
42	153	VERSUS3	9	10	9	9	9	10	10	10	9	10	9,25	9,67	9,50	9,50
43	133	VERSUS4HEIM	10	8	9	9	10	10	10	10	10	10	9,00	10,00	10,00	9,60
44	1	VERSUS4GAST	10	10	10	10	10	10	10	10	10	10	10,00	10,00	10,00	10,00
45	35	VERSUS4	10	10	10	10	9	10	10	10	10	10	10,00	9,67	10,00	9,90

46	1	VERSUS5HEIM	10 10 10 10 10 10 10 10 10 10 10 10 10,00 10,00 10,00 10,00
47	35	VERSUS5GAST	10 10 10 10 10 10 9 10 10 10 10 10 10,00 9,67 10,00 9,90
48	153	VERSUS5	10 9 9 9 9 9 10 10 10 10 10 10 9,25 9,33 10,00 9,50
49	75	TABGAST	10 10 10 10 10 9 9 10 10 10 10 10 10,00 9,33 10,00 9,80
50	75	TABVERGLEICH	10 10 9 9 10 10 10 10 10 10 10 10 9,50 10,00 10,00 9,80
51	1	FORM1HEIMHEIM	10 10 10 10 10 10 10 10 10 10 10 10 10,00 10,00 10,00 10,00
52	133	FORM1GASTAUSW	10 10 9 9 10 10 9 10 9 10 9 10 9,50 9,67 9,50 9,60
53	35	FORM1HEIM	10 10 10 10 10 10 10 10 10 9 10 10 10,00 10,00 9,50 9,90
54	1	FORM1GAST	10 10 10 10 10 10 10 10 10 10 10 10 10,00 10,00 10,00 10,00
55	133	DIFFFORM1	10 10 9 9 9 9 10 10 10 10 10 10 9,50 9,33 10,00 9,60
56	133	FORM2HEIMHEIM	10 10 9 9 10 10 9 9 10 10 9,50 9,67 9,50 9,60
57	75	FORM2GASTAUSW	10 10 10 10 10 10 10 10 10 8 10 10,00 10,00 9,00 9,80
58	35	FORM2HEIM	10 10 10 10 10 10 10 9 10 10 10,00 10,00 9,50 9,90
59	112	FORM2GAST	9 9 10 10 10 10 10 10 10 9 10,00 10,00 10,00 9,70
60	112	DIFFFORM2	9 9 10 10 10 9 10 10 10 10 9,50 9,67 10,00 9,70
61	1	FORM3HEIMHEIM	10 10 10 10 10 10 10 10 10 10 10 10 10,00 10,00 10,00 10,00
62	133	FORM3GASTAUSW	10 10 9 9 10 10 10 9 10 9 9,50 10,00 9,50 9,60
63	75	FORM3HEIM	10 10 10 10 10 10 8 10 10 10 10,00 9,33 10,00 9,80
64	35	FORM3GAST	10 10 10 10 10 10 10 10 10 9 10,00 10,00 10,00 9,90
65	1	DIFFFORM3	10 10 10 10 10 10 10 10 10 10 10 10,00 10,00 10,00 10,00
66	35	FORM4HEIMHEIM	10 10 10 10 10 10 10 10 10 10 9 10,00 10,00 10,00 9,90
67	1	FORM4GASTAUSW	10 10 10 10 10 10 10 10 10 10 10 10 10,00 10,00 10,00 10,00
68	35	FORM4HEIM	9 10 10 10 10 10 10 10 10 10 10 9,75 10,00 10,00 9,90
69	75	FORM4GAST	10 10 10 10 10 10 10 10 10 9 9 10,00 10,00 9,50 9,80
70	112	DIFFFORM4	10 9 10 10 10 10 10 10 10 9 9 9,75 10,00 9,50 9,70
71	133	FORM5HEIMHEIM	10 10 10 10 10 10 9 9 8 10,00 10,00 9,00 9,60
72	112	FORM5GASTAUSW	9 9 10 10 10 10 10 9 10 10 9,50 10,00 9,50 9,70
73	75	FORM5HEIM	10 10 9 9 10 10 10 10 10 10 9,50 10,00 10,00 9,80
74	112	FORM5GAST	10 9 10 10 10 9 10 10 10 10 9 9,75 9,67 10,00 9,70
75	75	DIFFFORM5	10 10 10 10 10 9 10 10 10 10 9 10,00 9,67 10,00 9,80
76	35	VIELETORE	10 10 10 10 10 10 9 10 10 10 10 10,00 10,00 9,50 9,90
77	153	AVG1PLATZHEIM	10 9 10 10 9 9 10 10 10 10 8 9,75 9,33 10,00 9,50
78	35	AVG1PLATZGAST	10 10 10 10 10 10 10 10 10 9 10 10,00 10,00 9,50 9,90
79	133	DIFFAVG1	10 9 10 10 9 8 10 10 10 10 10 9,75 9,00 10,00 9,60
80	153	AVG2PLATZHEIM	10 8 10 10 10 9 9 9 10 10 10 9,50 9,33 9,50 9,50
81	35	AVG2PLATZGAST	9 10 10 10 10 10 10 10 10 10 10 9,75 10,00 10,00 9,90
82	35	DIFFAVG2	10 10 10 10 10 9 10 10 10 10 10 10,00 9,67 10,00 9,90
83	35	AVG3PLATZHEIM	10 10 10 10 10 10 10 10 10 9 10 10,00 10,00 9,50 9,90
84	35	AVG3PLATZGAST	10 10 10 10 10 9 10 10 10 10 10 10,00 9,67 10,00 9,90
85	133	DIFFAVG3	10 8 10 10 10 9 10 10 10 9 10 9,50 9,67 9,50 9,60
86	165	AVG4PLATZHEIM	9 10 9 9 10 9 9 10 9 10 9,25 9,33 9,50 9,40
87	165	AVG4PLATZGAST	10 10 9 9 8 10 10 10 9 9 9,50 9,33 9,50 9,40
88	1	DIFFAVG4	10 10 10 10 10 10 10 10 10 10 10 10,00 10,00 10,00 10,00
89	35	AVG5PLATZHEIM	10 10 10 10 10 10 10 10 10 10 9 10,00 10,00 10,00 9,90
90	75	AVG5PLATZGAST	10 9 10 10 10 9 10 10 10 10 10 9,75 9,67 10,00 9,80
91	153	DIFFAVG5	10 10 9 9 10 10 10 9 9 9 9,50 10,00 9,00 9,50
92	35	HAEUFIGSTEPLAZIERUNGHEIM	10 10 10 10 10 10 9 10 10 10 10 10,00 10,00 9,50 9,90

93	75	HAEUFIGSTEPLAZIERUNGGAST	10 10 10 10 9 10 10 10 10 9 10	10,00 9,67 9,50 9,80
94	112	ADDNATIONALTOP9HEIM	10 9 10 10 9 10 10 10 10 9 10	9,75 9,67 9,50 9,70
95	1	ADDNATIONALTOP9GAST	10 10 10 10 10 10 10 10 10 10 10	10,00 10,00 10,00 10,00
96	1	ADDNATIONALLAENDERSP.TOP9H.	10 10 10 10 10 10 10 10 10 10 10	10,00 10,00 10,00 10,00
97	1	ADDNATIONALLAENDERSP.TOP9G.	10 10 10 10 10 10 10 10 10 10 10	10,00 10,00 10,00 10,00
98	112	ADDNATIONALTOP9POSHEIM	10 10 9 9 10 10 9 10 10 10 10	9,50 9,67 10,00 9,70
99	35	ADDNATIONALTOP9POSGAST	10 9 10 10 10 10 10 10 10 10 10	9,75 10,00 10,00 9,90
100	1	ADDNAT.LAENDERSP.TOPPOS9H.	10 10 10 10 10 10 10 10 10 10 10	10,00 10,00 10,00 10,00
101	75	ADDNAT.LAENDERSP.TOPPOS9G.	10 10 9 9 10 10 10 10 10 10 10	9,50 10,00 10,00 9,80
102	35	ADDMITTELNATLSPHEIM	9 10 10 10 10 10 10 10 10 10 10	9,75 10,00 10,00 9,90
103	133	ADDMITTELNATLSPGAST	9 10 10 10 9 10 9 9 10 10	9,75 9,33 9,50 9,60
104	1	ADDLSPKTHEIM	10 10 10 10 10 10 10 10 10 10 10	10,00 10,00 10,00 10,00
105	35	ADDLSPKTGAST	10 10 10 10 10 10 10 9 10 10 10	10,00 10,00 9,50 9,90
106	75	ADDGESCHOSSENTOREHEIM	9 10 10 10 10 10 10 9 10 10 10	9,75 10,00 9,50 9,80
107	35	ADDGESCHOSSENTOREGAST	10 10 10 10 10 10 10 9 10 10 10	10,00 10,00 9,50 9,90
108	1	ADDDURCHGELASSENTOREHEIM	10 10 10 10 10 10 10 10 10 10 10	10,00 10,00 10,00 10,00
109	1	ADDDURCHGELASSENTOREGAST	10 10 10 10 10 10 10 10 10 10 10	10,00 10,00 10,00 10,00
110	112	ADDPOSITIONHEIM	10 10 9 9 10 10 10 9 10 10 10	9,50 10,00 9,50 9,70
111	133	ADDPOSITIONGAST	10 10 9 9 10 10 10 8 10 10 10	9,50 10,00 9,00 9,60
112	112	ADDSTADIONGROESSEHEIM	8 10 10 10 10 10 10 10 10 9	9,50 10,00 10,00 9,70
113	1	ADDSTADIONGROESSEGAST	10 10 10 10 10 10 10 10 10 10 10	10,00 10,00 10,00 10,00
114	153	ADD10ABSTIEGHEIM	10 10 9 9 9 9 10 9 10 10	9,50 9,33 9,50 9,50
115	75	ADD10ABSTIEGGAST	10 10 9 9 10 10 10 10 10 10 10	9,50 10,00 10,00 9,80
116	75	ADDBRAZILHEIM	10 10 10 10 10 9 10 10 9 10	10,00 9,67 9,50 9,80
117	35	ADDBRAZILGAST	10 10 10 10 10 10 9 10 10 10	10,00 9,67 10,00 9,90
118	75	ADDBRAZILNATHEIM	10 10 10 10 10 10 10 10 9 9	10,00 10,00 9,50 9,80
119	1	ADDBRAZILNATGAST	10 10 10 10 10 10 10 10 10 10	10,00 10,00 10,00 10,00
120	35	ADDGERMANYNATHEIM	10 10 10 10 10 10 10 10 9 10	10,00 10,00 9,50 9,90
121	168	ADDGERMANYNATGAST	9 9 9 9 10 9 9 9 10 10	9,00 9,33 9,50 9,30
122	35	ADDAUSTRIAHEIM	10 10 10 10 9 10 10 10 10 10	10,00 9,67 10,00 9,90
123	153	ADDAUSTRIAGAST	10 10 9 9 10 9 10 10 8 10	9,50 9,67 9,00 9,50
124	35	ADDAUSTRIANATHEIM	10 10 10 10 9 10 10 10 10 10	10,00 9,67 10,00 9,90
125	1	ADDAUSTRIANATGAST	10 10 10 10 10 10 10 10 10 10	10,00 10,00 10,00 10,00
126	1	ADDEWIGETABELLEHEIM	10 10 10 10 10 10 10 10 10 10	10,00 10,00 10,00 10,00
127	75	ADDEWIGETABELLEGAST	10 9 10 10 10 10 10 10 10 9	9,75 10,00 10,00 9,80
128	75	ADDUEFA1JAHRHEIM	9 10 10 10 10 10 9 10 10 10	9,75 9,67 10,00 9,80
129	153	ADDUEFA1JAHRGAST	10 10 10 10 9 10 9 8 10 9	10,00 9,33 9,00 9,50
130	35	ADDUEFA5JAHREHEIM	10 10 10 10 10 10 10 9 10	10,00 10,00 9,50 9,90
131	35	ADDUEFA5JAHREGAST	10 9 10 10 10 10 10 10 10 10	9,75 10,00 10,00 9,90
132	133	ADDPUNKTEHEIMHEIM	10 10 9 9 10 10 9 10 9 10	9,50 9,67 9,50 9,60
133	133	ADDPUNKTEGASTGAST	10 9 10 10 10 9 10 10 9 9	9,75 9,67 9,50 9,60
134	165	ABSTIEG1HEIM	9 9 9 9 9 10 10 9 10 10	9,00 9,67 9,50 9,40
135	75	ABSTIEG1GAST	10 10 10 10 10 10 8 10 10 10	10,00 9,33 10,00 9,80
136	133	DIFFABSTIEG1	8 10 9 9 10 10 10 10 10 10	9,00 10,00 10,00 9,60
137	112	ABSTIEG2HEIM	10 10 9 9 10 10 10 9 10 10	9,50 10,00 9,50 9,70
138	112	ABSTIEG2GAST	10 10 10 10 9 10 10 10 10 8	10,00 9,67 10,00 9,70
139	112	DIFFABSTIEG2	10 8 10 10 10 10 9 10 10 10	9,50 9,67 10,00 9,70

140	1	ABSTIEG3HEIM	10 10 10 10 10 10 10 10 10 10 10 10 10,00 10,00 10,00 10,00
141	35	ABSTIEG3GAST	10 10 10 10 10 10 10 10 10 9 10 10,00 10,00 9,50 9,90
142	75	DIFFABSTIEG3	10 10 10 10 10 10 10 10 10 10 8 10,00 10,00 10,00 9,80
143	133	ABSTIEG4HEIM	10 9 10 10 9 9 10 10 10 9 10 9,75 9,33 9,50 9,60
144	35	ABSTIEG4GAST	10 10 10 10 10 10 10 10 10 10 9 10,00 10,00 10,00 9,90
145	133	DIFFABSTIEG4	10 10 10 10 10 9 9 10 10 10 8 10,00 9,33 10,00 9,60
146	35	ABSTIEG5HEIM	10 10 10 10 9 10 10 10 10 10 10 10,00 9,67 10,00 9,90
147	35	ABSTIEG5GAST	10 9 10 10 10 10 10 10 10 10 10 9,75 10,00 10,00 9,90
148	35	DIFFABSTIEG5	10 10 10 10 10 10 10 10 9 10 10 10,00 10,00 9,50 9,90
149	133	DIFFPUNKTEHEIMGAST	10 10 10 10 10 9 9 9 10 9 10,00 9,33 9,50 9,60
150	1	DIFFPKTLS	10 10 10 10 10 10 10 10 10 10 10 10,00 10,00 10,00 10,00
151	112	DIFFGSTLS	10 10 10 10 10 10 9 9 9 10 10,00 9,67 9,00 9,70
152	112	DIFFKASSTLS	10 10 10 10 10 10 9 10 9 9 10,00 9,67 9,50 9,70
153	169	DIFFPOSLS	8 9 9 9 10 10 9 10 9 9 8,75 9,67 9,50 9,20
154	75	DIFFEWIGETABELLE	9 10 10 10 10 10 10 9 10 10 9,75 10,00 9,50 9,80
155	133	DIFFNAT	10 10 9 9 10 9 10 9 10 10 9,50 9,67 9,50 9,60
156	35	DIFNAT2	10 10 10 10 10 10 10 10 9 10 10,00 10,00 9,50 9,90
157	35	DIFFNATPOS	10 10 10 10 10 10 10 9 10 10 10,00 10,00 9,50 9,90
158	75	DIFNATPOS2	10 9 10 10 10 9 10 10 10 10 9,75 9,67 10,00 9,80
159	35	DIFFNATPOSMIT	10 10 10 10 10 9 10 10 10 10 10 10,00 9,67 10,00 9,90
160	153	DIFFSTADGR	10 10 9 9 10 8 10 10 9 10 9,50 9,33 9,50 9,50
161	75	DIFF10ABSTIEG	10 9 10 10 10 10 10 10 9 10 9,75 10,00 9,50 9,80
162	1	DIFFBRAZIL	10 10 10 10 10 10 10 10 10 10 10,00 10,00 10,00 10,00
163	112	DIFBRAZILN	10 10 10 10 9 10 10 9 9 10 10,00 9,67 9,00 9,70
164	153	DIFFGERMANYN	8 10 10 10 10 10 10 10 9 8 9,50 10,00 9,50 9,50
165	153	DIFFAUSTRIA	10 9 9 9 9 10 10 10 10 9 9,25 9,67 10,00 9,50
166	112	DIFFAUSTRIAN	10 10 9 9 10 10 9 10 10 10 9,50 9,67 10,00 9,70
167	75	DIFFUEFA1J	10 10 10 10 8 10 10 10 10 10 10,00 9,33 10,00 9,80
168	112	DIFFUEFA5J	10 10 9 9 10 10 10 9 10 10 9,50 10,00 9,50 9,70
169	75	DIFFHAEUFPLAZ	10 8 10 10 10 10 10 10 10 10 9,50 10,00 10,00 9,80

Gesamtergebnis Wrapper mit Klassifizierer: „JRip“ über alle 3 Suchmethoden

NR	Platz	Name	Gesamt	Prozent
1	131	HEIM	3,63	36,33
2	94	GAST	4,10	41,00
3	117	POSHEIM	3,83	38,33
4	153	PLUSMINUSHEIM	3,50	35,00
5	135	SIEGEHEIM	3,60	36,00
6	135	REMISHEIM	3,60	36,00
7	143	NIEDERLAGENHEIM	3,53	35,33
8	131	PUNKTEHEIM	3,63	36,33
9	74	TDHEIM	4,40	44,00
10	88	POSGAST	4,17	41,67
11	141	PLUSMINUSGAST	3,57	35,67
12	85	SIEGEGAST	4,23	42,33
13	143	REMISGAST	3,53	35,33
14	143	NIEDERLAGENGAST	3,53	35,33
15	92	PUNKTEGAST	4,13	41,33
16	74	TDGAST	4,40	44,00
17	143	MONAT	3,53	35,33
18	143	JAHRESZEIT	3,53	35,33
19	35	TABDIFF	5,17	51,67
20	135	TABTENDENZ	3,60	36,00
21	162	TABDIFFFORM	3,40	34,00
22	162	TABDIFFFORMVERG	3,40	34,00
23	88	SIEGDIFF	4,17	41,67
24	164	REMISDIFF	3,37	33,67
25	88	NIEDDIFF	4,17	41,67
26	43	PUNKTEDIFF	4,93	49,33
27	23	TDDIFF	5,70	57,00
28	141	TORDIFFFORM	3,57	35,67
29	33	TFHEIM	5,27	52,67
30	131	TFGAST	3,63	36,33
31	78	HEIMSTARK	4,33	43,33
32	158	GASTSTARK	3,43	34,33
33	158	SIEGEHEIM	3,43	34,33
34	95	VERSUS1HEIM	4,07	40,67
35	95	VERSUS1GAST	4,07	40,67
36	68	VERSUS1	4,47	44,67
37	143	VERSUS2HEIM	3,53	35,33
38	95	VERSUS2GAST	4,07	40,67
39	86	VERSUS2	4,20	42,00
40	166	VERSUS3HEIM	3,33	33,33
41	113	VERSUS3GAST	3,90	39,00
42	103	VERSUS3	3,97	39,67
43	168	VERSUS4HEIM	3,30	33,00
44	117	VERSUS4GAST	3,83	38,33
45	95	VERSUS4	4,07	40,67

46	153	VERSUS5HEIM	3,50	35,00
47	122	VERSUS5GAST	3,80	38,00
48	92	VERSUS5	4,13	41,33
49	153	TABGAST	3,50	35,00
50	117	TABVERGLEICH	3,83	38,33
51	135	FORM1HEIMHEIM	3,60	36,00
52	58	FORM1GASTAUSW	4,63	46,33
53	52	FORM1HEIM	4,73	47,33
54	36	FORM1GAST	5,13	51,33
55	8	DIFFFORM1	6,13	61,33
56	143	FORM2HEIMHEIM	3,53	35,33
57	46	FORM2GASTAUSW	4,87	48,67
58	63	FORM2HEIM	4,57	45,67
59	58	FORM2GAST	4,63	46,33
60	6	DIFFFORM2	6,30	63,00
61	130	FORM3HEIMHEIM	3,73	37,33
62	74	FORM3GASTAUSW	4,40	44,00
63	110	FORM3HEIM	3,93	39,33
64	61	FORM3GAST	4,60	46,00
65	1	DIFFFORM3	7,13	71,33
66	127	FORM4HEIMHEIM	3,77	37,67
67	23	FORM4GASTAUSW	5,70	57,00
68	82	FORM4HEIM	4,27	42,67
69	61	FORM4GAST	4,60	46,00
70	3	DIFFFORM4	6,73	67,33
71	143	FORM5HEIMHEIM	3,53	35,33
72	40	FORM5GASTAUSW	4,97	49,67
73	65	FORM5HEIM	4,53	45,33
74	39	FORM5GAST	5,03	50,33
75	9	DIFFFORM5	6,07	60,67
76	153	VIELETORE	3,50	35,00
77	78	AVG1PLATZHEIM	4,33	43,33
78	27	AVG1PLATZGAST	5,43	54,33
79	2	DIFFAVG1	6,77	67,67
80	82	AVG2PLATZHEIM	4,27	42,67
81	70	AVG2PLATZGAST	4,43	44,33
82	9	DIFFAVG2	6,07	60,67
83	40	AVG3PLATZHEIM	4,97	49,67
84	37	AVG3PLATZGAST	5,10	51,00
85	11	DIFFAVG3	6,03	60,33
86	122	AVG4PLATZHEIM	3,80	38,00
87	58	AVG4PLATZGAST	4,63	46,33
88	18	DIFFAVG4	5,80	58,00
89	70	AVG5PLATZHEIM	4,43	44,33
90	78	AVG5PLATZGAST	4,33	43,33
91	15	DIFFAVG5	5,87	58,67
92	115	HAEUFIGSTEPLAZIERUNGHEIM	3,87	38,67

93	117	HAEUFIGSTEPLAZIERUNGGAST	3,83	38,33
94	110	ADDNATIONALTOP9HEIM	3,93	39,33
95	53	ADDNATIONALTOP9GAST	4,70	47,00
96	37	ADDNATIONALLAENDERSP.TOP9H.	5,10	51,00
97	99	ADDNATIONALLAENDERSP.TOP9G.	4,03	40,33
98	110	ADDNATIONALTOP9POSHEIM	3,93	39,33
99	50	ADDNATIONALTOP9POSGAST	4,77	47,67
100	74	ADDNAT.LAENDERSP.TOPPOS9H.	4,40	44,00
101	103	ADDNAT.LAENDERSP.TOPPOS9G.	3,97	39,67
102	68	ADDMITTELNATLSPHEIM	4,47	44,67
103	103	ADDMITTELNATLSPGAST	3,97	39,67
104	70	ADDLSPKTHEIM	4,43	44,33
105	53	ADDLSPKTGAST	4,70	47,00
106	48	ADDGESCHOSSENETOREHEIM	4,83	48,33
107	115	ADDGESCHOSSENETOREGAST	3,87	38,67
108	122	ADDDURCHGELASSENETOREHEIM	3,80	38,00
109	103	ADDDURCHGELASSENETOREGAST	3,97	39,67
110	46	ADDPOSITIONHEIM	4,87	48,67
111	113	ADDPOSITIONGAST	3,90	39,00
112	135	ADDSTADIONGROESSEHEIM	3,60	36,00
113	122	ADDSTADIONGROESSEGAST	3,80	38,00
114	158	ADD10ABSTIEGHEIM	3,43	34,33
115	103	ADD10ABSTIEGGAST	3,97	39,67
116	157	ADDBRAZILHEIM	3,47	34,67
117	122	ADDBRAZILGAST	3,80	38,00
118	131	ADDBRAZILNATHEIM	3,63	36,33
119	127	ADDBRAZILNATGAST	3,77	37,67
120	63	ADDGERMANYNATHEIM	4,57	45,67
121	49	ADDGERMANYNATGAST	4,80	48,00
122	158	ADDAUSTRIAHEIM	3,43	34,33
123	168	ADDAUSTRIAGAST	3,30	33,00
124	143	ADDAUSTRIANATHEIM	3,53	35,33
125	143	ADDAUSTRIANATGAST	3,53	35,33
126	67	ADDEWIGETABELLEHEIM	4,50	45,00
127	18	ADDEWIGETABELLEGAST	5,80	58,00
128	100	ADDUEFA1JAHRHEIM	4,00	40,00
129	31	ADDUEFA1JAHRGAST	5,30	53,00
130	82	ADDUEFA5JAHREHEIM	4,27	42,67
131	53	ADDUEFA5JAHREGAST	4,70	47,00
132	100	ADDPUNKTEHEIMHEIM	4,00	40,00
133	53	ADDPUNKTEGASTGAST	4,70	47,00
134	103	ABSTIEG1HEIM	3,97	39,67
135	15	ABSTIEG1GAST	5,87	58,67
136	44	DIFFABSTIEG1	4,90	49,00
137	88	ABSTIEG2HEIM	4,17	41,67
138	15	ABSTIEG2GAST	5,87	58,67
139	40	DIFFABSTIEG2	4,97	49,67

140	81	ABSTIEG3HEIM	4,30	43,00
141	18	ABSTIEG3GAST	5,80	58,00
142	27	DIFFABSTIEG3	5,43	54,33
143	86	ABSTIEG4HEIM	4,20	42,00
144	25	ABSTIEG4GAST	5,63	56,33
145	53	DIFFABSTIEG4	4,70	47,00
146	100	ABSTIEG5HEIM	4,00	40,00
147	29	ABSTIEG5GAST	5,40	54,00
148	65	DIFFABSTIEG5	4,53	45,33
149	13	DIFFPUNKTEHEIMGAST	6,00	60,00
150	26	DIFFPKTLS	5,60	56,00
151	34	DIFFGSTLS	5,20	52,00
152	44	DIFFKASSTLS	4,90	49,00
153	29	DIFFPOSLS	5,40	54,00
154	70	DIFFEWIGETABELLE	4,43	44,33
155	7	DIFFNAT	6,23	62,33
156	14	DIFNAT2	5,90	59,00
157	5	DIFFNATPOS	6,37	63,67
158	31	DIFNATPOS2	5,30	53,00
159	11	DIFFNATPOSMIT	6,03	60,33
160	135	DIFFSTADGR	3,60	36,00
161	103	DIFF10ABSTIEG	3,97	39,67
162	127	DIFFBRAZIL	3,77	37,67
163	117	DIFFBRAZILN	3,83	38,33
164	22	DIFFGERMANYN	5,77	57,67
165	166	DIFFAUSTRIA	3,33	33,33
166	164	DIFFAUSTRIAN	3,37	33,67
167	18	DIFFUEFA1J	5,80	58,00
168	4	DIFFUEFA5J	6,40	64,00
169	50	DIFFHAEUFPLAZ	4,77	47,67

Wrapper mit Klassifizierer: „J48“ und Suchmethode: „RankSearch“

NR	Platz	Name	24	35	46	57	25	36	47	26	37	27	3SD	4SD	5SD	Ges	
1	48	HEIM											0,00	0,00	0,00	0,00	
2	48	GAST											0,00	0,00	0,00	0,00	
3	48	POSHEIM											0,00	0,00	0,00	0,00	
4	48	PLUSMINUSHEIM											0,00	0,00	0,00	0,00	
5	48	SIEGEHEIM											0,00	0,00	0,00	0,00	
6	48	REMISHEIM											0,00	0,00	0,00	0,00	
7	35	NIEDERLAGENHEIM										1	0,00	0,33	0,00	0,10	
8	48	PUNKTEHEIM											0,00	0,00	0,00	0,00	
9	48	TDHEIM											0,00	0,00	0,00	0,00	
10	48	POSGAST											0,00	0,00	0,00	0,00	
11	48	PLUSMINUSGAST											0,00	0,00	0,00	0,00	
12	28	SIEGEGAST								1		1	0,25	0,33	0,00	0,20	
13	48	REMISGAST											0,00	0,00	0,00	0,00	
14	48	NIEDERLAGENGAST											0,00	0,00	0,00	0,00	
15	48	PUNKTEGAST											0,00	0,00	0,00	0,00	
16	48	TDGAST											0,00	0,00	0,00	0,00	
17	48	MONAT											0,00	0,00	0,00	0,00	
18	48	JAHRESZEIT											0,00	0,00	0,00	0,00	
19	48	TABDIFF											0,00	0,00	0,00	0,00	
20	48	TABTENDENZ											0,00	0,00	0,00	0,00	
21	48	TABDIFFFORM											0,00	0,00	0,00	0,00	
22	48	TABDIFFFORMVERG											0,00	0,00	0,00	0,00	
23	48	SIEGDIFF											0,00	0,00	0,00	0,00	
24	48	REMISDIFF											0,00	0,00	0,00	0,00	
25	48	NIEDDIFF											0,00	0,00	0,00	0,00	
26	48	PUNKTEDIFF											0,00	0,00	0,00	0,00	
27	17	TDDIFF					3					1	0,75	0,00	0,50	0,40	
28	48	TORDIFFFORM											0,00	0,00	0,00	0,00	
29	17	TFHEIM							1	1			1	0,50	0,33	0,50	0,40
30	48	TFGAST											0,00	0,00	0,00	0,00	
31	48	HEIMSTARK											0,00	0,00	0,00	0,00	
32	48	GASTSTARK											0,00	0,00	0,00	0,00	
33	48	SIEGEHEIM											0,00	0,00	0,00	0,00	
34	28	VERSUS1HEIM							2				0,50	0,00	0,00	0,20	
35	6	VERSUS1GAST					10			1			2,50	0,33	0,00	1,10	
36	48	VERSUS1											0,00	0,00	0,00	0,00	
37	17	VERSUS2HEIM						3		1			0,75	0,33	0,00	0,40	
38	6	VERSUS2GAST					10			1			2,50	0,33	0,00	1,10	
39	48	VERSUS2											0,00	0,00	0,00	0,00	
40	48	VERSUS3HEIM											0,00	0,00	0,00	0,00	
41	35	VERSUS3GAST							1				0,25	0,00	0,00	0,10	
42	48	VERSUS3											0,00	0,00	0,00	0,00	
43	48	VERSUS4HEIM											0,00	0,00	0,00	0,00	
44	48	VERSUS4GAST											0,00	0,00	0,00	0,00	
45	48	VERSUS4											0,00	0,00	0,00	0,00	

46	48	VERSUS5HEIM								0,00	0,00	0,00	0,00		
47	48	VERSUS5GAST								0,00	0,00	0,00	0,00		
48	48	VERSUS5								0,00	0,00	0,00	0,00		
49	48	TABGAST								0,00	0,00	0,00	0,00		
50	48	TABVERGLEICH								0,00	0,00	0,00	0,00		
51	48	FORM1HEIMHEIM								0,00	0,00	0,00	0,00		
52	48	FORM1GASTAUSW								0,00	0,00	0,00	0,00		
53	48	FORM1HEIM								0,00	0,00	0,00	0,00		
54	48	FORM1GAST								0,00	0,00	0,00	0,00		
55	1	DIFFFORM1	8	9			1		5	4,25	0,33	0,00	2,30		
56	48	FORM2HEIMHEIM								0,00	0,00	0,00	0,00		
57	48	FORM2GASTAUSW								0,00	0,00	0,00	0,00		
58	35	FORM2HEIM			1					0,25	0,00	0,00	0,10		
59	48	FORM2GAST								0,00	0,00	0,00	0,00		
60	4	DIFFFORM2	5	1		2	4		2	1	1,50	2,00	1,00	1,50	
61	48	FORM3HEIMHEIM								0,00	0,00	0,00	0,00		
62	48	FORM3GASTAUSW								0,00	0,00	0,00	0,00		
63	48	FORM3HEIM								0,00	0,00	0,00	0,00		
64	48	FORM3GAST								0,00	0,00	0,00	0,00		
65	2	DIFFFORM3	5	8		1	3	1	2	1	3,25	1,33	1,50	2,10	
66	48	FORM4HEIMHEIM								0,00	0,00	0,00	0,00		
67	48	FORM4GASTAUSW								0,00	0,00	0,00	0,00		
68	48	FORM4HEIM								0,00	0,00	0,00	0,00		
69	48	FORM4GAST								0,00	0,00	0,00	0,00		
70	6	DIFFFORM4	3	2		1	4		1		1,25	1,67	0,50	1,10	
71	48	FORM5HEIMHEIM								0,00	0,00	0,00	0,00		
72	48	FORM5GASTAUSW								0,00	0,00	0,00	0,00		
73	48	FORM5HEIM								0,00	0,00	0,00	0,00		
74	48	FORM5GAST								0,00	0,00	0,00	0,00		
75	4	DIFFFORM5	5	1		1	5	1		2		1,50	2,33	1,00	1,50
76	48	VIELETORE									0,00	0,00	0,00	0,00	
77	48	AVG1PLATZHEIM									0,00	0,00	0,00	0,00	
78	48	AVG1PLATZGAST									0,00	0,00	0,00	0,00	
79	10	DIFFAVG1	1		2	4		3			0,75	1,33	1,50	1,00	
80	24	AVG2PLATZHEIM	1			2					0,25	0,67	0,00	0,30	
81	48	AVG2PLATZGAST									0,00	0,00	0,00	0,00	
82	14	DIFFAVG2					3		4		0,00	0,00	1,50	0,70	
83	28	AVG3PLATZHEIM	1			1					0,25	0,33	0,00	0,20	
84	48	AVG3PLATZGAST									0,00	0,00	0,00	0,00	
85	17	DIFFAVG3							4		0,00	0,00	2,00	0,40	
86	48	AVG4PLATZHEIM									0,00	0,00	0,00	0,00	
87	48	AVG4PLATZGAST									0,00	0,00	0,00	0,00	
88	35	DIFFAVG4					1				0,00	0,33	0,00	0,10	
89	24	AVG5PLATZHEIM	1			2					0,25	0,67	0,00	0,30	
90	48	AVG5PLATZGAST									0,00	0,00	0,00	0,00	
91	14	DIFFAVG5						3	4		0,00	0,00	1,50	0,70	
92	48	HAEUFIGSTEPLAZIERUNGHEIM									0,00	0,00	0,00	0,00	

93	48	HAEUFIGSTEPLAZIERUNGSGAST					0,00	0,00	0,00	0,00	0,00
94	28	ADDNATIONALTOP9HEIM	2				0,50	0,00	0,00	0,20	
95	48	ADDNATIONALTOP9GAST					0,00	0,00	0,00	0,00	
96	6	ADDNATIONALLAENDERSP.TOP9H.		1		5	0,00	0,33	2,50	1,10	
97	48	ADDNATIONALLAENDERSP.TOP9G.					0,00	0,00	0,00	0,00	
98	35	ADDNATIONALTOP9POSHEIM	1				0,25	0,00	0,00	0,10	
99	48	ADDNATIONALTOP9POSGAST					0,00	0,00	0,00	0,00	
100	35	ADDNAT.LAENDERSP.TOPPOS9H.		1			0,00	0,00	0,50	0,10	
101	48	ADDNAT.LAENDERSP.TOPPOS9G.					0,00	0,00	0,00	0,00	
102	48	ADDMITTELNATLSPHEIM					0,00	0,00	0,00	0,00	
103	48	ADDMITTELNATLSPGAST					0,00	0,00	0,00	0,00	
104	48	ADDLSPKTHEIM					0,00	0,00	0,00	0,00	
105	48	ADDLSPKTGAST					0,00	0,00	0,00	0,00	
106	48	ADDGESCHOSSENETOREHEIM					0,00	0,00	0,00	0,00	
107	48	ADDGESCHOSSENETOREGAST					0,00	0,00	0,00	0,00	
108	48	ADD DURCHGELASSENETOREHEIM					0,00	0,00	0,00	0,00	
109	48	ADD DURCHGELASSENETOREGAST					0,00	0,00	0,00	0,00	
110	48	ADDPOSITIONHEIM					0,00	0,00	0,00	0,00	
111	48	ADDPOSITIONGAST					0,00	0,00	0,00	0,00	
112	48	ADDSTADIONGROESSEHEIM					0,00	0,00	0,00	0,00	
113	48	ADDSTADIONGROESSEGAST					0,00	0,00	0,00	0,00	
114	48	ADD10ABSTIEGHEIM					0,00	0,00	0,00	0,00	
115	48	ADD10ABSTIEGGAST					0,00	0,00	0,00	0,00	
116	48	ADDBRAZILHEIM					0,00	0,00	0,00	0,00	
117	48	ADDBRAZILGAST					0,00	0,00	0,00	0,00	
118	48	ADDBRAZILNATHEIM					0,00	0,00	0,00	0,00	
119	48	ADDBRAZILNATGAST					0,00	0,00	0,00	0,00	
120	48	ADDGERMANYNATHEIM					0,00	0,00	0,00	0,00	
121	48	ADDGERMANYNATGAST					0,00	0,00	0,00	0,00	
122	48	ADDAUSTRIAHEIM					0,00	0,00	0,00	0,00	
123	48	ADDAUSTRIAGAST					0,00	0,00	0,00	0,00	
124	48	ADDAUSTRIANATHEIM					0,00	0,00	0,00	0,00	
125	48	ADDAUSTRIANATGAST					0,00	0,00	0,00	0,00	
126	16	ADDEWIGETABELLEHEIM	1		1	1	0,25	0,67	1,50	0,60	
127	35	ADDEWIGETABELLEGAST				1	0,00	0,00	0,50	0,10	
128	13	ADDUEFA1JAHRHEIM	8				2,00	0,00	0,00	0,80	
129	35	ADDUEFA1JAHRGAST				1	0,00	0,00	0,50	0,10	
130	35	ADDUEFA5JAHREHEIM				1	0,00	0,33	0,00	0,10	
131	48	ADDUEFA5JAHREGAST					0,00	0,00	0,00	0,00	
132	35	ADDPUNKTEHEIMHEIM	1				0,25	0,00	0,00	0,10	
133	48	ADDPUNKTEGASTGAST					0,00	0,00	0,00	0,00	
134	48	ABSTIEG1HEIM					0,00	0,00	0,00	0,00	
135	24	ABSTIEG1GAST				1	2	0,00	0,00	1,50	0,30
136	48	DIFFABSTIEG1					0,00	0,00	0,00	0,00	
137	48	ABSTIEG2HEIM					0,00	0,00	0,00	0,00	
138	17	ABSTIEG2GAST	1		1	1	0,00	0,33	1,00	0,40	
139	48	DIFFABSTIEG2					0,00	0,00	0,00	0,00	

140	48	ABSTIEG3HEIM								0,00	0,00	0,00	0,00	
141	48	ABSTIEG3GAST								0,00	0,00	0,00	0,00	
142	48	DIFFABSTIEG3								0,00	0,00	0,00	0,00	
143	48	ABSTIEG4HEIM								0,00	0,00	0,00	0,00	
144	28	ABSTIEG4GAST		1			1			0,25	0,33	0,00	0,20	
145	48	DIFFABSTIEG4								0,00	0,00	0,00	0,00	
146	48	ABSTIEG5HEIM								0,00	0,00	0,00	0,00	
147	48	ABSTIEG5GAST								0,00	0,00	0,00	0,00	
148	48	DIFFABSTIEG5								0,00	0,00	0,00	0,00	
149	12	DIFFPUNKTEHEIMGAST		6					2	1	1,50	0,00	1,00	0,90
150	35	DIFFPKTLS		1							0,25	0,00	0,00	0,10
151	48	DIFFGSTLS									0,00	0,00	0,00	0,00
152	48	DIFFKASSTLS									0,00	0,00	0,00	0,00
153	17	DIFFPOSLS		4							1,00	0,00	0,00	0,40
154	48	DIFFEWIGETABELLE									0,00	0,00	0,00	0,00
155	24	DIFFNAT		1			2				0,00	0,33	1,00	0,30
156	35	DIFNAT2								1	0,00	0,00	0,00	0,10
157	17	DIFFNATPOS				2	1	1			0,00	0,67	0,50	0,40
158	35	DIFNATPOS2		1							0,25	0,00	0,00	0,10
159	28	DIFFNATPOSMIT					1	1			0,00	0,00	1,00	0,20
160	48	DIFFSTADGR									0,00	0,00	0,00	0,00
161	48	DIFF10ABSTIEG									0,00	0,00	0,00	0,00
162	48	DIFFBRAZIL									0,00	0,00	0,00	0,00
163	48	DIFBRAZILN									0,00	0,00	0,00	0,00
164	28	DIFFGERMANYN		1			1				0,25	0,33	0,00	0,20
165	48	DIFFAUSTRIA									0,00	0,00	0,00	0,00
166	48	DIFFAUSTRIAN									0,00	0,00	0,00	0,00
167	10	DIFFUEFA1J		2	6	1		1			2,00	0,33	0,50	1,00
168	2	DIFFUEFA5J		9		4	1	4	3		2,25	1,67	2,00	2,10
169	48	DIFFHAEUFPLAZ									0,00	0,00	0,00	0,00

Wrapper mit Klassifizierer: „J48“ und Suchmethode: „LinearForwardSelection“

NR	Platz	Name	24	35	46	57	25	36	47	26	37	27	3SD	4SD	5SD	Ges		
1	65	HEIM											0,00	0,00	0,00	0,00		
2	65	GAST											0,00	0,00	0,00	0,00		
3	65	POSHEIM											0,00	0,00	0,00	0,00		
4	65	PLUSMINUSHEIM											0,00	0,00	0,00	0,00		
5	65	SIEGEHEIM											0,00	0,00	0,00	0,00		
6	65	REMISHEIM											0,00	0,00	0,00	0,00		
7	65	NIEDERLAGENHEIM											0,00	0,00	0,00	0,00		
8	65	PUNKTEHEIM											0,00	0,00	0,00	0,00		
9	65	TDHEIM											0,00	0,00	0,00	0,00		
10	39	POSGAST											1	0,00	0,00	0,00	0,10	
11	65	PLUSMINUSGAST												0,00	0,00	0,00	0,00	
12	65	SIEGEGAST												0,00	0,00	0,00	0,00	
13	65	REMISGAST												0,00	0,00	0,00	0,00	
14	65	NIEDERLAGENGAST												0,00	0,00	0,00	0,00	
15	65	PUNKTEGAST												0,00	0,00	0,00	0,00	
16	18	TDGAST											1	1			0,50 0,33 0,00 0,30	
17	65	MONAT													0,00	0,00	0,00	0,00
18	65	JAHRESZEIT													0,00	0,00	0,00	0,00
19	65	TABDIFF													0,00	0,00	0,00	0,00
20	65	TABTENDENZ													0,00	0,00	0,00	0,00
21	65	TABDIFFFORM													0,00	0,00	0,00	0,00
22	65	TABDIFFFORMVERG													0,00	0,00	0,00	0,00
23	65	SIEGDIFF													0,00	0,00	0,00	0,00
24	65	REMISDIFF													0,00	0,00	0,00	0,00
25	65	NIEDDIFF													0,00	0,00	0,00	0,00
26	65	PUNKTEDIFF													0,00	0,00	0,00	0,00
27	25	TDDIFF											1				0,25 0,00 0,50 0,20	
28	65	TORDIFFFORM													0,00	0,00	0,00	0,00
29	65	TFHEIM													0,00	0,00	0,00	0,00
30	39	TFGAST													1	0,00 0,00 0,50 0,10		
31	65	HEIMSTARK													0,00	0,00	0,00	0,00
32	65	GASTSTARK													0,00	0,00	0,00	0,00
33	65	SIEGEHEIM													0,00	0,00	0,00	0,00
34	10	VERSUS1HEIM											2	1			0,75 0,33 0,00 0,40	
35	65	VERSUS1GAST													0,00	0,00	0,00	0,00
36	25	VERSUS1												1			0,25 0,33 0,00 0,20	
37	25	VERSUS2HEIM												1			0,25 0,33 0,00 0,20	
38	65	VERSUS2GAST													0,00	0,00	0,00	0,00
39	39	VERSUS2												1			0,25 0,00 0,00 0,10	
40	65	VERSUS3HEIM													0,00	0,00	0,00	0,00
41	65	VERSUS3GAST													0,00	0,00	0,00	0,00
42	39	VERSUS3													1	0,00 0,00 0,00 0,10		
43	65	VERSUS4HEIM													0,00	0,00	0,00	0,00
44	65	VERSUS4GAST													0,00	0,00	0,00	0,00
45	39	VERSUS4												1			0,25 0,00 0,00 0,10	

46	39	VERSUS5HEIM		1			0,25	0,00	0,00	0,10		
47	65	VERSUS5GAST					0,00	0,00	0,00	0,00		
48	25	VERSUS5		1		1		0,25	0,33	0,00	0,20	
49	65	TABGAST					0,00	0,00	0,00	0,00		
50	65	TABVERGLEICH					0,00	0,00	0,00	0,00		
51	65	FORM1HEIMHEIM					0,00	0,00	0,00	0,00		
52	65	FORM1GASTAUSW					0,00	0,00	0,00	0,00		
53	65	FORM1HEIM					0,00	0,00	0,00	0,00		
54	65	FORM1GAST					0,00	0,00	0,00	0,00		
55	3	DIFFFORM1		1		2	1	3	0,25	0,67	0,50	0,70
56	65	FORM2HEIMHEIM					0,00	0,00	0,00	0,00		
57	25	FORM2GASTAUSW		1		1			0,25	0,00	0,50	0,20
58	65	FORM2HEIM					0,00	0,00	0,00	0,00		
59	65	FORM2GAST					0,00	0,00	0,00	0,00		
60	10	DIFFFORM2		2		1	1		0,50	0,33	0,50	0,40
61	65	FORM3HEIMHEIM					0,00	0,00	0,00	0,00		
62	39	FORM3GASTAUSW					1		0,00	0,00	0,50	0,10
63	65	FORM3HEIM					0,00	0,00	0,00	0,00		
64	39	FORM3GAST		1					0,25	0,00	0,00	0,10
65	39	DIFFFORM3						1	0,00	0,00	0,50	0,10
66	65	FORM4HEIMHEIM					0,00	0,00	0,00	0,00		
67	3	FORM4GASTAUSW			1	3	2	1	0,25	1,00	1,00	0,70
68	65	FORM4HEIM					0,00	0,00	0,00	0,00		
69	65	FORM4GAST					0,00	0,00	0,00	0,00		
70	65	DIFFFORM4					0,00	0,00	0,00	0,00		
71	65	FORM5HEIMHEIM					0,00	0,00	0,00	0,00		
72	10	FORM5GASTAUSW		4					1,00	0,00	0,00	0,40
73	39	FORM5HEIM		1					0,25	0,00	0,00	0,10
74	65	FORM5GAST					0,00	0,00	0,00	0,00		
75	7	DIFFFORM5		1	2	1	1	1	0,75	0,67	0,50	0,60
76	65	VIELETORE					0,00	0,00	0,00	0,00		
77	65	AVG1PLATZHEIM					0,00	0,00	0,00	0,00		
78	18	AVG1PLATZGAST			1	1		1	0,25	0,33	0,50	0,30
79	3	DIFFAVG1		1	5		1		1,50	0,33	0,00	0,70
80	65	AVG2PLATZHEIM					0,00	0,00	0,00	0,00		
81	65	AVG2PLATZGAST					0,00	0,00	0,00	0,00		
82	39	DIFFAVG2					1		0,00	0,00	0,50	0,10
83	65	AVG3PLATZHEIM					0,00	0,00	0,00	0,00		
84	65	AVG3PLATZGAST					0,00	0,00	0,00	0,00		
85	39	DIFFAVG3			1				0,25	0,00	0,00	0,10
86	65	AVG4PLATZHEIM					0,00	0,00	0,00	0,00		
87	25	AVG4PLATZGAST		2					0,50	0,00	0,00	0,20
88	10	DIFFAVG4		1		1	2		0,25	0,33	1,00	0,40
89	65	AVG5PLATZHEIM					0,00	0,00	0,00	0,00		
90	65	AVG5PLATZGAST					0,00	0,00	0,00	0,00		
91	25	DIFFAVG5		1			1		0,25	0,00	0,50	0,20
92	65	HAEUFIGSTEPLAZIERUNGHEIM							0,00	0,00	0,00	0,00

93	25	HAEUFIGSTEPLAZIERUNGGAST			1	1	0,00	0,33	0,50	0,20
94	65	ADDNATIONALTOP9HEIM					0,00	0,00	0,00	0,00
95	65	ADDNATIONALTOP9GAST					0,00	0,00	0,00	0,00
96	65	ADDNATIONALLAENDERSP.TOP9H.					0,00	0,00	0,00	0,00
97	65	ADDNATIONALLAENDERSP.TOP9G.					0,00	0,00	0,00	0,00
98	65	ADDNATIONALTOP9POSHEIM					0,00	0,00	0,00	0,00
99	65	ADDNATIONALTOP9POSGAST					0,00	0,00	0,00	0,00
100	39	ADDNAT.LAENDERSP.TOPPOS9H.	1				0,25	0,00	0,00	0,10
101	25	ADDNAT.LAENDERSP.TOPPOS9G.	1	1			0,25	0,33	0,00	0,20
102	65	ADDMITTELNATLSPHEIM					0,00	0,00	0,00	0,00
103	39	ADDMITTELNATLSPGAST				1	0,00	0,00	0,00	0,10
104	65	ADDLSPKTHEIM					0,00	0,00	0,00	0,00
105	25	ADDLSPKTGAST	2				0,50	0,00	0,00	0,20
106	65	ADDGESCHOSSENETOREHEIM					0,00	0,00	0,00	0,00
107	39	ADDGESCHOSSENETOREGAST				1	0,00	0,00	0,50	0,10
108	65	ADD DURCHGELASSENETOREHEIM					0,00	0,00	0,00	0,00
109	65	ADD DURCHGELASSENETOREGAST					0,00	0,00	0,00	0,00
110	65	ADDPOSITIONHEIM					0,00	0,00	0,00	0,00
111	65	ADDPOSITIONGAST					0,00	0,00	0,00	0,00
112	65	ADDSTADIONGROESSEHEIM					0,00	0,00	0,00	0,00
113	39	ADDSTADIONGROESSEGAST			1		0,00	0,33	0,00	0,10
114	65	ADD10ABSTIEGHEIM					0,00	0,00	0,00	0,00
115	65	ADD10ABSTIEGGAST					0,00	0,00	0,00	0,00
116	65	ADDBRAZILHEIM					0,00	0,00	0,00	0,00
117	65	ADDBRAZILGAST					0,00	0,00	0,00	0,00
118	39	ADDBRAZILNATHEIM	1				0,25	0,00	0,00	0,10
119	39	ADDBRAZILNATGAST	1				0,25	0,00	0,00	0,10
120	65	ADDGERMANYNATHEIM					0,00	0,00	0,00	0,00
121	25	ADDGERMANYNATGAST			1		1	0,25	0,00	0,00
122	65	ADDAUSTRIAHEIM						0,00	0,00	0,00
123	65	ADDAUSTRIAGAST						0,00	0,00	0,00
124	65	ADDAUSTRIANATHEIM						0,00	0,00	0,00
125	65	ADDAUSTRIANATGAST						0,00	0,00	0,00
126	65	ADDEWIGETABELLEHEIM						0,00	0,00	0,00
127	10	ADDEWIGETABELLEGAST		3			1	0,75	0,00	0,00
128	65	ADDUEFA1JAHRHEIM						0,00	0,00	0,00
129	10	ADDUEFA1JAHRGAST	1	1	1	1		0,75	0,33	0,00
130	65	ADDUEFA5JAHREHEIM						0,00	0,00	0,00
131	65	ADDUEFA5JAHREGAST						0,00	0,00	0,00
132	65	ADDPUNKTEHEIMHEIM						0,00	0,00	0,00
133	39	ADDPUNKTEGASTGAST			1			0,25	0,00	0,00
134	65	ABSTIEG1HEIM						0,00	0,00	0,00
135	39	ABSTIEG1GAST			1			0,25	0,00	0,00
136	39	DIFFABSTIEG1				1		0,00	0,33	0,00
137	65	ABSTIEG2HEIM						0,00	0,00	0,00
138	18	ABSTIEG2GAST	1	1		1		0,50	0,33	0,00
139	39	DIFFABSTIEG2			1			0,25	0,00	0,00
								0,10		

140	65	ABSTIEG3HEIM									0,00	0,00	0,00	0,00			
141	39	ABSTIEG3GAST		1							0,25	0,00	0,00	0,10			
142	65	DIFFABSTIEG3									0,00	0,00	0,00	0,00			
143	65	ABSTIEG4HEIM									0,00	0,00	0,00	0,00			
144	18	ABSTIEG4GAST		1						2	0,25	0,00	0,00	0,30			
145	65	DIFFABSTIEG4									0,00	0,00	0,00	0,00			
146	65	ABSTIEG5HEIM									0,00	0,00	0,00	0,00			
147	10	ABSTIEG5GAST		3	1						1,00	0,00	0,00	0,40			
148	39	DIFFABSTIEG5				1					0,00	0,33	0,00	0,10			
149	18	DIFFPUNKTEHEIMGAST		1	1					1	0,50	0,00	0,00	0,30			
150	25	DIFFPKTLS			2						0,50	0,00	0,00	0,20			
151	25	DIFFGSTLS		1		1					0,50	0,00	0,00	0,20			
152	8	DIFFKASSTLS		1		2	1	1			0,75	0,33	0,50	0,50			
153	18	DIFFPOSLS				3					0,75	0,00	0,00	0,30			
154	65	DIFFEWIGETABELLE									0,00	0,00	0,00	0,00			
155	8	DIFFNAT			1	1	2			1	0,25	1,00	0,00	0,50			
156	18	DIFNAT2				1				1	0,00	0,33	0,50	0,30			
157	2	DIFFNATPOS		3	7	2	2	1			3,00	1,00	0,00	1,50			
158	39	DIFNATPOS2				1					0,00	0,33	0,00	0,10			
159	3	DIFFNATPOSMIT			2	2				2	1	1,00	0,00	1,00	0,70		
160	65	DIFFSTADGR			1		1					0,00	0,00	0,00	0,00		
161	25	DIFF10ABSTIEG										0,25	0,33	0,00	0,20		
162	65	DIFFBRAZIL										0,00	0,00	0,00	0,00		
163	65	DIFBRAZILN										0,00	0,00	0,00	0,00		
164	1	DIFFGERMANYN		6	4	1	8	8	10	3	10	10	7	4,75	7,00	10,00	6,70
165	65	DIFFAUSTRIA										0,00	0,00	0,00	0,00		
166	65	DIFFAUSTRIAN										0,00	0,00	0,00	0,00		
167	39	DIFFUEFA1J				1						0,25	0,00	0,00	0,10		
168	10	DIFFUEFA5J		1			1	1		1		0,25	0,67	0,50	0,40		
169	65	DIFFHAEUFPLAZ										0,00	0,00	0,00	0,00		

Wrapper mit Klassifizierer: „J48“ und Suchmethode: „BestFirst-Backward Elimination“

NR	Platz	Name	24	35	46	57	25	36	47	26	37	27	3SD	4SD	5SD	Ges
1	1	HEIM	10	10	9	10	10	10	9	10	8	10	9,75	9,67	9,00	9,60
2	1	GAST	10	8	10	9	9	10	10	10	10	10	9,25	9,67	10,00	9,60
3	36	POSHEIM	6	7	9	7	7	9	7	6	8	8	7,25	7,67	7,00	7,40
4	62	PLUSMINUSHEIM	7	5	6	8	7	9	5	8	8	7	6,50	7,00	8,00	7,00
5	118	SIEGEHEIM	6	6	10	6	5	7	7	6	4	5	7,00	6,33	5,00	6,20
6	14	REMISHEIM	7	8	6	9	9	8	7	8	8	9	7,50	8,00	8,00	7,90
7	113	NIEDERLAGENHEIM	1	8	7	2	10	6	6	9	7	7	4,50	7,33	8,00	6,30
8	78	PUNKTEHEIM	6	4	8	6	6	7	8	8	8	6	6,00	7,00	8,00	6,70
9	147	TDHEIM	6	4	9	6	5	5	5	5	6	5	6,25	5,00	5,50	5,60
10	62	POSGAST	4	7	6	10	6	8	7	6	6	10	6,75	7,00	6,00	7,00
11	30	PLUSMINUSGAST	4	8	7	8	9	9	7	9	6	8	6,75	8,33	7,50	7,50
12	36	SIEGEGAST	7	7	7	8	7	10	7	7	7	7	7,25	8,00	7,00	7,40
13	36	REMISGAST	9	8	6	8	9	8	9	1	9	7	7,75	8,67	5,00	7,40
14	43	NIEDERLAGENGAST	7	6	5	7	10	5	8	7	8	9	6,25	7,67	7,50	7,20
15	98	PUNKTEGAST	8	7	6	7	7	7	6	4	8	5	7,00	6,67	6,00	6,50
16	50	TDGAST	5	8	9	8	4	4	10	10	8	5	7,50	6,00	9,00	7,10
17	88	MONAT	7	7	6	5	7	6	6	7	9	6	6,25	6,33	8,00	6,60
18	68	JAHRESZEIT	6	9	8	6	8	6	7	6	7	6	7,25	7,00	6,50	6,90
19	50	TABDIFF	4	7	9	9	8	7	8	5	8	6	7,25	7,67	6,50	7,10
20	78	TABTENDENZ	7	8	8	5	9	6	5	6	5	8	7,00	6,67	5,50	6,70
21	68	TABDIFFFORM	4	7	8	8	5	8	8	6	7	8	6,75	7,00	6,50	6,90
22	98	TABDIFFFORMVERG	4	6	6	7	4	8	8	8	5	9	5,75	6,67	6,50	6,50
23	14	SIEGDIFF	8	8	10	10	7	7	9	7	6	7	9,00	7,67	6,50	7,90
24	7	REMISDIFF	6	8	10	9	8	9	8	10	6	8	8,25	8,33	8,00	8,20
25	43	NIEDDIFF	8	7	6	9	7	4	9	10	7	5	7,50	6,67	8,50	7,20
26	88	PUNKTEDIFF	7	4	8	6	8	6	7	6	6	8	6,25	7,00	6,00	6,60
27	43	TDDIFF	7	6	6	7	10	7	6	8	6	9	6,50	7,67	7,00	7,20
28	126	TORDIFFFORM	7	7	5	7	7	7	6	3	6	5	6,50	6,67	4,50	6,00
29	25	TFHEIM	7	8	10	7	7	7	9	6	8	7	8,00	7,67	7,00	7,60
30	50	TFGAST	5	7	7	9	5	7	6	7	10	8	7,00	6,00	8,50	7,10
31	73	HEIMSTARK	6	6	7	7	6	8	6	5	8	9	6,50	6,67	6,50	6,80
32	19	GASTSTARK	7	6	9	8	6	7	10	9	8	8	7,50	7,67	8,50	7,80
33	43	SIEGEHEIM	7	8	8	9	6	7	6	6	8	7	8,00	6,33	7,00	7,20
34	73	VERSUS1HEIM	8	8	6	7	8	7	6	6	5	7	7,25	7,00	5,50	6,80
35	139	VERSUS1GAST	5	3	6	3	7	8	4	7	8	7	4,25	6,33	7,50	5,80
36	78	VERSUS1	8	7	8	6	7	10	4	7	5	5	7,25	7,00	6,00	6,70
37	50	VERSUS2HEIM	8	8	7	7	8	6	6	9	5	7	7,50	6,67	7,00	7,10
38	139	VERSUS2GAST	7	6	7	6	7	6	5	5	5	4	6,50	6,00	5,00	5,80
39	113	VERSUS2	6	6	6	7	6	8	5	3	8	8	6,25	6,33	5,50	6,30
40	78	VERSUS3HEIM	5	8	8	6	5	9	8	7	5	6	6,75	7,33	6,00	6,70
41	155	VERSUS3GAST	4	5	3	7	4	6	3	6	8	7	4,75	4,33	7,00	5,30
42	164	VERSUS3	5	4	4	5	4	6	7	3	2	5	4,50	5,67	2,50	4,50
43	36	VERSUS4HEIM	5	9	9	7	9	8	6	6	7	8	7,50	7,67	6,50	7,40
44	88	VERSUS4GAST	6	7	5	7	7	7	7	6	5	9	6,25	7,00	5,50	6,60
45	50	VERSUS4	6	4	9	6	10	6	9	6	8	7	6,25	8,33	7,00	7,10

46	142	VERSUS5HEIM	7	6	5	6	6	5	5	6	6	5	6,00	5,33	6,00	5,70
47	88	VERSUS5GAST	6	9	7	6	5	8	8	5	4	8	7,00	7,00	4,50	6,60
48	142	VERSUS5	5	7	4	4	7	6	5	5	6	8	5,00	6,00	5,50	5,70
49	4	TABGAST	9	10	9	6	7	8	10	8	10	8	8,50	8,33	9,00	8,50
50	40	TABVERGLEICH	9	7	8	7	5	7	7	7	8	8	7,75	6,33	7,50	7,30
51	142	FORM1HEIMHEIM	2	5	8	5	4	7	5	8	7	6	5,00	5,33	7,50	5,70
52	78	FORM1GASTAUSW	6	9	4	9	6	9	6	6	6	6	7,00	7,00	6,00	6,70
53	50	FORM1HEIM	4	6	7	6	8	8	8	9	7	8	5,75	8,00	8,00	7,10
54	107	FORM1GAST	3	7	6	7	7	8	8	8	4	6	5,75	7,67	6,00	6,40
55	88	DIFFFORM1	6	7	7	4	9	7	5	7	6	8	6,00	7,00	6,50	6,60
56	50	FORM2HEIMHEIM	8	6	5	8	4	8	9	9	6	8	6,75	7,00	7,50	7,10
57	113	FORM2GASTAUSW	8	5	5	5	7	7	5	7	7	7	5,75	6,33	7,00	6,30
58	73	FORM2HEIM	7	5	8	7	8	7	7	7	7	5	6,75	7,33	7,00	6,80
59	152	FORM2GAST	5	7	5	5	3	6	3	6	8	6	5,50	4,00	7,00	5,40
60	78	DIFFFORM2	5	9	7	6	7	5	7	6	8	7	6,75	6,33	7,00	6,70
61	107	FORM3HEIMHEIM	5	6	6	5	6	7	5	8	8	8	5,50	6,00	8,00	6,40
62	12	FORM3GASTAUSW	9	7	9	9	7	7	10	7	8	7	8,50	8,00	7,50	8,00
63	147	FORM3HEIM	5	5	6	6	7	6	7	4	5	5	5,50	6,67	4,50	5,60
64	120	FORM3GAST	5	6	6	6	7	7	6	4	7	7	5,75	6,67	5,50	6,10
65	5	DIFFFORM3	9	9	7	8	9	10	9	9	8	6	8,25	9,33	8,50	8,40
66	126	FORM4HEIMHEIM	3	5	7	7	4	6	8	6	7	7	5,50	6,00	6,50	6,00
67	107	FORM4GASTAUSW	5	7	7	5	4	9	7	6	9	5	6,00	6,67	7,50	6,40
68	120	FORM4HEIM	6	4	7	8	4	9	5	5	7	6	6,25	6,00	6,00	6,10
69	130	FORM4GAST	5	6	4	6	4	4	10	4	8	8	5,25	6,00	6,00	5,90
70	50	DIFFFORM4	4	8	7	6	6	10	7	10	7	6	6,25	7,67	8,50	7,10
71	113	FORM5HEIMHEIM	7	8	6	9	4	4	7	5	8	5	7,50	5,00	6,50	6,30
72	62	FORM5GASTAUSW	9	7	10	5	5	6	7	7	8	6	7,75	6,00	7,50	7,00
73	130	FORM5HEIM	6	6	5	7	4	7	5	5	8	6	6,00	5,33	6,50	5,90
74	158	FORM5GAST	5	5	3	3	8	2	6	8	5	5	4,00	5,33	6,50	5,00
75	107	DIFFFORM5	5	8	7	3	6	8	5	6	8	8	5,75	6,33	7,00	6,40
76	169	VIELETORE	0	0	0	0	0	0	0	0	0	0	0,00	0,00	0,00	0,00
77	163	AVG1PLATZHEIM	4	1	2	5	5	4	8	5	6	6	3,00	5,67	5,50	4,60
78	43	AVG1PLATZGAST	6	7	6	6	9	9	8	6	8	7	6,25	8,67	7,00	7,20
79	98	DIFFAVG1	5	5	9	7	8	4	8	6	6	7	6,50	6,67	6,00	6,50
80	168	AVG2PLATZHEIM	2	0	5	2	4	2	2	2	4	2	2,25	2,67	3,00	2,50
81	166	AVG2PLATZGAST	5	5	5	1	3	4	4	4	3	4	4,00	3,67	3,50	3,80
82	167	DIFFAVG2	5	1	4	1	5	2	2	1	5	2	2,75	3,00	3,00	2,80
83	151	AVG3PLATZHEIM	6	3	4	7	9	5	3	6	7	5	5,00	5,67	6,50	5,50
84	155	AVG3PLATZGAST	5	5	6	6	4	6	4	6	5	6	5,50	4,67	5,50	5,30
85	88	DIFFAVG3	8	7	6	9	8	7	7	3	6	5	7,50	7,33	4,50	6,60
86	152	AVG4PLATZHEIM	5	4	6	3	6	4	7	6	7	6	4,50	5,67	6,50	5,40
87	120	AVG4PLATZGAST	7	9	5	4	4	5	6	7	6	8	6,25	5,00	6,50	6,10
88	130	DIFFAVG4	6	3	8	6	6	8	5	5	6	6	5,75	6,33	5,50	5,90
89	88	AVG5PLATZHEIM	5	3	5	7	9	5	8	8	9	7	5,00	7,33	8,50	6,60
90	130	AVG5PLATZGAST	7	6	3	8	6	7	5	7	6	4	6,00	6,00	6,50	5,90
91	62	DIFFAVG5	6	3	9	6	7	8	8	9	8	6	6,00	7,67	8,50	7,00
92	73	HAEUFIGSTEPLAZIERUNGHEIM	4	9	8	6	6	9	8	6	7	5	6,75	7,67	6,50	6,80

93	23	HAEUFIGSTEPLAZIERUNGSGAST	5	8	6	7	7	10	8	10	8	8	6,50	8,33	9,00	7,70
94	50	ADDNATIONALTOP9HEIM	5	8	5	5	9	9	6	9	7	8	5,75	8,00	8,00	7,10
95	147	ADDNATIONALTOP9GAST	9	9	6	5	6	5	5	2	5	4	7,25	5,33	3,50	5,60
96	160	ADDNATIONALLAENDERSP.TOP9H.	7	3	6	4	6	3	4	4	8	4	5,00	4,33	6,00	4,90
97	162	ADDNATIONALLAENDERSP.TOP9G.	2	9	3	6	2	3	6	7	5	5	5,00	3,67	6,00	4,80
98	40	ADDNATIONALTOP9POSHEIM	7	8	8	6	8	7	6	7	7	9	7,25	7,00	7,00	7,30
99	73	ADDNATIONALTOP9POSGAST	7	4	7	7	8	7	7	7	6	8	6,25	7,33	6,50	6,80
100	118	ADDNAT.LAENDERSP.TOPPOS9H.	6	8	6	2	6	9	4	5	9	7	5,50	6,33	7,00	6,20
101	78	ADDNAT.LAENDERSP.TOPPOS9G.	5	6	5	6	7	6	8	8	7	9	5,50	7,00	7,50	6,70
102	120	ADDMITTELNATLSPHEIM	7	6	7	9	6	7	7	4	4	4	7,25	6,67	4,00	6,10
103	165	ADDMITTELNATLSPGAST	5	4	7	5	4	4	4	1	5	4	5,25	4,00	3,00	4,30
104	142	ADDLSPKTHEIM	7	3	4	8	7	4	6	5	7	6	5,50	5,67	6,00	5,70
105	78	ADDLSPKTGAST	5	6	6	9	5	9	8	6	6	7	6,50	7,33	6,00	6,70
106	147	ADDGESCHOSSENETOREHEIM	6	7	5	5	6	6	5	4	7	5	5,75	5,67	5,50	5,60
107	30	ADDGESCHOSSENETOREGAST	8	4	6	8	7	7	10	8	8	9	6,50	8,00	8,00	7,50
108	120	ADDDURCHGELASSENETOREHEIM	5	3	6	8	8	5	5	8	8	5	5,50	6,00	8,00	6,10
109	152	ADDDURCHGELASSENETOREGAST	2	4	6	4	6	7	6	7	5	7	4,00	6,33	6,00	5,40
110	120	ADDPOSITIONHEIM	6	5	5	9	4	7	6	5	7	7	6,25	5,67	6,00	6,10
111	126	ADDPOSITIONGAST	6	4	8	5	5	7	8	3	7	7	5,75	6,67	5,00	6,00
112	155	ADDSTADIONGROESSEHEIM	4	5	5	5	7	3	5	6	9	4	4,75	5,00	7,50	5,30
113	160	ADDSTADIONGROESSEGAST	6	3	6	4	6	4	7	4	6	3	4,75	5,67	5,00	4,90
114	43	ADD10ABSTIEGHEIM	8	7	6	7	8	7	8	8	7	6	7,00	7,67	7,50	7,20
115	98	ADD10ABSTIEGGAST	5	8	6	8	8	4	8	8	4	6	6,75	6,67	6,00	6,50
116	25	ADDBRAZILHEIM	7	10	6	9	8	9	6	7	5	9	8,00	7,67	6,00	7,60
117	9	ADDBRAZILGAST	6	7	8	10	7	8	9	7	9	10	7,75	8,00	8,00	8,10
118	14	ADDBRAZILNATHEIM	7	9	9	6	8	8	8	7	8	9	7,75	8,00	7,50	7,90
119	12	ADDBRAZILNATGAST	9	7	7	6	9	6	9	9	9	9	7,25	8,00	9,00	8,00
120	30	ADDGERMANYNATHEIM	3	9	9	8	8	5	9	7	9	8	7,25	7,33	8,00	7,50
121	68	ADDGERMANYNATGAST	7	8	8	4	8	9	3	8	7	7	6,75	6,67	7,50	6,90
122	19	ADDAUSTRIAHEIM	6	7	8	8	7	8	9	7	9	9	7,25	8,00	8,00	7,80
123	107	ADDAUSTRIAGAST	5	4	7	6	7	7	8	5	7	8	5,50	7,33	6,00	6,40
124	3	ADDAUSTRIANATHEIM	10	7	7	10	9	8	8	8	10	9	8,50	8,33	9,00	8,60
125	5	ADDAUSTRIANATGAST	6	9	10	9	8	7	10	7	10	8	8,50	8,33	8,50	8,40
126	88	ADDEWIGETABELLEHEIM	8	5	5	6	6	8	6	7	7	8	6,00	6,67	7,00	6,60
127	68	ADDEWIGETABELLEGAST	3	6	6	8	8	9	8	7	7	7	5,75	8,33	7,00	6,90
128	14	ADDUEFA1JAHRHEIM	6	7	9	8	8	7	9	9	9	7	7,50	8,00	9,00	7,90
129	62	ADDUEFA1JAHRGAST	9	8	8	6	6	5	3	8	9	8	7,75	4,67	8,50	7,00
130	98	ADDUEFA5JAHREHEIM	7	8	7	6	4	6	9	6	5	7	7,00	6,33	5,50	6,50
131	130	ADDUEFA5JAHREGAST	5	4	7	8	6	7	7	4	6	5	6,00	6,67	5,00	5,90
132	50	ADDPUNKTEHEIMHEIM	9	7	7	7	8	7	4	9	6	7	7,50	6,33	7,50	7,10
133	50	ADDPUNKTEGASTGAST	9	7	8	4	7	6	6	8	10	6	7,00	6,33	9,00	7,10
134	126	ABSTIEG1HEIM	6	6	3	8	8	5	5	7	6	6	5,75	6,00	6,50	6,00
135	50	ABSTIEG1GAST	6	9	7	6	5	9	6	6	8	9	7,00	6,67	7,00	7,10
136	78	DIFFABSTIEG1	6	7	5	6	6	8	8	8	9	4	6,00	7,33	8,50	6,70
137	142	ABSTIEG2HEIM	6	4	3	5	8	5	6	6	7	7	4,50	6,33	6,50	5,70
138	130	ABSTIEG2GAST	7	3	7	6	7	5	5	8	6	5	5,75	5,67	7,00	5,90
139	130	DIFFABSTIEG2	8	3	6	5	7	7	5	3	8	7	5,50	6,33	5,50	5,90

140	139	ABSTIEG3HEIM
141	68	ABSTIEG3GAST
142	88	DIFFABSTIEG3
143	107	ABSTIEG4HEIM
144	98	ABSTIEG4GAST
145	130	DIFFABSTIEG4
146	98	ABSTIEG5HEIM
147	98	ABSTIEG5GAST
148	113	DIFFABSTIEG5
149	25	DIFFPUNKTEHEIMGAST
150	40	DIFFPKTLS
151	19	DIFFGSTLS
152	78	DIFFKASSTLS
153	62	DIFFPOSLS
154	88	DIFFEWIGETABELLE
155	7	DIFFNAT
156	19	DIFNAT2
157	25	DIFFNATPOS
158	98	DIFNATPOS2
159	30	DIFFNATPOSMIT
160	158	DIFFSTADGR
161	23	DIFF10ABSTIEG
162	14	DIFFBRAZIL
163	30	DIFBRAZILN
164	9	DIFFGERMANYN
165	25	DIFFAUSTRIA
166	9	DIFFAUSTRIAN
167	30	DIFFUEFA1J
168	130	DIFFUEFA5J
169	43	DIFFHAEUFPLAZ

7	3	6	6	7	8	4	8	4	5	5,50	6,33	6,00	5,80
8	4	7	8	9	7	6	5	8	7	6,75	7,33	6,50	6,90
7	5	8	8	7	6	8	7	6	4	7,00	7,00	6,50	6,60
6	4	7	6	9	6	4	6	8	8	5,75	6,33	7,00	6,40
7	7	8	5	8	7	6	5	5	7	6,75	7,00	5,00	6,50
7	4	6	6	3	6	8	4	8	7	5,75	5,67	6,00	5,90
7	4	6	6	3	6	8	4	8	7	5,75	5,67	6,00	5,90
6	6	6	8	6	6	6	5	7	9	6,50	6,00	6,00	6,50
7	6	5	6	7	8	5	6	8	7	6,00	6,67	7,00	6,50
7	7	5	7	6	7	7	7	4	6	6,50	6,67	5,50	6,30
4	9	9	10	8	6	7	6	8	9	8,00	7,00	7,00	7,60
5	8	9	7	8	7	10	4	9	6	7,25	8,33	6,50	7,30
7	7	7	10	9	7	7	9	8	7	7,75	7,67	8,50	7,80
8	4	7	7	6	8	8	6	7	6	6,50	7,33	6,50	6,70
4	7	7	10	8	6	5	7	9	7	7,00	6,33	8,00	7,00
6	7	7	7	6	7	8	4	9	5	6,75	7,00	6,50	6,60
8	8	7	9	8	9	10	8	9	6	8,00	9,00	8,50	8,20
7	8	8	9	5	9	6	9	8	9	8,00	6,67	8,50	7,80
8	7	7	9	7	7	9	5	8	9	7,75	7,67	6,50	7,60
6	6	5	7	9	4	7	8	7	6	6,00	6,67	7,50	6,50
4	7	8	7	7	8	9	9	8	8	6,50	8,00	8,50	7,50
6	7	6	6	4	2	2	6	5	6	6,25	2,67	5,50	5,00
8	6	6	8	8	10	8	7	8	8	7,00	8,67	7,50	7,70
8	7	8	8	7	8	10	6	8	9	7,75	8,33	7,00	7,90
7	8	9	4	5	8	10	7	8	9	7,00	7,67	7,50	7,50
1	9	8	10	8	9	7	10	10	9	7,00	8,00	10,00	8,10
5	8	8	9	6	7	8	8	7	10	7,50	7,00	7,50	7,60
6	9	9	9	8	8	9	6	10	7	8,25	8,33	8,00	8,10
6	8	8	7	6	7	10	6	10	7	7,25	7,67	8,00	7,50
1	5	7	6	8	9	9	5	5	4	4,75	8,67	5,00	5,90
7	7	7	9	10	7	6	6	7	6	7,50	7,67	6,50	7,20

Gesamtergebnis Wrapper mit Klassifizierer: „J48“ über alle 3 Suchmethoden

NR	Platz	Name	Gesamt	Prozent
1	3	HEIM	3,20	32,00
2	3	GAST	3,20	32,00
3	51	POSHEIM	2,47	24,67
4	79	PLUSMINUSHEIM	2,33	23,33
5	125	SIEGEHEIM	2,07	20,67
6	29	REMISHEIM	2,63	26,33
7	114	NIEDERLAGENHEIM	2,13	21,33
8	96	PUNKTEHEIM	2,23	22,33
9	149	TDHEIM	1,87	18,67
10	72	POSGAST	2,37	23,67
11	45	PLUSMINUSGAST	2,50	25,00
12	40	SIEGEGAST	2,53	25,33
13	51	REMISGAST	2,47	24,67
14	61	NIEDERLAGENGAST	2,40	24,00
15	109	PUNKTEGAST	2,17	21,67
16	51	TDGAST	2,47	24,67
17	101	MONAT	2,20	22,00
18	82	JAHRESZEIT	2,30	23,00
19	72	TABDIFF	2,37	23,67
20	96	TABTENDENZ	2,23	22,33
21	82	TABDIFFFORM	2,30	23,00
22	109	TABDIFFFORMVERG	2,17	21,67
23	29	SIEGDIFF	2,63	26,33
24	18	REMISDIFF	2,73	27,33
25	61	NIEDDIFF	2,40	24,00
26	101	PUNKTEDIFF	2,20	22,00
27	35	TDDIFF	2,60	26,00
28	131	TORDIFFFORM	2,00	20,00
29	26	TFHEIM	2,67	26,67
30	61	TFGAST	2,40	24,00
31	92	HEIMSTARK	2,27	22,67
32	35	GASTSTARK	2,60	26,00
33	61	SIEGEHEIM	2,40	24,00
34	51	VERSUS1HEIM	2,47	24,67
35	82	VERSUS1GAST	2,30	23,00
36	82	VERSUS1	2,30	23,00
37	38	VERSUS2HEIM	2,57	25,67
38	82	VERSUS2GAST	2,30	23,00
39	114	VERSUS2	2,13	21,33
40	96	VERSUS3HEIM	2,23	22,33
41	153	VERSUS3GAST	1,80	18,00
42	163	VERSUS3	1,53	15,33
43	51	VERSUS4HEIM	2,47	24,67
44	101	VERSUS4GAST	2,20	22,00
45	61	VERSUS4	2,40	24,00

46	143	VERSUS5HEIM	1,93	19,33
47	101	VERSUS5GAST	2,20	22,00
48	138	VERSUS5	1,97	19,67
49	13	TABGAST	2,83	28,33
50	59	TABVERGLEICH	2,43	24,33
51	145	FORM1HEIMHEIM	1,90	19,00
52	96	FORM1GASTAUSW	2,23	22,33
53	72	FORM1HEIM	2,37	23,67
54	114	FORM1GAST	2,13	21,33
55	3	DIFFFORM1	3,20	32,00
56	72	FORM2HEIMHEIM	2,37	23,67
57	109	FORM2GASTAUSW	2,17	21,67
58	82	FORM2HEIM	2,30	23,00
59	153	FORM2GAST	1,80	18,00
60	10	DIFFFORM2	2,87	28,67
61	114	FORM3HEIMHEIM	2,13	21,33
62	22	FORM3GASTAUSW	2,70	27,00
63	149	FORM3HEIM	1,87	18,67
64	125	FORM3GAST	2,07	20,67
65	2	DIFFFORM3	3,53	35,33
66	131	FORM4HEIMHEIM	2,00	20,00
67	72	FORM4GASTAUSW	2,37	23,67
68	127	FORM4HEIM	2,03	20,33
69	138	FORM4GAST	1,97	19,67
70	18	DIFFFORM4	2,73	27,33
71	123	FORM5HEIMHEIM	2,10	21,00
72	51	FORM5GASTAUSW	2,47	24,67
73	131	FORM5HEIM	2,00	20,00
74	159	FORM5GAST	1,67	16,67
75	13	DIFFFORM5	2,83	28,33
76	169	VIELETORE	0,00	0,00
77	163	AVG1PLATZHEIM	1,53	15,33
78	45	AVG1PLATZGAST	2,50	25,00
79	18	DIFFAVG1	2,73	27,33
80	168	AVG2PLATZHEIM	0,93	9,33
81	166	AVG2PLATZGAST	1,27	12,67
82	167	DIFFAVG2	1,20	12,00
83	145	AVG3PLATZHEIM	1,90	19,00
84	157	AVG3PLATZGAST	1,77	17,67
85	72	DIFFAVG3	2,37	23,67
86	153	AVG4PLATZHEIM	1,80	18,00
87	123	AVG4PLATZGAST	2,10	21,00
88	114	DIFFAVG4	2,13	21,33
89	82	AVG5PLATZHEIM	2,30	23,00
90	138	AVG5PLATZGAST	1,97	19,67
91	29	DIFFAVG5	2,63	26,33
92	92	HAEUFIGSTEPLAZIERUNGHEIM	2,27	22,67

93	29	HAEUFIGSTEPLAZIERUNGGAST	2,63	26,33
94	59	ADDNATIONALTOP9HEIM	2,43	24,33
95	149	ADDNATIONALTOP9GAST	1,87	18,67
96	131	ADDNATIONALLAENDERSP.TOP9H.	2,00	20,00
97	162	ADDNATIONALLAENDERSP.TOP9G.	1,60	16,00
98	51	ADDNATIONALTOP9POSHEIM	2,47	24,67
99	92	ADDNATIONALTOP9POSGAST	2,27	22,67
100	114	ADDNAT.LAENDERSP.TOPPOS9H.	2,13	21,33
101	82	ADDNAT.LAENDERSP.TOPPOS9G.	2,30	23,00
102	127	ADDMITTELNATLSPHEIM	2,03	20,33
103	165	ADDMITTELNATLSPGAST	1,47	14,67
104	145	ADDLSPKTHEIM	1,90	19,00
105	82	ADDLSPKTGAST	2,30	23,00
106	149	ADDGESCHOSSENETOREHEIM	1,87	18,67
107	40	ADDGESCHOSSENETOREGAST	2,53	25,33
108	127	ADDDURCHGELASSENETOREHEIM	2,03	20,33
109	153	ADDDURCHGELASSENETOREGAST	1,80	18,00
110	127	ADDPOSITIONHEIM	2,03	20,33
111	131	ADDPOSITIONGAST	2,00	20,00
112	157	ADDSTADIONGROESSEHEIM	1,77	17,67
113	159	ADDSTADIONGROESSEGAST	1,67	16,67
114	61	ADD10ABSTIEGHEIM	2,40	24,00
115	109	ADD10ABSTIEGGAST	2,17	21,67
116	40	ADDBRAZILHEIM	2,53	25,33
117	22	ADDBRAZILGAST	2,70	27,00
118	26	ADDBRAZILNATHEIM	2,67	26,67
119	22	ADDBRAZILNATGAST	2,70	27,00
120	45	ADDGERMANYNATHEIM	2,50	25,00
121	72	ADDGERMANYNATGAST	2,37	23,67
122	35	ADDAUSTRIAHEIM	2,60	26,00
123	114	ADDAUSTRIAGAST	2,13	21,33
124	10	ADDAUSTRIANATHEIM	2,87	28,67
125	15	ADDAUSTRIANATGAST	2,80	28,00
126	61	ADDEWIGETABELLEHEIM	2,40	24,00
127	51	ADDEWIGETABELLEGAST	2,47	24,67
128	9	ADDUEFA1JAHRHEIM	2,90	29,00
129	45	ADDUEFA1JAHRGAST	2,50	25,00
130	101	ADDUEFA5JAHREHEIM	2,20	22,00
131	138	ADDUEFA5JAHREGAST	1,97	19,67
132	61	ADDPUNKTEHEIMHEIM	2,40	24,00
133	61	ADDPUNKTEGASTGAST	2,40	24,00
134	131	ABSTIEG1HEIM	2,00	20,00
135	45	ABSTIEG1GAST	2,50	25,00
136	92	DIFFABSTIEG1	2,27	22,67
137	145	ABSTIEG2HEIM	1,90	19,00
138	101	ABSTIEG2GAST	2,20	22,00
139	131	DIFFABSTIEG2	2,00	20,00

140	143	ABSTIEG3HEIM	1,93	19,33
141	79	ABSTIEG3GAST	2,33	23,33
142	101	DIFFABSTIEG3	2,20	22,00
143	114	ABSTIEG4HEIM	2,13	21,33
144	79	ABSTIEG4GAST	2,33	23,33
145	138	DIFFABSTIEG4	1,97	19,67
146	109	ABSTIEG5HEIM	2,17	21,67
147	82	ABSTIEG5GAST	2,30	23,00
148	114	DIFFABSTIEG5	2,13	21,33
149	8	DIFFPUNKTEHEIMGAST	2,93	29,33
150	40	DIFFPKTLS	2,53	25,33
151	26	DIFFGSTLS	2,67	26,67
152	61	DIFFKASSTLS	2,40	24,00
153	38	DIFFPOSLS	2,57	25,67
154	101	DIFFEWIGETABELLE	2,20	22,00
155	7	DIFFNAT	3,00	30,00
156	18	DIFNAT2	2,73	27,33
157	6	DIFFNATPOS	3,17	31,67
158	96	DIFNATPOS2	2,23	22,33
159	15	DIFFNATPOSMIT	2,80	28,00
160	159	DIFFSTADGR	1,67	16,67
161	29	DIFF10ABSTIEG	2,63	26,33
162	29	DIFFBRAZIL	2,63	26,33
163	45	DIFBRAZILN	2,50	25,00
164	1	DIFFGERMANYN	5,00	50,00
165	40	DIFFAUSTRIA	2,53	25,33
166	22	DIFFAUSTRIAN	2,70	27,00
167	10	DIFFUEFA1J	2,87	28,67
168	15	DIFFUEFA5J	2,80	28,00
169	61	DIFFHAEUFPLAZ	2,40	24,00

Wrapper mit Klassifizierer: „SVM“ und Suchmethode: „RankSearch“

NR	Platz	Name	24	35	46	57	25	36	47	26	37	27	3SD	4SD	5SD	Ges
1	93	HEIM		1									0,25	0,00	0,00	0,10
2	115	GAST											0,00	0,00	0,00	0,00
3	93	POSHEIM					1						0,25	0,00	0,00	0,10
4	115	PLUSMINUSHEIM											0,00	0,00	0,00	0,00
5	115	SIEGEHEIM											0,00	0,00	0,00	0,00
6	115	REMISHEIM											0,00	0,00	0,00	0,00
7	93	NIEDERLAGENHEIM							1				0,00	0,33	0,00	0,10
8	115	PUNKTEHEIM											0,00	0,00	0,00	0,00
9	56	TDHEIM					4	1					1,25	0,00	0,00	0,50
10	115	POSGAST											0,00	0,00	0,00	0,00
11	115	PLUSMINUSGAST											0,00	0,00	0,00	0,00
12	64	SIEGEGAST					1	1	2				0,25	1,00	0,00	0,40
13	115	REMISGAST											0,00	0,00	0,00	0,00
14	115	NIEDERLAGENGAST											0,00	0,00	0,00	0,00
15	72	PUNKTEGAST		1					2				0,25	0,67	0,00	0,30
16	72	TDGAST		2	1								0,75	0,00	0,00	0,30
17	115	MONAT											0,00	0,00	0,00	0,00
18	115	JAHRESZEIT											0,00	0,00	0,00	0,00
19	93	TABDIFF				1							0,25	0,00	0,00	0,10
20	115	TABTENDENZ											0,00	0,00	0,00	0,00
21	115	TABDIFFFORM											0,00	0,00	0,00	0,00
22	115	TABDIFFFORMVERG											0,00	0,00	0,00	0,00
23	85	SIEGDIFF					1	1					0,25	0,33	0,00	0,20
24	115	REMISDIFF											0,00	0,00	0,00	0,00
25	115	NIEDDIFF											0,00	0,00	0,00	0,00
26	64	PUNKTEDIFF				2	2						1,00	0,00	0,00	0,40
27	32	TDDIFF				4	2	5		5	2		1,50	1,67	2,50	1,80
28	115	TORDIFFFORM											0,00	0,00	0,00	0,00
29	19	TFHEIM					10	1	3	2	8	2	2,75	4,33	2,00	2,80
30	115	TFGAST											0,00	0,00	0,00	0,00
31	115	HEIMSTARK											0,00	0,00	0,00	0,00
32	115	GASTSTARK											0,00	0,00	0,00	0,00
33	115	SIEGEHEIM											0,00	0,00	0,00	0,00
34	41	VERSUS1HEIM					5		5				1,25	1,67	0,00	1,00
35	72	VERSUS1GAST						1	2				0,25	0,67	0,00	0,30
36	93	VERSUS1					1						0,25	0,00	0,00	0,10
37	51	VERSUS2HEIM					1	3		2			1,00	0,67	0,00	0,60
38	85	VERSUS2GAST							2				0,00	0,67	0,00	0,20
39	93	VERSUS2					1						0,25	0,00	0,00	0,10
40	93	VERSUS3HEIM					1						0,25	0,00	0,00	0,10
41	115	VERSUS3GAST											0,00	0,00	0,00	0,00
42	93	VERSUS3							1				0,00	0,33	0,00	0,10
43	115	VERSUS4HEIM											0,00	0,00	0,00	0,00
44	115	VERSUS4GAST											0,00	0,00	0,00	0,00
45	72	VERSUS4					1		2				0,25	0,67	0,00	0,30

46	115	VERSUS5HEIM															0,00	0,00	0,00	0,00									
47	93	VERSUS5GAST														1	0,00	0,33	0,00	0,10									
48	93	VERSUS5													1		0,25	0,00	0,00	0,10									
49	115	TABGAST															0,00	0,00	0,00	0,00									
50	115	TABVERGLEICH															0,00	0,00	0,00	0,00									
51	115	FORM1HEIMHEIM															0,00	0,00	0,00	0,00									
52	51	FORM1GASTAUSW													1	4		1	0,25	1,33	0,50	0,60							
53	43	FORM1HEIM													1	2	3	1	1	0,25	1,67	1,00	0,80						
54	48	FORM1GAST													1	1	4	1		0,50	1,33	0,50	0,70						
55	6	DIFFFORM1													1	10	10	3	5	5	4	7	1	9	6,00	4,67	4,00	5,50	
56	115	FORM2HEIMHEIM																			0,00	0,00	0,00	0,00					
57	64	FORM2GASTAUSW													1		2		1		0,25	0,67	0,50	0,40					
58	56	FORM2HEIM													2	1	2				0,50	1,00	0,00	0,50					
59	56	FORM2GAST													2		3				0,50	1,00	0,00	0,50					
60	4	DIFFFORM2													3	10	10	1	4	7	7	5	5	8	6,00	6,00	5,00	6,00	
61	115	FORM3HEIMHEIM																			0,00	0,00	0,00	0,00					
62	64	FORM3GASTAUSW													1		1		2		0,25	0,33	1,00	0,40					
63	85	FORM3HEIM															2				0,00	0,67	0,00	0,20					
64	72	FORM3GAST													1		2				0,25	0,67	0,00	0,30					
65	1	DIFFFORM3													10	10	3	9	9	5	8	6	6	5,75	7,67	7,00	6,60		
66	85	FORM4HEIMHEIM																	1	1		0,00	0,33	0,50	0,20				
67	23	FORM4GASTAUSW													1	3	1	5	2		10		4	1,25	2,33	5,00	2,60		
68	115	FORM4HEIM																			0,00	0,00	0,00	0,00					
69	72	FORM4GAST													1		2				0,25	0,67	0,00	0,30					
70	2	DIFFFORM4													10	10	1	7	9	9	6	7	5	5,25	8,33	6,50	6,40		
71	115	FORM5HEIMHEIM																			0,00	0,00	0,00	0,00					
72	48	FORM5GASTAUSW													1		4		2			0,25	1,33	1,00	0,70				
73	56	FORM5HEIM															1	1	1	1	2		0,25	0,67	1,00	0,50			
74	51	FORM5GAST													1	1		4				0,50	1,33	0,00	0,60				
75	8	DIFFFORM5													1	8	10	1	7	8	6	2	4	2	5,00	7,00	3,00	4,90	
76	115	VIELETORE																			0,00	0,00	0,00	0,00					
77	64	AVG1PLATZHEIM													1		1		2			0,50	0,67	0,00	0,40				
78	22	AVG1PLATZGAST													1	7	5		4	7	3			3,25	4,67	0,00	2,70		
79	3	DIFFAVG1													4	7	8	4	10	7	7	6	4	6	5,75	8,00	5,00	6,30	
80	36	AVG2PLATZHEIM														4	3		7	1					1,75	2,67	0,00	1,50	
81	85	AVG2PLATZGAST																2					0,00	0,67	0,00	0,20			
82	15	DIFFAVG2													3	5	1	9	1	3	8	1	8		2,25	4,33	4,50	3,90	
83	39	AVG3PLATZHEIM													4		1	6		2					1,25	2,67	0,00	1,30	
84	43	AVG3PLATZGAST													1	2		2	2	1					0,75	1,67	0,00	0,80	
85	12	DIFFAVG3													1	5	9	1	6	3	9		6	2	4,00	6,00	3,00	4,20	
86	64	AVG4PLATZHEIM													1	2		1							1,00	0,00	0,00	0,40	
87	51	AVG4PLATZGAST														1	3	1			1					1,25	0,33	0,00	0,60
88	16	DIFFAVG4													1	2	7	2	7		8	3	4		3,00	5,00	3,50	3,40	
89	36	AVG5PLATZHEIM													4	3		7	1						1,75	2,67	0,00	1,50	
90	85	AVG5PLATZGAST																2					0,00	0,67	0,00	0,20			
91	18	DIFFAVG5													2	5	1	8	1	3	3	1	8		2,00	4,00	2,00	3,20	
92	115	HAEUFIGSTEPLAZIERUNGHEIM																					0,00	0,00	0,00	0,00			

93	115	HAEUFIGSTEPLAZIERUNGSGAST								0,00	0,00	0,00	0,00
94	72	ADDNATIONALTOP9HEIM		2	1					0,50	0,33	0,00	0,30
95	64	ADDNATIONALTOP9GAST	1	1	2					0,50	0,67	0,00	0,40
96	27	ADDNATIONALLAENDERSP.TOP9H.	1		9		7	1	5	0,25	3,00	4,00	2,30
97	115	ADDNATIONALLAENDERSP.TOP9G.								0,00	0,00	0,00	0,00
98	85	ADDNATIONALTOP9POSHEIM			1	1				0,25	0,33	0,00	0,20
99	56	ADDNATIONALTOP9POSGAST	2	1	2					0,75	0,67	0,00	0,50
100	72	ADDNAT.LAENDERSP.TOPPOS9H.	1		1		1			0,25	0,33	0,50	0,30
101	115	ADDNAT.LAENDERSP.TOPPOS9G.								0,00	0,00	0,00	0,00
102	93	ADDMITTELNATLSPHEIM			1					0,00	0,33	0,00	0,10
103	93	ADDMITTELNATLSPGAST	1							0,25	0,00	0,00	0,10
104	56	ADDLSPKTHEIM		1	1	3				0,25	1,33	0,00	0,50
105	72	ADDLSPKTGAST		1	2					0,25	0,67	0,00	0,30
106	35	ADDGESCHOSSENTOREHEIM		6		2	6		2	1,50	2,67	1,00	1,60
107	115	ADDGESCHOSSENTOREGAST								0,00	0,00	0,00	0,00
108	115	ADD DURCHGELASSEN TOREHEIM								0,00	0,00	0,00	0,00
109	115	ADD DURCHGELASSEN TOREGAST								0,00	0,00	0,00	0,00
110	41	ADDPOSITIONHEIM		1	1	1	3		1	0,75	2,00	0,50	1,00
111	115	ADDPOSITIONGAST								0,00	0,00	0,00	0,00
112	115	ADDSTADIONGROESSEHEIM								0,00	0,00	0,00	0,00
113	93	ADDSTADIONGROESSEGAST	1							0,25	0,00	0,00	0,10
114	115	ADD10ABSTIEGHEIM								0,00	0,00	0,00	0,00
115	93	ADD10ABSTIEGGAST				1				0,00	0,33	0,00	0,10
116	115	ADDBRAZILHEIM								0,00	0,00	0,00	0,00
117	93	ADDBRAZILGAST	1							0,25	0,00	0,00	0,10
118	115	ADDBRAZILNATHEIM								0,00	0,00	0,00	0,00
119	115	ADDBRAZILNATGAST								0,00	0,00	0,00	0,00
120	28	ADDGERMANYNATHEIM		4	9		6	1		3,25	2,33	0,50	2,10
121	72	ADDGERMANYNATGAST		1		2				0,25	0,67	0,00	0,30
122	115	ADDAUSTRIAHEIM								0,00	0,00	0,00	0,00
123	115	ADDAUSTRIAGAST								0,00	0,00	0,00	0,00
124	115	ADDAUSTRIANATHEIM								0,00	0,00	0,00	0,00
125	115	ADDAUSTRIANATGAST								0,00	0,00	0,00	0,00
126	31	ADDEWIGETABELLEHEIM		3	3	2	6	1	2	1,50	3,00	2,00	1,90
127	39	ADDEWIGETABELLEGAST		3	2	1	3		2	1,50	1,67	1,00	1,30
128	43	ADDUEFA1JAHRHEIM				8				2,00	0,00	0,00	0,80
129	56	ADDUEFA1JAHRGAST		1		3			1	0,25	1,00	0,50	0,50
130	72	ADDUEFA5JAHRHEIM		1	1		1			0,50	0,33	0,00	0,30
131	56	ADDUEFA5JAHREGAST		1	1	1	2			0,75	0,67	0,00	0,50
132	48	ADDPUNKTEHEIMHEIM		4	1		2			1,25	0,67	0,00	0,70
133	72	ADDPUNKTEGASTGAST		1	1	1				0,50	0,33	0,00	0,30
134	93	ABSTIEG1HEIM				1				0,00	0,33	0,00	0,10
135	8	ABSTIEG1GAST		10	5	1	7	7	2	7	6	4	4,00
136	64	DIFFABSTIEG1				2		1		1	0,00	1,00	0,00
137	93	ABSTIEG2HEIM				1					0,00	0,33	0,00
138	7	ABSTIEG2GAST		10	5	1	7	7	2	7	6	5	4,00
139	43	DIFFABSTIEG2				1	3	1	2	1	0,25	1,33	1,00
													0,80

140	93	ABSTIEG3HEIM		1						0,00	0,33	0,00	0,10				
141	11	ABSTIEG3GAST	10	5	1	6	8	2	4	6	3	4,00	5,33	5,00	4,50		
142	34	DIFFABSTIEG3	2	1	3		3	1	2	3		2	1,50	2,00	1,50	1,70	
143	93	ABSTIEG4HEIM				1						0,00	0,33	0,00	0,10		
144	14	ABSTIEG4GAST	7	4	1	5	7	2	5	6	3	3,00	4,67	5,50	4,00		
145	72	DIFFABSTIEG4				3						0,00	1,00	0,00	0,30		
146	93	ABSTIEG5HEIM				1						0,00	0,33	0,00	0,10		
147	28	ABSTIEG5GAST	4	1	1	5	5		3		2	1,50	3,33	1,50	2,10		
148	85	DIFFABSTIEG5	1			1						0,25	0,33	0,00	0,20		
149	4	DIFFPUNKTEHEIMGAST	5	8	8	5	10	7	7	7	3	5,25	7,33	7,00	6,00		
150	19	DIFFPKTLS	1	6	8	1	3	2	7			4,00	4,00	0,00	2,80		
151	43	DIFFGSTLS		4		3			1			1,00	1,00	0,50	0,80		
152	32	DIFFKASSTLS	2	3	6		4	2	1			2,75	2,33	0,00	1,80		
153	19	DIFFPOSLS	2	7	7	3	1	7		1		4,00	3,67	0,50	2,80		
154	93	DIFFEWIGETABELLE				1						0,00	0,33	0,00	0,10		
155	16	DIFFNAT	5	4	8	1	3	4	4	3	1	1	4,50	3,67	2,00	3,40	
156	23	DIFNAT2	2		6	1	3		7	1	3	3	2,25	3,33	2,00	2,60	
157	13	DIFFNATPOS	2	5	8	1	4	5	8	3	2	3	4,00	5,67	2,50	4,10	
158	38	DIFNATPOS2		2	1	3			7	1			0,75	3,33	0,50	1,40	
159	23	DIFFNATPOSMIT		3	5	1	6	1	2	6	2		2,25	3,00	4,00	2,60	
160	115	DIFFSTADGR											0,00	0,00	0,00	0,00	
161	115	DIFF10ABSTIEG											0,00	0,00	0,00	0,00	
162	115	DIFFBRAZIL											0,00	0,00	0,00	0,00	
163	93	DIFBRAZILN			1								0,25	0,00	0,00	0,10	
164	28	DIFFGERMANYN		1	2	6		3	1	1	2	2	3	2,25	1,67	2,00	2,10
165	115	DIFFAUSTRIA											0,00	0,00	0,00	0,00	
166	115	DIFFAUSTRIAN											0,00	0,00	0,00	0,00	
167	26	DIFFUEFA1J		6	2	6	8		1	1	1		3,50	3,00	1,00	2,50	
168	8	DIFFUEFA5J		1	4	9	2	8	6	6	10		3	4,00	6,67	5,00	4,90
169	51	DIFFHAEUFPLAZ		1	1	1		2		1			0,75	1,00	0,00	0,60	

Wrapper mit Klassifizierer: „SVM“ und Suchmethode: „LinearForwardSelection“

NR	Platz	Name	24	35	46	57	25	36	47	26	37	27	3SD	4SD	5SD	Ges
1	95	HEIM											0,00	0,00	0,00	0,00
2	12	GAST	1				1	2		3	3	3	0,25	1,00	3,00	1,30
3	58	POSHEIM					1			1	1		0,25	0,33	0,50	0,30
4	52	PLUSMINUSHEIM					1			1		2	0,25	0,33	0,00	0,40
5	95	SIEGEHEIM											0,00	0,00	0,00	0,00
6	95	REMISHEIM											0,00	0,00	0,00	0,00
7	76	NIEDERLAGENHEIM								1			0,00	0,00	0,50	0,10
8	95	PUNKTEHEIM											0,00	0,00	0,00	0,00
9	65	TDHEIM							1			1	0,00	0,33	0,00	0,20
10	76	POSGAST							1				0,00	0,33	0,00	0,10
11	76	PLUSMINUSGAST								1			0,00	0,00	0,50	0,10
12	76	SIEGEGAST					1						0,25	0,00	0,00	0,10
13	95	REMISGAST											0,00	0,00	0,00	0,00
14	95	NIEDERLAGENGAST											0,00	0,00	0,00	0,00
15	95	PUNKTEGAST											0,00	0,00	0,00	0,00
16	44	TDGAST					1	2		1		1	0,75	0,33	0,50	0,50
17	95	MONAT											0,00	0,00	0,00	0,00
18	95	JAHRESZEIT											0,00	0,00	0,00	0,00
19	76	TABDIFF						1					0,25	0,00	0,00	0,10
20	17	TABTENDENZ		3	4			3	1		1		1,75	1,33	0,50	1,20
21	95	TABDIFFFORM											0,00	0,00	0,00	0,00
22	95	TABDIFFFORMVERG											0,00	0,00	0,00	0,00
23	95	SIEGDIFF											0,00	0,00	0,00	0,00
24	95	REMISDIFF											0,00	0,00	0,00	0,00
25	95	NIEDDIFF											0,00	0,00	0,00	0,00
26	95	PUNKTEDIFF											0,00	0,00	0,00	0,00
27	31	TDDIFF					2			5		1	0,50	1,67	0,50	0,80
28	95	TORDIFFFORM											0,00	0,00	0,00	0,00
29	95	TFHEIM											0,00	0,00	0,00	0,00
30	58	TFGAST					1			1		1	0,25	0,33	0,50	0,30
31	95	HEIMSTARK											0,00	0,00	0,00	0,00
32	95	GASTSTARK											0,00	0,00	0,00	0,00
33	95	SIEGEHEIM											0,00	0,00	0,00	0,00
34	95	VERSUS1HEIM											0,00	0,00	0,00	0,00
35	95	VERSUS1GAST											0,00	0,00	0,00	0,00
36	95	VERSUS1											0,00	0,00	0,00	0,00
37	95	VERSUS2HEIM											0,00	0,00	0,00	0,00
38	95	VERSUS2GAST											0,00	0,00	0,00	0,00
39	76	VERSUS2								1			0,00	0,33	0,00	0,10
40	95	VERSUS3HEIM											0,00	0,00	0,00	0,00
41	95	VERSUS3GAST											0,00	0,00	0,00	0,00
42	95	VERSUS3											0,00	0,00	0,00	0,00
43	95	VERSUS4HEIM											0,00	0,00	0,00	0,00
44	95	VERSUS4GAST											0,00	0,00	0,00	0,00
45	95	VERSUS4											0,00	0,00	0,00	0,00

46	95	VERSUS5HEIM								0,00	0,00	0,00	0,00
47	95	VERSUS5GAST								0,00	0,00	0,00	0,00
48	95	VERSUS5								0,00	0,00	0,00	0,00
49	95	TABGAST								0,00	0,00	0,00	0,00
50	95	TABVERGLEICH								0,00	0,00	0,00	0,00
51	95	FORM1HEIMHEIM								0,00	0,00	0,00	0,00
52	76	FORM1GASTAUSW		1						0,25	0,00	0,00	0,10
53	95	FORM1HEIM		1						0,00	0,00	0,00	0,00
54	76	FORM1GAST		3	4					0,25	0,00	0,00	0,10
55	34	DIFFFORM1		1	2					1,75	0,00	0,00	0,70
56	95	FORM2HEIMHEIM								0,00	0,00	0,00	0,00
57	58	FORM2GASTAUSW								0,75	0,00	0,00	0,30
58	95	FORM2HEIM								0,00	0,00	0,00	0,00
59	76	FORM2GAST		1						0,25	0,00	0,00	0,10
60	44	DIFFFORM2		2		1	1	1		0,50	0,00	1,00	0,50
61	95	FORM3HEIMHEIM								0,00	0,00	0,00	0,00
62	95	FORM3GASTAUSW								0,00	0,00	0,00	0,00
63	95	FORM3HEIM								0,00	0,00	0,00	0,00
64	95	FORM3GAST								0,00	0,00	0,00	0,00
65	76	DIFFFORM3					1			0,00	0,00	0,50	0,10
66	95	FORM4HEIMHEIM								0,00	0,00	0,00	0,00
67	31	FORM4GASTAUSW		2	3		1	1	1	1,25	0,33	0,50	0,80
68	95	FORM4HEIM								0,00	0,00	0,00	0,00
69	95	FORM4GAST								0,00	0,00	0,00	0,00
70	31	DIFFFORM4		3	2			3		0,75	0,67	1,50	0,80
71	95	FORM5HEIMHEIM								0,00	0,00	0,00	0,00
72	76	FORM5GASTAUSW						1		0,00	0,00	0,50	0,10
73	95	FORM5HEIM								0,00	0,00	0,00	0,00
74	95	FORM5GAST								0,00	0,00	0,00	0,00
75	37	DIFFFORM5		1	1	3		1		1,25	0,33	0,00	0,60
76	95	VIELETORE								0,00	0,00	0,00	0,00
77	95	AVG1PLATZHEIM								0,00	0,00	0,00	0,00
78	65	AVG1PLATZGAST			2					0,50	0,00	0,00	0,20
79	2	DIFFAVG1		4	10	4		3	9	4	5	4,50	4,00
80	95	AVG2PLATZHEIM								0,00	0,00	0,00	0,00
81	95	AVG2PLATZGAST								0,00	0,00	0,00	0,00
82	58	DIFFAVG2		1			1		1			0,25	0,67
83	95	AVG3PLATZHEIM								0,00	0,00	0,00	0,00
84	27	AVG3PLATZGAST		4	1	4				1,25	1,33	0,00	0,90
85	10	DIFFAVG3		1	2	2		1	2	1	5	1	1,25
86	34	AVG4PLATZHEIM		3	1	3						1,00	1,00
87	65	AVG4PLATZGAST		1		1						0,25	0,33
88	7	DIFFAVG4		2	2	1	5	2	3	2	1	2,50	1,67
89	58	AVG5PLATZHEIM						2		1		0,00	0,67
90	8	AVG5PLATZGAST		2	3	1	2	3	2	3	1	1,50	2,33
91	17	DIFFAVG5		4	1	4	1	1			1	1,25	2,00
92	76	HAEUFIGSTEPLAZIERUNGHEIM							1			0,00	0,33

93	20	HAEUFIGSTEPLAZIERUNGGAST	2	2	2	2	2	1	1,00	1,33	1,50	1,10			
94	37	ADDNATIONALTOP9HEIM			1	1	3	1	0,00	0,67	1,50	0,60			
95	95	ADDNATIONALTOP9GAST							0,00	0,00	0,00	0,00			
96	27	ADDNATIONALLAENDERSP.TOP9H.			1	1	5	1	1	0,00	0,67	3,00	0,90		
97	37	ADDNATIONALLAENDERSP.TOP9G.	1	3				1	1	1,00	0,00	1,00	0,60		
98	44	ADDNATIONALTOP9POSHEIM	1	3		1				1,00	0,33	0,00	0,50		
99	52	ADDNATIONALTOP9POSGAST	1	1		1			1	0,50	0,33	0,50	0,40		
100	76	ADDNAT.LAENDERSP.TOPPOS9H.						1		0,00	0,00	0,50	0,10		
101	44	ADDNAT.LAENDERSP.TOPPOS9G.			1		1	2	1	0,25	0,33	1,00	0,50		
102	58	ADDMITTELNATLSPHEIM		1			1	1		0,25	0,33	0,50	0,30		
103	17	ADDMITTELNATLSPGAST	3	1	1	3	1	2	1	1,25	2,00	0,50	1,20		
104	52	ADDLSPKTHEIM		2				1	1	0,50	0,00	0,50	0,40		
105	44	ADDLSPKTGAST	2	1		2				0,75	0,67	0,00	0,50		
106	52	ADDGESCHOSSENTOREHEIM		2	1				1	0,75	0,00	0,50	0,40		
107	34	ADDGESCHOSSENTOREGAST	2	1		2		2		0,75	0,67	1,00	0,70		
108	20	ADDDURCHGELASSENTOREHEIM	3	1	1	2	3		1	1,75	1,00	0,50	1,10		
109	11	ADDDURCHGELASSENTOREGAST	3		1	3	3	1	2	1	1,00	2,33	1,00	1,40	
110	65	ADDPOSITIONHEIM		1		1				0,25	0,33	0,00	0,20		
111	12	ADDPOSITIONGAST	2	1	1	2	2	3	1	1	1,50	1,67	1,00	1,30	
112	12	ADDSTADIONGROESSEHEIM	4	1		4	2		1	1	1,25	2,00	1,00	1,30	
113	8	ADDSTADIONGROESSEGAST	5	1	1	5	1		3	1	1,75	2,00	1,50	1,70	
114	20	ADD10ABSTIEGHEIM	3	1	1	3	2		1		1,25	1,67	0,50	1,10	
115	4	ADD10ABSTIEGGAST	4	3		1	6	4	3	1	3	2,00	4,33	2,00	2,80
116	6	ADDBRAZILHEIM	3	4		3	1	3	1	2	2	1,75	2,33	1,50	1,90
117	44	ADDBRAZILGAST	1	1		1	1		1			0,75	0,33	0,50	0,50
118	20	ADDBRAZILNATHEIM	1			1	1	1	5	1	1	0,25	1,00	3,00	1,10
119	76	ADDBRAZILNATGAST					1					0,00	0,33	0,00	0,10
120	52	ADDGERMANYNATHEIM	1	1		1		1				0,50	0,67	0,00	0,40
121	65	ADDGERMANYNATGAST				1		1				0,25	0,33	0,00	0,20
122	37	ADDAUSTRIAHEIM	1			1	1	1	1	1		0,25	1,00	0,50	0,60
123	12	ADDAUSTRIAGAST	3	1	1	3		1	2	2		1,25	1,33	2,00	1,30
124	44	ADDAUSTRIANATHEIM	1			1		2		1		0,25	0,33	1,00	0,50
125	12	ADDAUSTRIANATGAST	1	2	2	1	1	1	4	1		1,25	1,00	2,00	1,30
126	27	ADDEWIGETABELLEHEIM	3	1	1	3		1				1,25	1,33	0,00	0,90
127	37	ADDEWIGETABELLEGAST		1	1				2	2		0,50	0,00	1,00	0,60
128	58	ADDUEFA1JAHRHEIM			1			1	1			0,25	0,33	0,50	0,30
129	95	ADDUEFA1JAHRGAST										0,00	0,00	0,00	0,00
130	76	ADDUEFA5JAHRHEIM							1			0,00	0,00	0,50	0,10
131	76	ADDUEFA5JAHREGAST								1		0,00	0,00	0,50	0,10
132	95	ADDPUNKTEHEIMHEIM										0,00	0,00	0,00	0,00
133	95	ADDPUNKTEGASTGAST										0,00	0,00	0,00	0,00
134	65	ABSTIEG1HEIM			2							0,50	0,00	0,00	0,20
135	95	ABSTIEG1GAST										0,00	0,00	0,00	0,00
136	95	DIFFABSTIEG1										0,00	0,00	0,00	0,00
137	95	ABSTIEG2HEIM										0,00	0,00	0,00	0,00
138	95	ABSTIEG2GAST										0,00	0,00	0,00	0,00
139	95	DIFFABSTIEG2										0,00	0,00	0,00	0,00

140	95	ABSTIEG3HEIM										0,00	0,00	0,00	0,00				
141	95	ABSTIEG3GAST										0,00	0,00	0,00	0,00				
142	95	DIFFABSTIEG3										0,00	0,00	0,00	0,00				
143	95	ABSTIEG4HEIM										0,00	0,00	0,00	0,00				
144	95	ABSTIEG4GAST										0,00	0,00	0,00	0,00				
145	95	DIFFABSTIEG4										0,00	0,00	0,00	0,00				
146	95	ABSTIEG5HEIM										0,00	0,00	0,00	0,00				
147	95	ABSTIEG5GAST										0,00	0,00	0,00	0,00				
148	95	DIFFABSTIEG5										0,00	0,00	0,00	0,00				
149	37	DIFFPUNKTEHEIMGAST				4		1		1		1,00	0,33	0,50	0,60				
150	65	DIFFPKTLS						1		1		0,00	0,33	0,50	0,20				
151	65	DIFFGSTLS				1	1					0,50	0,00	0,00	0,20				
152	37	DIFFKASSTLS				3		3				0,75	1,00	0,00	0,60				
153	27	DIFFPOSLS				2	2	2	2		1	1,50	0,67	0,50	0,90				
154	65	DIFFEWIGETABELLE				1		1				0,25	0,33	0,00	0,20				
155	25	DIFFNAT				1	1	3	2	1	1	1,75	0,67	0,50	1,00				
156	65	DIFNAT2				1	1					0,50	0,00	0,00	0,20				
157	1	DIFFNATPOS				7	6	2	1	7	7	2	6	7	7	4,00	5,33	6,50	5,20
158	52	DIFNATPOS2				1	1		1			1		0,50	0,33	0,50	0,40		
159	20	DIFFNATPOSMIT				2	2	1		1		3	2	1,25	0,33	1,50	1,10		
160	95	DIFFSTADGR												0,00	0,00	0,00	0,00		
161	95	DIFF10ABSTIEG												0,00	0,00	0,00	0,00		
162	76	DIFFBRAZIL						1						0,25	0,00	0,00	0,10		
163	76	DIFBRAZILN								1				0,00	0,33	0,00	0,10		
164	3	DIFFGERMANYN				4	4	2	2	4	6	1	1	2	4	3,00	3,67	1,50	3,00
165	65	DIFFAUSTRIA										2			0,00	0,67	0,00	0,20	
166	76	DIFFAUSTRIAN										1			0,00	0,33	0,00	0,10	
167	5	DIFFUEFA1J				3	1	2		5	1	2	4	2	1,50	2,00	3,00	2,00	
168	25	DIFFUEFA5J				1	1		1	1	2	1		3	0,50	1,33	0,50	1,00	
169	44	DIFFHAEUFPLAZ						1	2			1	1		0,75	0,00	1,00	0,50	

Wrapper mit Klassifizierer: „SVM“ und Suchmethode: „BestFirst-Backward Elimination“

NR	Platz	Name	24	35	46	57	25	36	47	26	37	27	3SD	4SD	5SD	Ges
1	167	HEIM	9	9	7	10	9	8	4	9	2	3	8,75	7,00	5,50	7,00
2	168	GAST	7	9	5	10	8	1	9	8	4	5	7,75	6,00	6,00	6,60
3	70	POSHEIM	9	8	10	8	9	10	10	10	9	10	8,75	9,67	9,50	9,30
4	124	PLUSMINUSHEIM	7	10	10	10	9	8	8	8	10	10	9,25	8,33	9,00	9,00
5	32	SIEGEHEIM	10	10	9	9	10	10	10	10	9	8	9,50	10,00	9,50	9,50
6	94	REMISHEIM	10	9	9	10	10	10	7	10	9	8	9,50	9,00	9,50	9,20
7	9	NIEDERLAGENHEIM	10	9	9	10	10	10	10	10	10	9	9,50	10,00	10,00	9,70
8	49	PUNKTEHEIM	10	10	10	10	9	10	8	9	8	10	10,00	9,00	8,50	9,40
9	108	TDHEIM	10	8	10	9	10	10	9	9	8	8	9,25	9,67	8,50	9,10
10	94	POSGAST	10	7	8	10	9	9	9	10	10	10	8,75	9,00	10,00	9,20
11	94	PLUSMINUSGAST	9	10	10	8	8	10	9	10	10	8	9,25	9,00	10,00	9,20
12	70	SIEGEGAST	9	9	10	10	10	8	9	10	9	9	9,50	9,00	9,50	9,30
13	18	REMISGAST	10	9	9	10	10	10	9	10	10	9	9,50	9,67	10,00	9,60
14	94	NIEDERLAGENGAST	8	9	10	10	9	8	10	10	9	9	9,25	9,00	9,50	9,20
15	9	PUNKTEGAST	10	10	10	10	10	10	9	9	10	9	10,00	9,67	9,50	9,70
16	18	TDGAST	8	10	9	10	10	9	10	10	10	10	9,25	9,67	10,00	9,60
17	165	MONAT	10	7	8	9	9	8	6	5	8	6	8,50	7,67	6,50	7,60
18	164	JAHRESZEIT	8	9	9	8	9	7	6	9	7	8	8,50	7,33	8,00	8,00
19	132	TABDIFF	9	6	8	10	9	10	10	10	8	9	8,25	9,67	9,00	8,90
20	132	TABTENDENZ	9	9	9	9	8	9	8	9	9	10	9,00	8,33	9,00	8,90
21	152	TABDIFFFORM	8	7	9	9	10	8	9	8	10	9	8,25	9,00	9,00	8,70
22	152	TABDIFFFORMVERG	8	9	9	7	8	10	9	9	10	8	8,25	9,00	9,50	8,70
23	32	SIEGDIFF	10	10	10	9	8	9	10	10	9	10	9,75	9,00	9,50	9,50
24	132	REMISDIFF	10	10	10	8	9	10	7	8	8	9	9,50	8,67	8,00	8,90
25	94	NIEDDIFF	10	8	10	8	9	10	9	10	9	9	9,00	9,33	9,50	9,20
26	32	PUNKTEDIFF	10	9	9	10	10	10	8	10	10	9	9,50	9,33	10,00	9,50
27	32	TDDIFF	9	10	10	10	9	10	8	10	9	10	9,75	9,00	9,50	9,50
28	132	TORDIFFFORM	9	9	9	9	8	8	10	9	9	9	9,00	8,67	9,00	8,90
29	49	TFHEIM	7	10	10	10	10	10	8	10	10	9	9,25	9,33	10,00	9,40
30	156	TFGAST	8	10	9	6	9	10	8	9	6	10	8,25	9,00	7,50	8,50
31	32	HEIMSTARK	10	10	10	10	9	9	10	9	9	9	10,00	9,33	9,00	9,50
32	124	GASTSTARK	10	8	10	8	7	9	10	10	8	10	9,00	8,67	9,00	9,00
33	165	SIEGEHEIM	7	8	7	6	10	7	8	9	8	6	7,00	8,33	8,50	7,60
34	132	VERSUS1HEIM	9	9	6	9	9	9	8	10	10	10	8,25	8,67	10,00	8,90
35	70	VERSUS1GAST	9	10	9	9	9	10	9	9	10	9	9,25	9,33	9,50	9,30
36	49	VERSUS1	10	9	10	9	10	9	9	9	9	10	9,50	9,33	9,00	9,40
37	32	VERSUS2HEIM	10	9	10	10	10	9	10	7	10	10	9,75	9,67	8,50	9,50
38	49	VERSUS2GAST	10	9	9	10	9	10	9	10	9	9	9,50	9,33	9,50	9,40
39	148	VERSUS2	8	9	8	8	9	10	9	9	8	10	8,25	9,33	8,50	8,80
40	70	VERSUS3HEIM	9	10	9	10	10	10	7	9	9	10	9,50	9,00	9,00	9,30
41	32	VERSUS3GAST	9	10	10	9	10	9	10	9	9	10	9,50	9,67	9,00	9,50
42	148	VERSUS3	7	10	10	9	9	9	8	8	8	10	9,00	8,67	8,00	8,80
43	32	VERSUS4HEIM	10	10	10	9	10	8	10	8	10	10	9,75	9,33	9,00	9,50
44	70	VERSUS4GAST	9	10	7	10	9	10	10	9	9	10	9,00	9,67	9,00	9,30
45	94	VERSUS4	10	9	8	9	8	10	9	10	9	10	9,00	9,00	9,50	9,20

46	124	VERSUS5HEIM	9	9	9	9	9	9	9	9	8	10	9,00	9,00	8,50	9,00
47	108	VERSUS5GAST	10	9	10	7	10	8	9	9	10	9	9,00	9,00	9,50	9,10
48	159	VERSUS5	9	8	8	7	9	10	6	8	10	9	8,00	8,33	9,00	8,40
49	156	TABGAST	8	9	8	8	6	10	7	10	9	10	8,25	7,67	9,50	8,50
50	160	TABVERGLEICH	10	9	8	8	7	8	9	7	8	9	8,75	8,00	7,50	8,30
51	70	FORM1HEIMHEIM	9	10	10	8	9	10	10	8	10	9	9,25	9,67	9,00	9,30
52	70	FORM1GASTAUSW	10	9	9	10	10	9	10	8	8	10	9,50	9,67	8,00	9,30
53	70	FORM1HEIM	9	9	10	8	10	9	9	10	9	10	9,00	9,33	9,50	9,30
54	94	FORM1GAST	9	9	9	9	9	10	8	9	10	10	9,00	9,00	9,50	9,20
55	154	DIFFFORM1	10	7	10	8	8	8	8	9	10	8	8,75	8,00	9,50	8,60
56	9	FORM2HEIMHEIM	10	10	9	10	10	9	10	10	9	10	9,75	9,67	9,50	9,70
57	132	FORM2GASTAUSW	9	8	10	10	10	8	8	9	8	9	9,25	8,67	8,50	8,90
58	9	FORM2HEIM	10	10	8	10	10	10	10	9	10	10	9,50	10,00	9,50	9,70
59	49	FORM2GAST	8	10	10	10	10	9	9	9	10	9	9,50	9,33	9,50	9,40
60	108	DIFFFORM2	10	10	9	9	8	10	8	8	9	10	9,50	8,67	8,50	9,10
61	132	FORM3HEIMHEIM	7	10	10	10	8	8	10	10	8	8	9,25	8,67	9,00	8,90
62	32	FORM3GASTAUSW	9	10	8	9	10	10	10	10	9	10	9,00	10,00	9,50	9,50
63	94	FORM3HEIM	9	10	10	10	10	9	6	10	10	8	9,75	8,33	10,00	9,20
64	132	FORM3GAST	9	9	9	7	9	10	7	9	10	10	8,50	8,67	9,50	8,90
65	132	DIFFFORM3	8	8	10	7	9	10	9	9	9	10	8,25	9,33	9,00	8,90
66	3	FORM4HEIMHEIM	9	10	10	10	10	10	9	10	10	10	9,75	9,67	10,00	9,80
67	9	FORM4GASTAUSW	10	10	9	10	9	10	9	10	10	10	9,75	9,33	10,00	9,70
68	3	FORM4HEIM	10	10	10	10	10	10	9	10	10	9	10,00	9,67	10,00	9,80
69	108	FORM4GAST	7	10	9	10	8	10	9	10	9	9	9,00	9,00	9,50	9,10
70	70	DIFFFORM4	9	10	10	9	10	10	8	8	9	10	9,50	9,33	8,50	9,30
71	70	FORM5HEIMHEIM	10	8	10	8	10	9	8	10	10	10	9,00	9,00	10,00	9,30
72	49	FORM5GASTAUSW	10	8	10	10	10	10	9	9	9	9	9,50	9,67	9,00	9,40
73	49	FORM5HEIM	8	10	9	10	9	10	9	10	9	10	9,25	9,33	9,50	9,40
74	94	FORM5GAST	9	9	10	8	9	10	10	9	9	9	9,00	9,67	9,00	9,20
75	162	DIFFFORM5	10	6	8	10	8	10	6	8	7	9	8,50	8,00	7,50	8,20
76	169	VIELETORE	0	0	0	0	0	0	0	0	0	0	0,00	0,00	0,00	0,00
77	124	AVG1PLATZHEIM	10	9	9	10	9	9	9	7	8	10	9,50	9,00	7,50	9,00
78	3	AVG1PLATZGAST	10	10	10	10	9	10	10	10	9	10	10,00	9,67	9,50	9,80
79	108	DIFFAVG1	9	8	9	10	9	8	10	9	10	9	9,00	9,00	9,50	9,10
80	124	AVG2PLATZHEIM	9	9	10	9	9	9	9	8	9	9	9,25	9,00	8,50	9,00
81	3	AVG2PLATZGAST	10	10	9	10	10	10	10	10	9	10	9,75	10,00	9,50	9,80
82	124	DIFFAVG2	10	9	9	9	9	10	10	8	8	8	9,25	9,67	8,00	9,00
83	70	AVG3PLATZHEIM	10	7	10	9	9	8	10	10	10	10	9,00	9,00	10,00	9,30
84	18	AVG3PLATZGAST	9	10	9	9	9	10	10	10	10	10	9,25	9,67	10,00	9,60
85	124	DIFFAVG3	9	10	8	8	9	10	8	9	9	10	8,75	9,00	9,00	9,00
86	70	AVG4PLATZHEIM	9	9	9	10	9	10	8	10	10	9	9,25	9,00	10,00	9,30
87	49	AVG4PLATZGAST	10	9	10	10	10	10	10	9	9	7	9,75	10,00	9,00	9,40
88	49	DIFFAVG4	10	10	8	9	9	9	10	9	10	10	9,25	9,33	9,50	9,40
89	94	AVG5PLATZHEIM	9	10	8	10	10	9	8	9	10	9	9,25	9,00	9,50	9,20
90	3	AVG5PLATZGAST	10	10	10	10	10	9	10	10	10	9	10,00	9,67	10,00	9,80
91	108	DIFFAVG5	10	9	10	9	8	10	9	10	9	7	9,50	9,00	9,50	9,10
92	18	HAEUFIGSTEPLAZIERUNGHEIM	9	9	10	10	10	9	10	10	9	10	9,50	9,67	9,50	9,60

93	49	HAEUFIGSTEPLAZIERUNGSGAST	9	8	10	10	9	9	10	9	10	10	9,25	9,33	9,50	9,40
94	18	ADDNATIONALTOP9HEIM	9	9	10	9	9	10	10	10	10	10	9,25	9,67	10,00	9,60
95	32	ADDNATIONALTOP9GAST	9	8	10	10	9	10	10	10	10	10	9,25	9,67	10,00	9,50
96	70	ADDNATIONALLAENDERSP.TOP9H.	10	9	8	10	9	9	10	9	10	9	9,25	9,33	9,50	9,30
97	32	ADDNATIONALLAENDERSP.TOP9G.	9	9	9	10	10	9	10	10	10	10	9,25	9,67	10,00	9,50
98	70	ADDNATIONALTOP9POSHEIM	9	10	9	9	10	10	9	10	8	9	9,25	9,67	9,00	9,30
99	49	ADDNATIONALTOP9POSGAST	8	9	10	9	10	9	9	10	10	10	9,00	9,33	10,00	9,40
100	70	ADDNAT.LAENDERSP.TOPPOS9H.	9	10	9	9	9	9	9	10	9	10	9,25	9,00	9,50	9,30
101	32	ADDNAT.LAENDERSP.TOPPOS9G.	10	10	10	10	7	10	10	10	10	8	10,00	9,00	10,00	9,50
102	70	ADDMITTELNATLSPHEIM	10	10	10	10	7	10	10	10	10	8	10,00	9,00	9,00	9,30
103	108	ADDMITTELNATLSPGAST	9	9	8	9	10	9	10	8	9	10	8,75	9,67	8,50	9,10
104	18	ADDLSPKTHEIM	10	10	10	9	10	8	9	10	10	10	9,75	9,00	10,00	9,60
105	132	ADDLSPKTGAST	8	8	10	9	9	9	9	8	10	9	8,75	9,00	9,00	8,90
106	9	ADDGESCHOSSENETOREHEIM	10	9	9	10	10	10	9	10	10	10	9,50	9,67	10,00	9,70
107	108	ADDGESCHOSSENETOREGAST	8	10	10	8	10	9	9	8	10	9	9,00	9,33	9,00	9,10
108	9	ADDDURCHGELASSENETOREHEIM	9	10	9	10	10	10	9	10	10	10	9,50	9,67	10,00	9,70
109	18	ADDDURCHGELASSENETOREGAST	10	10	9	10	10	10	8	9	10	10	9,75	9,33	9,50	9,60
110	9	ADDPOSITIONHEIM	10	10	10	9	10	10	9	10	10	9	9,75	9,67	10,00	9,70
111	132	ADDPOSITIONGAST	9	8	10	8	7	10	10	9	9	9	8,75	9,00	9,00	8,90
112	49	ADDSTADIONGROESSEHEIM	8	9	10	10	10	10	10	8	10	9	9,25	10,00	9,00	9,40
113	132	ADDSTADIONGROESSEGAST	9	9	8	8	10	9	8	8	10	10	8,50	9,00	9,00	8,90
114	49	ADD10ABSTIEGHEIM	10	9	10	10	9	8	10	9	9	10	9,75	9,00	9,00	9,40
115	18	ADD10ABSTIEGGAST	9	10	9	9	10	10	10	9	10	10	9,25	10,00	9,50	9,60
116	32	ADDBRAZILHEIM	10	9	9	10	10	8	10	10	10	9	9,50	9,33	10,00	9,50
117	18	ADDBRAZILGAST	10	9	9	10	10	9	10	9	10	10	9,50	9,67	9,50	9,60
118	70	ADDBRAZILNATHEIM	10	10	8	10	10	8	10	9	9	9	9,50	9,33	9,00	9,30
119	70	ADDBRAZILNATGAST	10	9	10	9	9	7	10	10	10	9	9,50	8,67	10,00	9,30
120	49	ADDGERMANYNATHEIM	10	8	8	9	10	9	10	10	10	10	8,75	9,67	10,00	9,40
121	94	ADDGERMANYNATGAST	10	9	8	10	10	10	8	9	9	9	9,25	9,33	9,00	9,20
122	32	ADDAUSTRIAHEIM	10	10	10	10	10	9	10	8	8	10	10,00	9,67	8,00	9,50
123	49	ADDAUSTRIAGAST	9	9	9	10	10	9	9	10	10	9	9,25	9,33	10,00	9,40
124	9	ADDAUSTRIANATHEIM	10	10	9	10	9	10	10	10	9	10	9,75	9,67	9,50	9,70
125	18	ADDAUSTRIANATGAST	10	10	9	10	9	8	10	10	10	10	9,75	9,00	10,00	9,60
126	18	ADDEWIGETABELLEHEIM	10	10	9	10	10	9	10	10	10	8	9,75	9,67	10,00	9,60
127	132	ADDEWIGETABELLEGAST	9	8	9	9	8	10	10	9	8	9	8,75	9,33	8,50	8,90
128	94	ADDUEFA1JAHRHEIM	10	9	10	8	9	9	8	9	10	10	9,25	8,67	9,50	9,20
129	70	ADDUEFA1JAHRGAST	10	10	9	9	10	7	9	10	9	10	9,50	8,67	9,50	9,30
130	1	ADDUEFA5JAHREHEIM	10	10	10	10	9	10	10	10	10	10	10,00	9,67	10,00	9,90
131	70	ADDUEFA5JAHREGAST	8	10	10	9	10	9	9	9	9	10	9,25	9,33	9,00	9,30
132	3	ADDPUNKTEHEIMHEIM	10	10	10	10	9	10	10	10	10	9	10,00	9,67	10,00	9,80
133	49	ADDPUNKTEGASTGAST	10	9	10	10	9	10	8	10	10	8	9,75	9,00	10,00	9,40
134	70	ABSTIEG1HEIM	10	10	8	9	9	8	10	10	10	9	9,25	9,00	10,00	9,30
135	70	ABSTIEG1GAST	10	8	8	10	8	10	10	10	9	10	9,00	9,33	9,50	9,30
136	108	DIFFABSTIEG1	8	10	9	8	10	10	7	10	9	10	8,75	9,00	9,50	9,10
137	1	ABSTIEG2HEIM	10	10	10	9	10	10	10	10	10	10	9,75	10,00	10,00	9,90
138	148	ABSTIEG2GAST	9	6	9	7	9	10	9	10	10	9	7,75	9,33	10,00	8,80
139	94	DIFFABSTIEG2	9	8	10	8	10	10	8	10	10	9	8,75	9,33	10,00	9,20

140	108	ABSTIEG3HEIM
141	162	ABSTIEG3GAST
142	160	DIFFABSTIEG3
143	154	ABSTIEG4HEIM
144	108	ABSTIEG4GAST
145	156	DIFFABSTIEG4
146	132	ABSTIEG5HEIM
147	148	ABSTIEG5GAST
148	108	DIFFABSTIEG5
149	32	DIFFPUNKTEHEIMGAST
150	32	DIFFPKTLS
151	132	DIFFGSTLS
152	108	DIFFKASSTLS
153	49	DIFFPOSLS
154	49	DIFFEWIGETABELLE
155	108	DIFFNAT
156	124	DIFNAT2
157	70	DIFFNATPOS
158	108	DIFNATPOS2
159	132	DIFFNATPOSMIT
160	18	DIFFSTADGR
161	49	DIFF10ABSTIEG
162	108	DIFFBRAZIL
163	18	DIFBRAZILN
164	18	DIFFGERMANYN
165	49	DIFFAUSTRIA
166	32	DIFFAUSTRIAN
167	70	DIFFUEFA1J
168	94	DIFFUEFA5J
169	49	DIFFHAEUFPLAZ

10	6	9	9	9	10	10	9	10	9	8,50	9,67	9,50	9,10
7	8	8	8	10	8	8	9	8	8	7,75	8,67	8,50	8,20
10	9	8	8	9	9	8	5	8	9	8,75	8,67	6,50	8,30
10	7	8	8	9	9	9	9	10	7	8,25	9,00	9,50	8,60
10	10	10	8	9	7	9	9	10	9	9,50	8,33	9,50	9,10
8	8	10	9	7	9	8	9	7	10	8,75	8,00	8,00	8,50
9	8	9	9	9	9	8	10	9	9	8,75	8,67	9,50	8,90
9	8	7	8	9	10	8	9	10	10	8,00	9,00	9,50	8,80
10	9	9	8	9	10	7	9	10	10	9,00	8,67	9,50	9,10
10	9	10	10	10	10	8	8	10	10	9,75	9,33	9,00	9,50
10	10	10	8	10	9	9	10	10	9	9,50	9,33	10,00	9,50
10	8	9	8	7	9	10	9	9	10	8,75	8,67	9,00	8,90
10	8	9	10	9	9	10	8	9	9	9,25	9,33	8,50	9,10
9	10	10	9	10	10	9	8	9	10	9,50	9,67	8,50	9,40
9	9	10	10	9	9	10	8	10	10	9,50	9,33	9,00	9,40
9	8	10	9	9	10	9	9	10	8	9,00	9,33	9,50	9,10
9	8	9	9	9	7	10	10	10	9	8,75	8,67	10,00	9,00
9	9	10	10	9	9	9	8	10	10	9,50	9,00	9,00	9,30
9	8	10	8	9	9	10	9	10	9	8,75	9,33	9,50	9,10
9	8	7	9	9	10	10	9	8	10	8,25	9,67	8,50	8,90
10	9	10	10	10	10	8	10	9	10	9,75	9,33	9,50	9,60
10	9	9	10	9	8	9	10	10	10	9,50	8,67	10,00	9,40
9	8	8	9	10	9	10	9	9	10	8,50	9,67	9,00	9,10
10	9	10	10	10	10	10	9	9	9	9,75	10,00	9,00	9,60
10	8	10	9	10	10	10	10	10	9	9,25	10,00	10,00	9,60
10	9	10	10	9	10	7	9	10	10	9,75	8,67	9,50	9,40
10	10	9	9	9	10	8	10	10	10	9,50	9,00	10,00	9,50
10	7	9	10	10	9	9	9	10	10	9,00	9,33	9,50	9,30
10	7	9	9	9	9	9	9	10	10	8,75	9,00	10,00	9,20
10	10	10	10	9	10	9	10	7	9	10,00	9,33	8,50	9,40

Gesamtergebnis Wrapper mit Klassifizierer: „SVM“ über alle 3 Suchmethoden

NR	Platz	Name	Gesamt	Prozent
1	168	HEIM	2,37	23,67
2	165	GAST	2,63	26,33
3	101	POSHEIM	3,23	32,33
4	126	PLUSMINUSHEIM	3,13	31,33
5	119	SIEGEHEIM	3,17	31,67
6	140	REMISHEIM	3,07	30,67
7	86	NIEDERLAGENHEIM	3,30	33,00
8	126	PUNKTEHEIM	3,13	31,33
9	94	TDHEIM	3,27	32,67
10	134	POSGAST	3,10	31,00
11	134	PLUSMINUSGAST	3,10	31,00
12	94	SIEGEGAST	3,27	32,67
13	111	REMISGAST	3,20	32,00
14	140	NIEDERLAGENGAST	3,07	30,67
15	75	PUNKTEGAST	3,33	33,33
16	57	TDGAST	3,47	34,67
17	166	MONAT	2,53	25,33
18	164	JAHRESZEIT	2,67	26,67
19	147	TABDIFF	3,03	30,33
20	70	TABTENDENZ	3,37	33,67
21	158	TABDIFFFORM	2,90	29,00
22	158	TABDIFFFORMVERG	2,90	29,00
23	101	SIEGDIFF	3,23	32,33
24	152	REMISDIFF	2,97	29,67
25	140	NIEDDIFF	3,07	30,67
26	86	PUNKTEDIFF	3,30	33,00
27	30	TDDIFF	4,03	40,33
28	152	TORDIFFFORM	2,97	29,67
29	29	TFHEIM	4,07	40,67
30	156	TFGAST	2,93	29,33
31	119	HEIMSTARK	3,17	31,67
32	148	GASTSTARK	3,00	30,00
33	166	SIEGEHEIM	2,53	25,33
34	86	VERSUS1HEIM	3,30	33,00
35	111	VERSUS1GAST	3,20	32,00
36	119	VERSUS1	3,17	31,67
37	70	VERSUS2HEIM	3,37	33,67
38	111	VERSUS2GAST	3,20	32,00
39	148	VERSUS2	3,00	30,00
40	126	VERSUS3HEIM	3,13	31,33
41	119	VERSUS3GAST	3,17	31,67
42	152	VERSUS3	2,97	29,67
43	119	VERSUS4HEIM	3,17	31,67
44	134	VERSUS4GAST	3,10	31,00
45	119	VERSUS4	3,17	31,67

46	148	VERSUS5HEIM	3,00	30,00
47	140	VERSUS5GAST	3,07	30,67
48	161	VERSUS5	2,83	28,33
49	161	TABGAST	2,83	28,33
50	163	TABVERGLEICH	2,77	27,67
51	134	FORM1HEIMHEIM	3,10	31,00
52	75	FORM1GASTAUSW	3,33	33,33
53	70	FORM1HEIM	3,37	33,67
54	75	FORM1GAST	3,33	33,33
55	8	DIFFFORM1	4,93	49,33
56	101	FORM2HEIMHEIM	3,23	32,33
57	111	FORM2GASTAUSW	3,20	32,00
58	64	FORM2HEIM	3,40	34,00
59	75	FORM2GAST	3,33	33,33
60	5	DIFFFORM2	5,20	52,00
61	152	FORM3HEIMHEIM	2,97	29,67
62	86	FORM3GASTAUSW	3,30	33,00
63	126	FORM3HEIM	3,13	31,33
64	140	FORM3GAST	3,07	30,67
65	5	DIFFFORM3	5,20	52,00
66	75	FORM4HEIMHEIM	3,33	33,33
67	19	FORM4GASTAUSW	4,37	43,67
68	94	FORM4HEIM	3,27	32,67
69	126	FORM4GAST	3,13	31,33
70	3	DIFFFORM4	5,50	55,00
71	134	FORM5HEIMHEIM	3,10	31,00
72	64	FORM5GASTAUSW	3,40	34,00
73	86	FORM5HEIM	3,30	33,00
74	94	FORM5GAST	3,27	32,67
75	15	DIFFFORM5	4,57	45,67
76	169	VIELETORE	0,00	0,00
77	126	AVG1PLATZHEIM	3,13	31,33
78	22	AVG1PLATZGAST	4,23	42,33
79	1	DIFFAVG1	6,60	66,00
80	50	AVG2PLATZHEIM	3,50	35,00
81	75	AVG2PLATZGAST	3,33	33,33
82	18	DIFFAVG2	4,40	44,00
83	49	AVG3PLATZHEIM	3,53	35,33
84	37	AVG3PLATZGAST	3,77	37,67
85	9	DIFFAVG3	4,90	49,00
86	57	AVG4PLATZHEIM	3,47	34,67
87	64	AVG4PLATZGAST	3,40	34,00
88	11	DIFFAVG4	4,87	48,67
89	38	AVG5PLATZHEIM	3,67	36,67
90	33	AVG5PLATZGAST	3,90	39,00
91	16	DIFFAVG5	4,50	45,00
92	101	HAEUFIGSTEPLAZIERUNGHEIM	3,23	32,33

93	50	HAEUFIGSTEPLAZIERUNGGAST	3,50	35,00
94	50	ADDNATIONALTOP9HEIM	3,50	35,00
95	86	ADDNATIONALTOP9GAST	3,30	33,00
96	25	ADDNATIONALLAENDERSP.TOP9H.	4,17	41,67
97	70	ADDNATIONALLAENDERSP.TOP9G.	3,37	33,67
98	75	ADDNATIONALTOP9POSHEIM	3,33	33,33
99	61	ADDNATIONALTOP9POSGAST	3,43	34,33
100	101	ADDNAT.LAENDERSP.TOPPOS9H.	3,23	32,33
101	75	ADDNAT.LAENDERSP.TOPPOS9G.	3,33	33,33
102	101	ADDMITTELNATLSPHEIM	3,23	32,33
103	57	ADDMITTELNATLSPGAST	3,47	34,67
104	50	ADDLSPKTHEIM	3,50	35,00
105	101	ADDLSPKTGAST	3,23	32,33
106	33	ADDGESCHOSSENETOREHEIM	3,90	39,00
107	94	ADDGESCHOSSENETOREGAST	3,27	32,67
108	44	ADDDURCHGELASSENETOREHEIM	3,60	36,00
109	38	ADDDURCHGELASSENETOREGAST	3,67	36,67
110	40	ADDPOSITIONHEIM	3,63	36,33
111	64	ADDPOSITIONGAST	3,40	34,00
112	46	ADDSTADIONGROESSEHEIM	3,57	35,67
113	46	ADDSTADIONGROESSEGAST	3,57	35,67
114	50	ADD10ABSTIEGHEIM	3,50	35,00
115	25	ADD10ABSTIEGGAST	4,17	41,67
116	36	ADDBRAZILHEIM	3,80	38,00
117	64	ADDBRAZILGAST	3,40	34,00
118	57	ADDBRAZILNATHEIM	3,47	34,67
119	126	ADDBRAZILNATGAST	3,13	31,33
120	31	ADDGERMANYNATHEIM	3,97	39,67
121	101	ADDGERMANYNATGAST	3,23	32,33
122	70	ADDAUSTRIAHEIM	3,37	33,67
123	46	ADDAUSTRIAGAST	3,57	35,67
124	64	ADDAUSTRIANATHEIM	3,40	34,00
125	40	ADDAUSTRIANATGAST	3,63	36,33
126	28	ADDEWIGETABELLEHEIM	4,13	41,33
127	44	ADDEWIGETABELLEGAST	3,60	36,00
128	61	ADDUEFA1JAHRHEIM	3,43	34,33
129	94	ADDUEFA1JAHRGAST	3,27	32,67
130	61	ADDUEFA5JAHREHEIM	3,43	34,33
131	86	ADDUEFA5JAHREGAST	3,30	33,00
132	50	ADDPUNKTEHEIMHEIM	3,50	35,00
133	101	ADDPUNKTEGASTGAST	3,23	32,33
134	111	ABSTIEG1HEIM	3,20	32,00
135	12	ABSTIEG1GAST	4,73	47,33
136	119	DIFFABSTIEG1	3,17	31,67
137	75	ABSTIEG2HEIM	3,33	33,33
138	13	ABSTIEG2GAST	4,60	46,00
139	75	DIFFABSTIEG2	3,33	33,33

140	140	ABSTIEG3HEIM	3,07	30,67
141	22	ABSTIEG3GAST	4,23	42,33
142	75	DIFFABSTIEG3	3,33	33,33
143	158	ABSTIEG4HEIM	2,90	29,00
144	19	ABSTIEG4GAST	4,37	43,67
145	156	DIFFABSTIEG4	2,93	29,33
146	148	ABSTIEG5HEIM	3,00	30,00
147	40	ABSTIEG5GAST	3,63	36,33
148	134	DIFFABSTIEG5	3,10	31,00
149	4	DIFFPUNKTEHEIMGAST	5,37	53,67
150	25	DIFFPKTLS	4,17	41,67
151	86	DIFFGSTLS	3,30	33,00
152	35	DIFFKASSTLS	3,83	38,33
153	19	DIFFPOSLS	4,37	43,67
154	101	DIFFEWIGETABELLE	3,23	32,33
155	16	DIFFNAT	4,50	45,00
156	32	DIFNAT2	3,93	39,33
157	2	DIFFNATPOS	6,20	62,00
158	40	DIFNATPOS2	3,63	36,33
159	24	DIFFNATPOSMIT	4,20	42,00
160	111	DIFFSTADGR	3,20	32,00
161	126	DIFF10ABSTIEG	3,13	31,33
162	140	DIFFBRAZIL	3,07	30,67
163	94	DIFFBRAZILN	3,27	32,67
164	9	DIFFGERMANYN	4,90	49,00
165	111	DIFFAUSTRIA	3,20	32,00
166	111	DIFFAUSTRIAN	3,20	32,00
167	13	DIFFUEFA1J	4,60	46,00
168	7	DIFFUEFA5J	5,03	50,33
169	50	DIFFHAEUFPLAZ	3,50	35,00

Wrapper mit Klassifizierer: „Random Forest“ und Suchmethode: „RankSearch“

NR	Platz	Name	24	35	46	57	25	36	47	26	37	27	3SD	4SD	5SD	Ges
1	137	HEIM	2	4	1		3	1	2	4	2	1	1,75	2,00	3,00	2,00
2	134	GAST	2	4	1		3	1	2	4	3	1	1,75	2,00	3,50	2,10
3	117	POSHEIM	2	4	3	1	3	2	2	4	3	2	2,50	2,33	3,50	2,60
4	154	PLUSMINUSHEIM	2	2			1	1		3	1		1,00	0,67	2,00	1,00
5	139	SIEGEHEIM	2	3	1		1	1	2	4	3	1	1,50	1,33	3,50	1,80
6	154	REMISHEIM	2	2			1	1	1	2	1		1,00	1,00	1,50	1,00
7	149	NIEDERLAGENHEIM	2	2	1		1	1	1	3	1		1,25	1,00	2,00	1,20
8	149	PUNKTEHEIM	2	2	1		1	2	1	2	1		1,25	1,33	1,50	1,20
9	84	TDHEIM	2	5	5	2	4	4	2	4	4	5	3,50	3,33	4,00	3,70
10	84	POSGAST	2	7	1	1	5	4	2	5	3	7	2,75	3,67	4,00	3,70
11	153	PLUSMINUSGAST	2	2	1		1	1		3	1		1,25	0,67	2,00	1,10
12	96	SIEGEGAST	2	7	1	1	5	3	3	5	4	4	2,75	3,67	4,50	3,50
13	146	REMISGAST	2	2	1		1	1	1	4	1		1,25	1,00	2,50	1,30
14	146	NIEDERLAGENGAST	2	2			2	1	1	4	1		1,00	1,33	2,50	1,30
15	122	PUNKTEGAST	2	5	1		3	2	2	4	3	2	2,00	2,33	3,50	2,40
16	92	TDGAST	2	7	3	2	5	2	2	4	4	5	3,50	3,00	4,00	3,60
17	156	MONAT	2	2					1	1	1		1,00	0,33	1,00	0,70
18	165	JAHRESZEIT	2	1							1		0,75	0,00	0,50	0,40
19	73	TABDIFF	5	7	3	1	5	3	2	5	5	4	4,00	3,33	5,00	4,00
20	134	TABTENDENZ	2	4	1		3	1	2	4	3	1	1,75	2,00	3,50	2,10
21	169	TABDIFFFORM	2										0,50	0,00	0,00	0,20
22	162	TABDIFFFORMVERG	1	1	1				1			1	0,75	0,33	0,50	0,50
23	77	SIEGDIFF	3	7	2	2	5	5	2	5	5	3	3,50	4,00	5,00	3,90
24	156	REMISDIFF	2	1	1		1	1			1		1,00	0,67	0,50	0,70
25	82	NIEDDIFF	3	7	2	1	5	4	2	5	6	3	3,25	3,67	5,50	3,80
26	60	PUNKTEDIFF	2	8	3	5	5	7	2	5	5	4	4,50	4,67	5,00	4,60
27	32	TDDIFF	6	8	5	1	7	6	2	7	5	7	5,00	5,00	6,00	5,40
28	126	TORDIFFFORM	2	4	1		3	2	2	5	3	1	1,75	2,33	4,00	2,30
29	28	TFHEIM	2	7	7	3	7	5	5	7	7	7	4,75	5,67	7,00	5,70
30	143	TFGAST	1	1			2	4			3	5	1,00	1,33	1,50	1,60
31	96	HEIMSTARK	2	5	3	2	4	2	2	5	4	6	3,00	2,67	4,50	3,50
32	165	GASTSTARK	2						2				0,50	0,67	0,00	0,40
33	162	SIEGEHEIM	2	2	1								1,25	0,00	0,00	0,50
34	115	VERSUS1HEIM	3	8			8	1	2	3	1	1	2,75	3,67	2,00	2,70
35	112	VERSUS1GAST	7	4	2		4	1	2	4	3	1	3,25	2,33	3,50	2,80
36	73	VERSUS1	2	8	3	1	5	5	2	5	5	4	3,50	4,00	5,00	4,00
37	139	VERSUS2HEIM	2	4			5	1	1	3	1	1	1,50	2,33	2,00	1,80
38	115	VERSUS2GAST	8	4			4	1	2	4	3	1	3,00	2,33	3,50	2,70
39	77	VERSUS2	2	7	5	1	5	5	2	5	4	3	3,75	4,00	4,50	3,90
40	146	VERSUS3HEIM	2	2			2	1	1	3	2		1,00	1,33	2,50	1,30
41	126	VERSUS3GAST	4	4			3	1	2	4	3	2	2,00	2,00	3,50	2,30
42	77	VERSUS3	2	7	3		6	5	2	5	4	5	3,00	4,33	4,50	3,90
43	149	VERSUS4HEIM	2	1			2	1	1	3	2		0,75	1,33	2,50	1,20
44	117	VERSUS4GAST	4	4			5	1	2	4	3	3	2,00	2,67	3,50	2,60
45	101	VERSUS4	3	6	1		6	3	2	5	4	4	2,50	3,67	4,50	3,40

46	149	VERSUS5HEIM	2	1	2	1	1	3	2	0,75	1,33	2,50	1,20			
47	112	VERSUS5GAST	4	4	7	1	2	4	3	3	2,00	3,33	3,50	2,80		
48	92	VERSUS5	2	7	1	1	6	5	2	5	4	3	2,75	4,33	4,50	3,60
49	145	TABGAST	2	4	1	1	2	2	2		1,50	1,33	2,00	1,40		
50	107	TABVERGLEICH	2	6	1	1	4	2	2	5	5	4	2,50	2,67	5,00	3,20
51	130	FORM1HEIMHEIM	2	4	1	3	1	2	4	3	2	1,75	2,00	3,50	2,20	
52	51	FORM1GASTAUSW	4	8	1	1	7	6	2	6	6	7	3,50	5,00	6,00	4,80
53	60	FORM1HEIM	4	6	4	2	5	4	3	6	6	6	4,00	4,00	6,00	4,60
54	41	FORM1GAST	6	8	1	1	8	6	2	6	6	7	4,00	5,33	6,00	5,10
55	8	DIFFFORM1	6	10	6	4	8	7	4	7	7	8	6,50	6,33	7,00	6,70
56	122	FORM2HEIMHEIM	2	4	1	1	3	2	2	4	3	2	2,00	2,33	3,50	2,40
57	41	FORM2GASTAUSW	3	8	3	1	8	7	2	6	6	7	3,75	5,67	6,00	5,10
58	60	FORM2HEIM	4	7	5	4	3	4	3	5	4	7	5,00	3,33	4,50	4,60
59	41	FORM2GAST	4	9	3	1	8	5	2	6	6	7	4,25	5,00	6,00	5,10
60	5	DIFFFORM2	6	10	6	5	8	7	5	6	7	8	6,75	6,67	6,50	6,80
61	138	FORM3HEIMHEIM	1	3			3	1	1	4	3	3	1,00	1,67	3,50	1,90
62	51	FORM3GASTAUSW	4	8	4	1	7	5	2	6	4	7	4,25	4,67	5,00	4,80
63	105	FORM3HEIM	4	5	1	1	4	2	2	5	4	5	2,75	2,67	4,50	3,30
64	55	FORM3GAST	2	7	3	2	7	6	2	6	5	7	3,50	5,00	5,50	4,70
65	1	DIFFFORM3	6	10	9	5	8	7	5	7	8	7	7,50	6,67	7,50	7,20
66	142	FORM4HEIMHEIM	2	2			2	1	1	4	3	2	1,00	1,33	3,50	1,70
67	31	FORM4GASTAUSW	4	8	6	3	7	7	2	6	5	7	5,25	5,33	5,50	5,50
68	77	FORM4HEIM	5	7	1	1	5	2	2	6	4	6	3,50	3,00	5,00	3,90
69	40	FORM4GAST	5	8	4	2	7	5	2	6	6	7	4,75	4,67	6,00	5,20
70	2	DIFFFORM4	5	10	6	6	8	7	6	7	7	7	6,75	7,00	7,00	6,90
71	130	FORM5HEIMHEIM	2	3	1	1	3	1	2	4	3	2	1,75	2,00	3,50	2,20
72	34	FORM5GASTAUSW	4	8	4	1	8	7	2	6	6	7	4,25	5,67	6,00	5,30
73	49	FORM5HEIM	3	7	5	2	6	5	3	6	6	6	4,25	4,67	6,00	4,90
74	32	FORM5GAST	5	9	4	1	8	6	2	6	6	7	4,75	5,33	6,00	5,40
75	8	DIFFFORM5	6	10	6	6	8	8	3	6	7	7	7,00	6,33	6,50	6,70
76	156	VIELETORE	2	2			1	1			1		1,00	0,67	0,50	0,70
77	73	AVG1PLATZHEIM	3	5	4	2	5	2	3	5	4	7	3,50	3,33	4,50	4,00
78	26	AVG1PLATZGAST	4	10	6	1	7	7	3	6	7	7	5,25	5,67	6,50	5,80
79	2	DIFFAVG1	6	9	6	6	8	7	6	7	7	7	6,75	7,00	7,00	6,90
80	84	AVG2PLATZHEIM	2	7	2	1	7	3	2	5	3	5	3,00	4,00	4,00	3,70
81	60	AVG2PLATZGAST	4	7	2	1	7	5	2	6	6	6	3,50	4,67	6,00	4,60
82	18	DIFFAVG2	6	8	5	4	8	5	5	7	7	7	5,75	6,00	7,00	6,20
83	66	AVG3PLATZHEIM	4	8	2	3	7	2	4	5	4	5	4,25	4,33	4,50	4,40
84	34	AVG3PLATZGAST	4	9	4	2	7	6	2	6	6	7	4,75	5,00	6,00	5,30
85	2	DIFFAVG3	6	10	6	6	8	7	6	6	7	7	7,00	7,00	6,50	6,90
86	101	AVG4PLATZHEIM	2	6	2	2	3	2	2	5	4	6	3,00	2,33	4,50	3,40
87	51	AVG4PLATZGAST	3	8	6	1	5	6	2	5	6	6	4,50	4,33	5,50	4,80
88	10	DIFFAVG4	6	8	6	6	8	7	6	6	6	7	6,50	7,00	6,00	6,60
89	84	AVG5PLATZHEIM	2	7	2	1	7	3	2	5	3	5	3,00	4,00	4,00	3,70
90	60	AVG5PLATZGAST	4	7	2	1	7	5	2	6	6	6	3,50	4,67	6,00	4,60
91	23	DIFFAVG5	6	8	5	4	8	5	2	7	7	7	5,75	5,00	7,00	5,90
92	120	HAEUFIGSTEPLAZIERUNGHEIM	2	4	1	1	3	2	2	4	4	2	2,00	2,33	4,00	2,50

93	110	HAEUFIGSTEPLAZIERUNGSGAST	2	6	1	1	5	1	2	5	3	3	2,50	2,67	4,00	2,90
94	77	ADDNATIONALTOP9HEIM	3	6	2	2	6	2	2	5	4	7	3,25	3,33	4,50	3,90
95	41	ADDNATIONALTOP9GAST	5	9	2	1	7	6	2	6	6	7	4,25	5,00	6,00	5,10
96	55	ADDNATIONALLAENDERSP.TOP9H.	6	7	3	1	8	2	2	7	4	7	4,25	4,00	5,50	4,70
97	105	ADDNATIONALLAENDERSP.TOP9G.	3	6	1	1	5	2	2	5	3	5	2,75	3,00	4,00	3,30
98	92	ADDNATIONALTOP9POSHEIM	3	5	2	1	5	2	2	5	4	7	2,75	3,00	4,50	3,60
99	41	ADDNATIONALTOP9POSGAST	5	9	2	1	7	6	2	6	6	7	4,25	5,00	6,00	5,10
100	82	ADDNAT.LAENDERSP.TOPPOS9H.	6	7	2		6	3	2	5	4	3	3,75	3,67	4,50	3,80
101	117	ADDNAT.LAENDERSP.TOPPOS9G.	2	5	1	1	4	1	2	5	3	2	2,25	2,33	4,00	2,60
102	84	ADDMITTELNATLSPHEIM	3	7	3	1	6	2	2	5	3	5	3,50	3,33	4,00	3,70
103	96	ADDMITTELNATLSPGAST	3	7	1	1	5	2	2	5	4	5	3,00	3,00	4,50	3,50
104	70	ADDLSPKTHEIM	4	7	2	2	5	2	4	5	6	4	3,75	3,67	5,50	4,10
105	55	ADDLSPKTGAST	2	8	3	1	7	6	2	5	6	7	3,50	5,00	5,50	4,70
106	68	ADDGESCHOSSENTOREHEIM	2	7	6	1	3	6	5	4	6	2	4,00	4,67	5,00	4,20
107	112	ADDGESCHOSSENTOREGAST	2	7		1	5	1	2	5	3	2	2,50	2,67	4,00	2,80
108	122	ADD DURCHGELASSEN TOREHEIM	2	4	1	1	3	2	2	4	3	2	2,00	2,33	3,50	2,40
109	126	ADD DURCHGELASSEN TOREGAST	2	5	1		3	1	2	4	3	2	2,00	2,00	3,50	2,30
110	46	ADDPOSITIONHEIM	2	8	6	2	6	6	4	5	7	4	4,50	5,33	6,00	5,00
111	101	ADDPOSITIONGAST	2	7	1	1	6	3	2	5	3	4	2,75	3,67	4,00	3,40
112	134	ADDSTADIONGROESSEHEIM	2	2	2	1	3	1	2	4	2	2	1,75	2,00	3,00	2,10
113	120	ADDSTADIONGROESSEGAST	2	4	1		5	1	2	5	3	2	1,75	2,67	4,00	2,50
114	139	ADD10ABSTIEGHEIM	2	3			2	1	2	4	3	1	1,25	1,67	3,50	1,80
115	70	ADD10ABSTIEGGAST	4	7	1	1	7	3	2	5	4	7	3,25	4,00	4,50	4,10
116	144	ADDBRAZILHEIM	1	4	1				1	2	4	2	1,50	1,00	3,00	1,50
117	130	ADDBRAZILGAST	2	4	1		3	2	2	4	3	1	1,75	2,33	3,50	2,20
118	130	ADDBRAZILNATHEIM	2	4	1		3	1	2	4	3	2	1,75	2,00	3,50	2,20
119	126	ADDBRAZILNATGAST	2	5	1		4	1	2	4	3	1	2,00	2,33	3,50	2,30
120	73	ADDGERMANYNATHEIM	2	9	6		4	6	3	4	4	2	4,25	4,33	4,00	4,00
121	34	ADDGERMANYNATGAST	5	9	2	1	8	6	2	6	7	7	4,25	5,33	6,50	5,30
122	162	ADDAUSTRIAHEIM	2				1			1	1		0,50	0,33	1,00	0,50
123	156	ADDAUSTRIAGAST	2	1			1			2	1		0,75	0,33	1,50	0,70
124	160	ADDAUSTRIANATHEIM	2	2				1			1		1,00	0,33	0,50	0,60
125	168	ADDAUSTRIANATGAST	2						1				0,50	0,33	0,00	0,30
126	51	ADDEWIGETABELLEHEIM	3	7	2	1	7	7	2	6	6	7	3,25	5,33	6,00	4,80
127	23	ADDEWIGETABELLEGAST	3	8	4	5	7	7	5	6	7	7	5,00	6,33	6,50	5,90
128	101	ADDUEFA1JAHRHEIM	1	5	2	8	3	2	2	5	3	3	4,00	2,33	4,00	3,40
129	29	ADDUEFA1JAHRGAST	5	9	3	2	8	7	2	6	7	7	4,75	5,67	6,50	5,60
130	70	ADDUEFA5JAHREHEIM	2	6	3	5	4	2	3	5	4	7	4,00	3,00	4,50	4,10
131	55	ADDUEFA5JAHREGAST	4	8	3	1	7	6	2	5	5	6	4,00	5,00	5,00	4,70
132	108	ADDPUNKTEHEIMHEIM	2	8	2	2	2	3	2	4	3	2	3,50	2,33	3,50	3,00
133	60	ADDPUNKTEGASTGAST	2	8	4	1	7	6	2	5	5	6	3,75	5,00	5,00	4,60
134	84	ABSTIEG1HEIM	4	5	3		5	2	2	5	4	7	3,00	3,00	4,50	3,70
135	18	ABSTIEG1GAST	4	10	6	3	8	7	4	6	7	7	5,75	6,33	6,50	6,20
136	34	DIFFABSTIEG1	4	7	4	5	7	6	2	6	6	6	5,00	5,00	6,00	5,30
137	92	ABSTIEG2HEIM	3	5	3		5	2	2	5	4	7	2,75	3,00	4,50	3,60
138	15	ABSTIEG2GAST	4	10	6	2	8	8	4	7	7	7	5,50	6,67	7,00	6,30
139	46	DIFFABSTIEG2	4	7	4	1	7	6	2	6	6	7	4,00	5,00	6,00	5,00

140	84	ABSTIEG3HEIM	4	5	3	5	2	2	5	4	7	3,00	3,00	4,50	3,70	
141	21	ABSTIEG3GAST	4	10	6	2	8	7	3	6	7	7	5,50	6,00	6,50	6,00
142	34	DIFFABSTIEG3	5	8	5	1	7	6	2	6	6	7	4,75	5,00	6,00	5,30
143	84	ABSTIEG4HEIM	4	6	2	5	2	2	5	4	7	3,00	3,00	4,50	3,70	
144	20	ABSTIEG4GAST	4	10	6	2	8	7	3	7	7	7	5,50	6,00	7,00	6,10
145	55	DIFFABSTIEG4	3	7	3	1	7	6	2	6	6	6	3,50	5,00	6,00	4,70
146	96	ABSTIEG5HEIM	3	6	1	5	2	2	5	4	7	3,25	3,00	4,50	3,50	
147	29	ABSTIEG5GAST	3	9	5	2	8	7	2	6	7	7	4,75	5,67	6,50	5,60
148	67	DIFFABSTIEG5	2	7	3	1	5	5	2	6	5	7	3,25	4,00	5,50	4,30
149	5	DIFFPUNKTEHEIMGAST	4	10	6	6	7	7	5	7	9	7	6,50	6,33	8,00	6,80
150	26	DIFFPKTLS	4	8	6	3	7	6	5	6	6	7	5,25	6,00	6,00	5,80
151	34	DIFFGSTLS	4	10	3	1	8	6	2	6	6	7	4,50	5,33	6,00	5,30
152	49	DIFFKASSTLS	6	8	5	1	7	5	2	6	3	6	5,00	4,67	4,50	4,90
153	10	DIFFPOSLS	4	9	6	7	7	7	6	6	7	7	6,50	6,67	6,50	6,60
154	68	DIFFEWIGETABELLE	4	8	3	1	5	4	2	5	4	6	4,00	3,67	4,50	4,20
155	13	DIFFNAT	6	10	6	3	7	7	4	8	7	7	6,25	6,00	7,50	6,50
156	15	DIFNAT2	4	7	6	6	7	7	6	6	7	7	5,75	6,67	6,50	6,30
157	14	DIFFNATPOS	6	9	6	4	7	6	6	6	7	7	6,25	6,33	6,50	6,40
158	23	DIFNATPOS2	6	7	5	5	7	4	6	5	7	7	5,75	5,67	6,00	5,90
159	10	DIFFNATPOSMIT	6	10	6	5	7	7	5	7	6	7	6,75	6,33	6,50	6,60
160	110	DIFFSTADGR	4	3	3	5	1	2	5	3	3	3	2,50	2,67	4,00	2,90
161	96	DIFF10ABSTIEG	3	7	1	1	5	2	2	5	3	6	3,00	3,00	4,00	3,50
162	122	DIFFBRAZIL	2	5	2	3	2	2	4	3	1	2,25	2,33	3,50	2,40	
163	108	DIFBRAZILN	5	5	1	5	2	2	5	3	2	2	2,75	3,00	4,00	3,00
164	21	DIFFGERMANYN	6	9	5	3	7	8	2	6	7	7	5,75	5,67	6,50	6,00
165	160	DIFFAUSTRIA	2	2				1			1		1,00	0,33	0,50	0,60
166	165	DIFFAUSTRIAN	1				1	1		1			0,25	0,67	0,50	0,40
167	15	DIFFUEFA1J	4	9	5	8	8	6	2	7	7	7	6,50	5,33	7,00	6,30
168	5	DIFFUEFA5J	6	9	6	5	8	7	5	7	7	8	6,50	6,67	7,00	6,80
169	46	DIFFHAEUFPLAZ	5	7	4	1	7	5	2	6	6	7	4,25	4,67	6,00	5,00

Wrapper mit Klassifizierer: „Random Forest“ und Suchmethode: „LinearForwardSelection“

NR	Platz	Name	24	35	46	57	25	36	47	26	37	27	3SD	4SD	5SD	Ges
1	22	HEIM											0,00	0,00	0,00	0,00
2	12	GAST					1			1			0,25	0,33	0,00	0,20
3	22	POSHEIM											0,00	0,00	0,00	0,00
4	22	PLUSMINUSHEIM											0,00	0,00	0,00	0,00
5	22	SIEGEHEIM											0,00	0,00	0,00	0,00
6	22	REMISHEIM											0,00	0,00	0,00	0,00
7	22	NIEDERLAGENHEIM											0,00	0,00	0,00	0,00
8	22	PUNKTEHEIM											0,00	0,00	0,00	0,00
9	22	TDHEIM											0,00	0,00	0,00	0,00
10	22	POSGAST											0,00	0,00	0,00	0,00
11	22	PLUSMINUSGAST											0,00	0,00	0,00	0,00
12	22	SIEGEGAST											0,00	0,00	0,00	0,00
13	22	REMISGAST											0,00	0,00	0,00	0,00
14	22	NIEDERLAGENGAST											0,00	0,00	0,00	0,00
15	22	PUNKTEGAST											0,00	0,00	0,00	0,00
16	22	TDGAST											0,00	0,00	0,00	0,00
17	22	MONAT											0,00	0,00	0,00	0,00
18	22	JAHRESZEIT											0,00	0,00	0,00	0,00
19	22	TABDIFF											0,00	0,00	0,00	0,00
20	22	TABTENDENZ											0,00	0,00	0,00	0,00
21	22	TABDIFFFORM											0,00	0,00	0,00	0,00
22	22	TABDIFFFORMVERG											0,00	0,00	0,00	0,00
23	22	SIEGDIFF											0,00	0,00	0,00	0,00
24	22	REMISDIFF											0,00	0,00	0,00	0,00
25	22	NIEDDIFF											0,00	0,00	0,00	0,00
26	22	PUNKTEDIFF											0,00	0,00	0,00	0,00
27	22	TDDIFF											0,00	0,00	0,00	0,00
28	22	TORDIFFFORM											0,00	0,00	0,00	0,00
29	3	TFHEIM	1	2	6	1	1						2,50	0,33	0,00	1,10
30	4	TFGAST	3		2				1				1,25	0,00	0,50	0,60
31	2	HEIMSTARK	4		2	4	4			3	2	2	1,50	2,67	2,50	2,10
32	8	GASTSTARK	1	3									1,00	0,00	0,00	0,40
33	18	SIEGEHEIM					1						0,25	0,00	0,00	0,10
34	22	VERSUS1HEIM											0,00	0,00	0,00	0,00
35	22	VERSUS1GAST											0,00	0,00	0,00	0,00
36	22	VERSUS1											0,00	0,00	0,00	0,00
37	22	VERSUS2HEIM											0,00	0,00	0,00	0,00
38	22	VERSUS2GAST											0,00	0,00	0,00	0,00
39	22	VERSUS2											0,00	0,00	0,00	0,00
40	22	VERSUS3HEIM											0,00	0,00	0,00	0,00
41	22	VERSUS3GAST											0,00	0,00	0,00	0,00
42	22	VERSUS3											0,00	0,00	0,00	0,00
43	22	VERSUS4HEIM											0,00	0,00	0,00	0,00
44	22	VERSUS4GAST											0,00	0,00	0,00	0,00
45	22	VERSUS4											0,00	0,00	0,00	0,00

46	22	VERSUS5HEIM		0,00	0,00	0,00	0,00
47	22	VERSUS5GAST		0,00	0,00	0,00	0,00
48	22	VERSUS5		0,00	0,00	0,00	0,00
49	22	TABGAST		0,00	0,00	0,00	0,00
50	22	TABVERGLEICH		0,00	0,00	0,00	0,00
51	22	FORM1HEIMHEIM		0,00	0,00	0,00	0,00
52	22	FORM1GASTAUSW		0,00	0,00	0,00	0,00
53	22	FORM1HEIM		0,00	0,00	0,00	0,00
54	22	FORM1GAST		0,00	0,00	0,00	0,00
55	22	DIFFFORM1		0,00	0,00	0,00	0,00
56	22	FORM2HEIMHEIM		0,00	0,00	0,00	0,00
57	22	FORM2GASTAUSW		0,00	0,00	0,00	0,00
58	22	FORM2HEIM		0,00	0,00	0,00	0,00
59	22	FORM2GAST		0,00	0,00	0,00	0,00
60	22	DIFFFORM2		0,00	0,00	0,00	0,00
61	22	FORM3HEIMHEIM		0,00	0,00	0,00	0,00
62	22	FORM3GASTAUSW		0,00	0,00	0,00	0,00
63	22	FORM3HEIM		0,00	0,00	0,00	0,00
64	22	FORM3GAST		0,00	0,00	0,00	0,00
65	22	DIFFFORM3		0,00	0,00	0,00	0,00
66	22	FORM4HEIMHEIM		0,00	0,00	0,00	0,00
67	22	FORM4GASTAUSW		0,00	0,00	0,00	0,00
68	22	FORM4HEIM		0,00	0,00	0,00	0,00
69	22	FORM4GAST		0,00	0,00	0,00	0,00
70	22	DIFFFORM4		0,00	0,00	0,00	0,00
71	22	FORM5HEIMHEIM		0,00	0,00	0,00	0,00
72	22	FORM5GASTAUSW		0,00	0,00	0,00	0,00
73	22	FORM5HEIM		0,00	0,00	0,00	0,00
74	22	FORM5GAST		0,00	0,00	0,00	0,00
75	22	DIFFFORM5		0,00	0,00	0,00	0,00
76	12	VIELETORE	1 1	0,50	0,00	0,00	0,20
77	22	AVG1PLATZHEIM		0,00	0,00	0,00	0,00
78	22	AVG1PLATZGAST		0,00	0,00	0,00	0,00
79	22	DIFFAVG1		0,00	0,00	0,00	0,00
80	22	AVG2PLATZHEIM		0,00	0,00	0,00	0,00
81	22	AVG2PLATZGAST		0,00	0,00	0,00	0,00
82	22	DIFFAVG2		0,00	0,00	0,00	0,00
83	22	AVG3PLATZHEIM		0,00	0,00	0,00	0,00
84	22	AVG3PLATZGAST		0,00	0,00	0,00	0,00
85	22	DIFFAVG3		0,00	0,00	0,00	0,00
86	22	AVG4PLATZHEIM		0,00	0,00	0,00	0,00
87	22	AVG4PLATZGAST		0,00	0,00	0,00	0,00
88	22	DIFFAVG4		0,00	0,00	0,00	0,00
89	22	AVG5PLATZHEIM		0,00	0,00	0,00	0,00
90	22	AVG5PLATZGAST		0,00	0,00	0,00	0,00
91	22	DIFFAVG5		0,00	0,00	0,00	0,00
92	22	HAEUFIGSTEPLAZIERUNGHEIM		0,00	0,00	0,00	0,00

93	22	HAEUFIGSTEPLAZIERUNGSGAST				0,00	0,00	0,00	0,00
94	22	ADDNATIONALTOP9HEIM				0,00	0,00	0,00	0,00
95	22	ADDNATIONALTOP9GAST				0,00	0,00	0,00	0,00
96	22	ADDNATIONALLAENDERSP.TOP9H.				0,00	0,00	0,00	0,00
97	22	ADDNATIONALLAENDERSP.TOP9G.				0,00	0,00	0,00	0,00
98	22	ADDNATIONALTOP9POSHEIM				0,00	0,00	0,00	0,00
99	22	ADDNATIONALTOP9POSGAST				0,00	0,00	0,00	0,00
100	22	ADDNAT.LAENDERSP.TOPPOS9H.				0,00	0,00	0,00	0,00
101	22	ADDNAT.LAENDERSP.TOPPOS9G.				0,00	0,00	0,00	0,00
102	22	ADDMITTELNATLSPHEIM				0,00	0,00	0,00	0,00
103	22	ADDMITTELNATLSPGAST				0,00	0,00	0,00	0,00
104	22	ADDLSPKTHEIM				0,00	0,00	0,00	0,00
105	22	ADDLSPKTGAST				0,00	0,00	0,00	0,00
106	22	ADDGESCHOSSENETOREHEIM				0,00	0,00	0,00	0,00
107	22	ADDGESCHOSSENETOREGAST				0,00	0,00	0,00	0,00
108	22	ADD DURCHGELASSENETOREHEIM				0,00	0,00	0,00	0,00
109	22	ADD DURCHGELASSENETOREGAST				0,00	0,00	0,00	0,00
110	22	ADDPOSITIONHEIM				0,00	0,00	0,00	0,00
111	22	ADDPOSITIONGAST				0,00	0,00	0,00	0,00
112	22	ADDSTADIONGROESSEHEIM				0,00	0,00	0,00	0,00
113	10	ADDSTADIONGROESSEGAST		3		0,75	0,00	0,00	0,30
114	22	ADD10ABSTIEGHEIM				0,00	0,00	0,00	0,00
115	12	ADD10ABSTIEGGAST		1		0,25	0,33	0,00	0,20
116	22	ADDBRAZILHEIM				0,00	0,00	0,00	0,00
117	22	ADDBRAZILGAST				0,00	0,00	0,00	0,00
118	22	ADDBRAZILNATHEIM				0,00	0,00	0,00	0,00
119	22	ADDBRAZILNATGAST				0,00	0,00	0,00	0,00
120	22	ADDGERMANYNATHEIM				0,00	0,00	0,00	0,00
121	22	ADDGERMANYNATGAST				0,00	0,00	0,00	0,00
122	12	ADDAUSTRIAHEIM		2		0,50	0,00	0,00	0,20
123	22	ADDAUSTRIAGAST				0,00	0,00	0,00	0,00
124	10	ADDAUSTRIANATHEIM		3		0,75	0,00	0,00	0,30
125	4	ADDAUSTRIANATGAST		3 1 1 1		1,50	0,00	0,00	0,60
126	22	ADDEWIGETABELLEHEIM				0,00	0,00	0,00	0,00
127	22	ADDEWIGETABELLEGAST				0,00	0,00	0,00	0,00
128	22	ADDUEFA1JAHRHEIM				0,00	0,00	0,00	0,00
129	22	ADDUEFA1JAHRGAST				0,00	0,00	0,00	0,00
130	22	ADDUEFA5JAHRHEIM				0,00	0,00	0,00	0,00
131	18	ADDUEFA5JAHREGAST		1		0,25	0,00	0,00	0,10
132	22	ADDPUNKTEHEIMHEIM				0,00	0,00	0,00	0,00
133	22	ADDPUNKTEGASTGAST				0,00	0,00	0,00	0,00
134	22	ABSTIEG1HEIM				0,00	0,00	0,00	0,00
135	4	ABSTIEG1GAST		1 5		1,50	0,00	0,00	0,60
136	22	DIFFABSTIEG1				0,00	0,00	0,00	0,00
137	22	ABSTIEG2HEIM				0,00	0,00	0,00	0,00
138	7	ABSTIEG2GAST		1 2 2		1,25	0,00	0,00	0,50
139	22	DIFFABSTIEG2				0,00	0,00	0,00	0,00

140	22	ABSTIEG3HEIM												0,00	0,00	0,00	0,00											
141	12	ABSTIEG3GAST												0,50	0,00	0,00	0,20											
142	22	DIFFABSTIEG3												0,00	0,00	0,00	0,00											
143	22	ABSTIEG4HEIM												0,00	0,00	0,00	0,00											
144	18	ABSTIEG4GAST												0,25	0,00	0,00	0,10											
145	22	DIFFABSTIEG4												0,00	0,00	0,00	0,00											
146	22	ABSTIEG5HEIM												0,00	0,00	0,00	0,00											
147	18	ABSTIEG5GAST												0,25	0,00	0,00	0,10											
148	22	DIFFABSTIEG5												0,00	0,00	0,00	0,00											
149	22	DIFFPUNKTEHEIMGAST												0,00	0,00	0,00	0,00											
150	22	DIFFPKTLS												0,00	0,00	0,00	0,00											
151	22	DIFFGSTLS												0,00	0,00	0,00	0,00											
152	22	DIFFKASSTLS												0,00	0,00	0,00	0,00											
153	22	DIFFPOSLS												0,00	0,00	0,00	0,00											
154	22	DIFFEWIGETABELLE												0,00	0,00	0,00	0,00											
155	22	DIFFNAT												0,00	0,00	0,00	0,00											
156	22	DIFNAT2												0,00	0,00	0,00	0,00											
157	8	DIFFNATPOS											1	3	0,25	1,00	0,00	0,40										
158	12	DIFNATPOS2												2	0,00	0,67	0,00	0,20										
159	22	DIFFNATPOSMIT													0,00	0,00	0,00	0,00										
160	22	DIFFSTADGR													0,00	0,00	0,00	0,00										
161	22	DIFF10ABSTIEG													0,00	0,00	0,00	0,00										
162	22	DIFFBRAZIL													0,00	0,00	0,00	0,00										
163	22	DIFBRAZILN													0,00	0,00	0,00	0,00										
164	1	DIFFGERMANYN											10	6	1	3	5	10	9	10	10	10	5,00	8,00	10,00	7,40		
165	22	DIFFAUSTRIA														0,00	0,00	0,00	0,00									
166	22	DIFFAUSTRIAN														0,00	0,00	0,00	0,00									
167	22	DIFFUEFA1J														0,00	0,00	0,00	0,00									
168	22	DIFFUEFA5J														0,00	0,00	0,00	0,00									
169	22	DIFFHAEUFPLAZ															0,00	0,00	0,00	0,00								

Wrapper mit Klas.: „Random Forest“ und Suchmethode: „BestFirst-Backward Elimination“

NR	Platz	Name	24	35	46	57	25	36	47	26	37	27	3SD	4SD	5SD	Ges
1	110	HEIM	10	10	10	9	10	9	10	10	9	10	9,75	9,67	9,50	9,70
2	22	GAST	9	10	10	10	10	10	10	10	10	10	9,75	10,00	10,00	9,90
3	110	POSHEIM	10	10	10	10	10	9	10	9	10	9	10,00	9,67	9,50	9,70
4	110	PLUSMINUSHEIM	10	10	10	9	10	10	10	9	10	9	9,75	10,00	9,50	9,70
5	57	SIEGEHEIM	10	10	10	10	10	10	10	9	9	10	10,00	10,00	9,00	9,80
6	144	REMISHEIM	10	10	10	10	10	9	10	9	8	10	10,00	9,67	8,50	9,60
7	22	NIEDERLAGENHEIM	10	10	10	10	10	9	10	10	10	10	10,00	9,67	10,00	9,90
8	110	PUNKTEHEIM	10	10	10	10	10	8	10	10	9	10	10,00	9,33	9,50	9,70
9	22	TDHEIM	10	10	10	10	10	9	10	10	10	10	10,00	9,67	10,00	9,90
10	162	POSGAST	10	10	8	10	9	10	10	10	9	9	9,50	9,67	9,50	9,50
11	22	PLUSMINUSGAST	10	10	10	10	10	10	10	9	10	10	10,00	10,00	9,50	9,90
12	22	SIEGEGAST	10	9	10	10	10	10	10	10	10	10	9,75	10,00	10,00	9,90
13	22	REMISGAST	9	10	10	10	10	10	10	10	10	10	9,75	10,00	10,00	9,90
14	57	NIEDERLAGENGAST	10	10	10	10	10	9	10	10	9	10	10,00	9,67	9,50	9,80
15	1	PUNKTEGAST	10	10	10	10	10	10	10	10	10	10	10,00	10,00	10,00	10,00
16	22	TDGAST	10	10	9	10	10	10	10	10	10	10	9,75	10,00	10,00	9,90
17	1	MONAT	10	10	10	10	10	10	10	10	10	10	10,00	10,00	10,00	10,00
18	22	JAHRESZEIT	10	10	10	10	10	9	10	10	10	10	10,00	9,67	10,00	9,90
19	110	TABDIFF	10	9	10	10	9	10	10	10	10	9	9,75	9,67	10,00	9,70
20	57	TABTENDENZ	10	9	10	10	10	10	10	10	10	9	9,75	10,00	10,00	9,80
21	22	TABDIFFFORM	10	10	10	10	10	9	10	10	10	10	10,00	9,67	10,00	9,90
22	110	TABDIFFFORMVERG	10	9	10	10	10	10	9	10	9	10	9,75	9,67	9,50	9,70
23	144	SIEGDIFF	9	10	10	10	10	10	10	9	9	9	9,75	10,00	9,00	9,60
24	22	REMISDIFF	10	10	10	10	10	9	10	10	10	10	10,00	9,67	10,00	9,90
25	110	NIEDDIFF	10	9	10	9	10	9	10	10	10	10	9,50	9,67	10,00	9,70
26	57	PUNKTEDIFF	10	10	10	9	10	10	10	10	9	10	9,75	10,00	9,50	9,80
27	110	TDDIFF	10	10	10	8	9	10	10	10	10	10	9,50	9,67	10,00	9,70
28	110	TORDIFFFORM	8	10	10	10	10	10	9	10	10	10	9,50	9,67	10,00	9,70
29	110	TFHEIM	10	10	10	10	10	9	10	8	10	10	10,00	9,67	9,00	9,70
30	1	TFGAST	10	10	10	10	10	10	10	10	10	10	10,00	10,00	10,00	10,00
31	57	HEIMSTARK	9	10	10	10	10	10	10	10	10	9	9,75	10,00	10,00	9,80
32	110	GASTSTARK	10	9	10	10	9	9	10	10	10	10	9,75	9,33	10,00	9,70
33	144	SIEGEHEIM	10	10	10	9	10	9	8	10	10	10	9,75	9,00	10,00	9,60
34	22	VERSUS1HEIM	10	9	10	10	10	10	10	10	10	10	9,75	10,00	10,00	9,90
35	1	VERSUS1GAST	10	10	10	10	10	10	10	10	10	10	10,00	10,00	10,00	10,00
36	166	VERSUS1	10	10	9	10	9	10	9	10	8	9	9,75	9,33	9,00	9,40
37	22	VERSUS2HEIM	10	10	10	10	10	10	9	10	10	10	10,00	9,67	10,00	9,90
38	144	VERSUS2GAST	9	10	10	9	10	9	10	10	10	9	9,50	9,67	10,00	9,60
39	110	VERSUS2	10	10	10	9	10	10	10	10	8	10	9,75	10,00	9,00	9,70
40	110	VERSUS3HEIM	10	9	10	10	9	10	10	10	10	9	9,75	9,67	10,00	9,70
41	1	VERSUS3GAST	10	10	10	10	10	10	10	10	10	10	10,00	10,00	10,00	10,00
42	57	VERSUS3	9	10	10	10	10	9	10	10	10	10	9,75	9,67	10,00	9,80
43	1	VERSUS4HEIM	10	10	10	10	10	10	10	10	10	10	10,00	10,00	10,00	10,00
44	162	VERSUS4GAST	10	10	9	9	10	10	10	10	9	9	9,50	9,67	9,50	9,50
45	57	VERSUS4	10	10	10	10	10	8	10	10	10	10	10,00	9,33	10,00	9,80

46	22	VERSUS5HEIM	10 10 10 10 10 9 10 10 10 10 10 10 10,00 9,67 10,00 9,90
47	110	VERSUS5GAST	10 10 10 9 10 9 10 10 9 10 9,75 9,67 9,50 9,70
48	57	VERSUS5	10 9 10 10 10 10 10 10 9 10 10 9,75 10,00 9,50 9,80
49	22	TABGAST	10 10 10 10 10 10 10 10 10 10 9 10,00 10,00 10,00 9,90
50	1	TABVERGLEICH	10 10 10 10 10 10 10 10 10 10 10 10,00 10,00 10,00 10,00
51	22	FORM1HEIMHEIM	10 10 9 10 10 10 10 10 10 10 10 9,75 10,00 10,00 9,90
52	144	FORM1GASTAUSW	10 10 8 10 10 10 9 9 10 10 10 9,50 9,67 9,50 9,60
53	144	FORM1HEIM	10 8 10 10 9 9 10 10 10 10 10 9,50 9,33 10,00 9,60
54	110	FORM1GAST	10 9 10 10 10 9 10 9 10 10 9,75 9,67 9,50 9,70
55	57	DIFFFORM1	10 10 10 10 9 10 10 10 10 10 9 10,00 9,67 10,00 9,80
56	57	FORM2HEIMHEIM	10 10 10 9 10 10 10 10 10 10 9 9,75 10,00 10,00 9,80
57	1	FORM2GASTAUSW	10 10 10 10 10 10 10 10 10 10 10 10,00 10,00 10,00 10,00
58	110	FORM2HEIM	10 9 10 10 9 10 10 10 10 9 10 9,75 9,67 9,50 9,70
59	57	FORM2GAST	10 9 10 10 9 10 10 10 10 10 10 9,75 9,67 10,00 9,80
60	57	DIFFFORM2	10 10 10 10 10 10 10 9 9 10 10,00 10,00 9,00 9,80
61	57	FORM3HEIMHEIM	9 10 10 10 9 10 10 10 10 10 10 9,75 9,67 10,00 9,80
62	144	FORM3GASTAUSW	10 10 10 10 10 9 9 9 9 10 10,00 9,33 9,00 9,60
63	1	FORM3HEIM	10 10 10 10 10 10 10 10 10 10 10 10,00 10,00 10,00 10,00
64	57	FORM3GAST	9 10 10 10 10 9 10 10 10 10 10 9,75 9,67 10,00 9,80
65	144	DIFFFORM3	10 10 9 9 10 10 10 9 9 10 9,50 10,00 9,00 9,60
66	57	FORM4HEIMHEIM	10 10 10 9 9 10 10 10 10 10 10 9,75 9,67 10,00 9,80
67	22	FORM4GASTAUSW	10 10 10 10 10 10 9 10 10 10 10 10,00 9,67 10,00 9,90
68	22	FORM4HEIM	10 10 10 10 10 10 10 9 10 10 10 10,00 10,00 9,50 9,90
69	57	FORM4GAST	9 10 9 10 10 10 10 10 10 10 10 9,50 10,00 10,00 9,80
70	57	DIFFFORM4	10 10 10 10 8 10 10 10 10 10 10 10,00 9,33 10,00 9,80
71	57	FORM5HEIMHEIM	9 10 10 9 10 10 10 10 10 10 10 9,50 10,00 10,00 9,80
72	166	FORM5GASTAUSW	10 10 9 10 9 9 9 10 10 8 9,75 9,00 10,00 9,40
73	110	FORM5HEIM	10 9 10 9 10 10 9 10 10 10 10 9,50 9,67 10,00 9,70
74	1	FORM5GAST	10 10 10 10 10 10 10 10 10 10 10 10,00 10,00 10,00 10,00
75	144	DIFFFORM5	10 10 9 9 10 10 10 10 10 9 9 9,50 10,00 9,50 9,60
76	110	VIELETORE	10 10 9 9 10 10 10 10 10 10 9 9,50 10,00 10,00 9,70
77	110	AVG1PLATZHEIM	10 10 10 10 8 10 10 10 10 9 10 10,00 9,33 9,50 9,70
78	57	AVG1PLATZGAST	10 9 10 9 10 10 10 10 10 10 10 9,50 10,00 10,00 9,80
79	57	DIFFAVG1	10 10 9 10 10 9 10 10 10 10 10 9,75 9,67 10,00 9,80
80	110	AVG2PLATZHEIM	10 9 10 9 9 10 10 10 10 10 10 9,50 9,67 10,00 9,70
81	110	AVG2PLATZGAST	10 10 10 10 10 9 10 10 10 10 8 10,00 9,67 10,00 9,70
82	144	DIFFAVG2	10 10 10 10 9 10 10 9 8 10 10,00 9,67 8,50 9,60
83	144	AVG3PLATZHEIM	9 10 10 9 9 10 10 9 10 10 9,50 9,67 9,50 9,60
84	162	AVG3PLATZGAST	10 10 9 10 9 9 10 9 9 10 9,75 9,33 9,00 9,50
85	57	DIFFAVG3	9 10 10 10 10 9 10 9 10 10 10 9,75 9,67 10,00 9,80
86	57	AVG4PLATZHEIM	10 10 9 10 10 10 10 10 9 10 10 9,75 10,00 9,50 9,80
87	57	AVG4PLATZGAST	10 10 9 10 10 10 10 9 10 10 9,75 10,00 9,50 9,80
88	57	DIFFAVG4	10 10 10 10 10 9 10 9 10 10 10 10,00 9,67 9,50 9,80
89	110	AVG5PLATZHEIM	9 10 10 10 10 10 10 10 10 10 8 9,75 10,00 10,00 9,70
90	57	AVG5PLATZGAST	10 10 10 10 10 9 10 10 10 10 9 10,00 9,67 10,00 9,80
91	22	DIFFAVG5	10 10 10 9 10 10 10 10 10 10 10 9,75 10,00 10,00 9,90
92	22	HAEUFIGSTEPLAZIERUNGHEIM	10 10 10 9 10 10 10 10 10 10 10 9,75 10,00 10,00 9,90

93	22	HAEUFIGSTEPLAZIERUNGGAST	9 10 10 10 10 10 10 10 10 10 10 10 9,75 10,00 10,00 9,90
94	110	ADDNATIONALTOP9HEIM	10 9 10 9 10 10 10 10 10 9 10 9,50 10,00 9,50 9,70
95	1	ADDNATIONALTOP9GAST	10 10 10 10 10 10 10 10 10 10 10 10 10,00 10,00 10,00 10,00
96	1	ADDNATIONALLAENDERSP.TOP9H.	10 10 10 10 10 10 10 10 10 10 10 10 10,00 10,00 10,00 10,00
97	22	ADDNATIONALLAENDERSP.TOP9G.	10 9 10 10 10 10 10 10 10 10 10 10 9,75 10,00 10,00 9,90
98	57	ADDNATIONALTOP9POSHEIM	10 10 9 9 10 10 10 10 10 10 10 10 9,50 10,00 10,00 9,80
99	1	ADDNATIONALTOP9POSGAST	10 10 10 10 10 10 10 10 10 10 10 10 10,00 10,00 10,00 10,00
100	22	ADDNAT.LAENDERSP.TOPPOS9H.	10 10 10 10 10 10 10 9 10 10 10 10 10,00 10,00 9,50 9,90
101	57	ADDNAT.LAENDERSP.TOPPOS9G.	9 10 10 10 10 10 9 10 10 10 10 10 9,75 9,67 10,00 9,80
102	57	ADDMITTELNATLSPHEIM	10 10 10 10 10 10 9 10 9 10 10 10 10,00 9,67 9,50 9,80
103	57	ADDMITTELNATLSPGAST	9 10 10 10 9 10 10 10 10 10 10 10 9,75 9,67 10,00 9,80
104	110	ADDLSPKTHEIM	9 10 10 10 10 9 10 10 10 9 10 9 9,75 9,67 10,00 9,70
105	57	ADDLSPKTGAST	9 10 10 10 10 10 10 10 9 10 10 9 9,75 10,00 9,50 9,80
106	110	ADDGESCHOSSENTOREHEIM	10 10 9 10 9 10 10 9 10 10 9 10 9,75 9,67 9,50 9,70
107	57	ADDGESCHOSSENTOREGAST	10 10 10 10 10 9 10 9 10 10 10 10 10,00 9,67 9,50 9,80
108	110	ADDDURCHGELASSENTOREHEIM	9 10 10 10 10 9 10 9 10 10 9 10 9,75 9,67 9,50 9,70
109	1	ADDDURCHGELASSENTOREGAST	10 10 10 10 10 10 10 10 10 10 10 10 10,00 10,00 10,00 10,00
110	22	ADDPOSITIONHEIM	9 10 10 10 10 10 10 10 10 10 10 10 9,75 10,00 10,00 9,90
111	22	ADDPOSITIONGAST	10 10 10 10 10 9 10 10 10 10 10 10 10,00 9,67 10,00 9,90
112	1	ADDSTADIONGROESSEHEIM	10 10 10 10 10 10 10 10 10 10 10 10 10,00 10,00 10,00 10,00
113	57	ADDSTADIONGROESSEGAST	10 10 10 10 10 10 10 8 10 10 10 10 10,00 10,00 9,00 9,80
114	169	ADD10ABSTIEGHEIM	9 10 9 8 9 9 10 10 10 10 9 9,00 9,33 10,00 9,30
115	110	ADD10ABSTIEGGAST	9 10 10 9 10 10 10 10 9 10 9 9,50 10,00 9,50 9,70
116	144	ADDBRAZILHEIM	10 10 8 10 10 9 10 10 9 10 9 9,50 9,67 9,50 9,60
117	22	ADDBRAZILGAST	10 10 10 10 10 10 9 10 10 10 10 10 10,00 9,67 10,00 9,90
118	22	ADDBRAZILNATHEIM	9 10 10 10 10 10 10 10 10 10 10 10 9,75 10,00 10,00 9,90
119	57	ADDBRAZILNATGAST	10 10 10 10 10 10 9 10 10 10 9 10,00 9,67 10,00 9,80
120	22	ADDGERMANYNATHEIM	9 10 10 10 10 10 10 10 10 10 10 10 9,75 10,00 10,00 9,90
121	144	ADDGERMANYNATGAST	10 10 8 10 10 10 10 9 10 9 9,50 10,00 9,50 9,60
122	57	ADDAUSTRIAHEIM	10 10 10 10 10 10 10 10 10 9 9 10,00 10,00 9,50 9,80
123	110	ADDAUSTRIAGAST	9 9 10 10 10 9 10 10 10 10 10 9,50 9,67 10,00 9,70
124	110	ADDAUSTRIANATHEIM	10 9 10 10 10 9 10 10 9 10 9 9,75 9,67 9,50 9,70
125	57	ADDAUSTRIANATGAST	9 10 10 10 10 10 9 10 10 10 10 9,75 9,67 10,00 9,80
126	166	ADDEWIGETABELLEHEIM	9 10 9 9 9 10 9 10 10 9 9,25 9,33 10,00 9,40
127	22	ADDEWIGETABELLEGAST	10 9 10 10 10 10 10 10 10 10 10 9,75 10,00 10,00 9,90
128	110	ADDUEFA1JAHRHEIM	10 9 10 10 9 9 10 10 10 10 10 9,75 9,33 10,00 9,70
129	22	ADDUEFA1JAHRGAST	10 10 10 10 10 10 9 10 10 10 10 10,00 10,00 9,50 9,90
130	22	ADDUEFA5JAHREHEIM	10 10 10 10 10 10 10 10 10 10 10 9 10,00 10,00 10,00 9,90
131	110	ADDUEFA5JAHREGAST	10 10 9 8 10 10 10 10 10 10 10 9,25 10,00 10,00 9,70
132	57	ADDPUNKTEHEIMHEIM	10 10 10 9 10 10 10 9 10 10 10 9,75 10,00 9,50 9,80
133	57	ADDPUNKTEGASTGAST	10 9 10 10 10 9 10 10 10 10 10 9,75 9,67 10,00 9,80
134	22	ABSTIEG1HEIM	10 10 10 10 10 9 10 9 10 10 10 10,00 9,67 10,00 9,90
135	57	ABSTIEG1GAST	10 10 10 10 9 10 9 10 10 10 10 10,00 9,33 10,00 9,80
136	144	DIFFABSTIEG1	10 10 7 9 10 10 10 10 10 10 10 9,00 10,00 10,00 9,60
137	57	ABSTIEG2HEIM	10 10 9 10 9 10 10 10 10 10 10 9,75 9,67 10,00 9,80
138	57	ABSTIEG2GAST	9 10 10 9 10 10 10 10 10 10 10 9,50 10,00 10,00 9,80
139	22	DIFFABSTIEG2	10 10 10 10 10 9 10 10 10 10 10 10,00 9,67 10,00 9,90

140	1	ABSTIEG3HEIM	10 10 10 10 10 10 10 10 10 10 10 10 10,00 10,00 10,00 10,00
141	110	ABSTIEG3GAST	10 9 10 10 10 10 9 10 10 10 9 9,75 9,67 10,00 9,70
142	57	DIFFABSTIEG3	10 10 10 10 9 10 10 10 10 10 9 10,00 9,67 10,00 9,80
143	57	ABSTIEG4HEIM	9 10 10 10 9 10 10 10 10 10 10 9,75 9,67 10,00 9,80
144	144	ABSTIEG4GAST	9 10 10 10 10 9 10 10 10 8 10 9,75 9,67 9,00 9,60
145	57	DIFFABSTIEG4	10 10 9 10 10 10 9 10 10 10 10 9,75 9,67 10,00 9,80
146	57	ABSTIEG5HEIM	10 9 10 10 10 10 10 10 10 9 10 9,75 10,00 9,50 9,80
147	57	ABSTIEG5GAST	10 10 10 10 9 10 9 10 10 10 10 10,00 9,33 10,00 9,80
148	144	DIFFABSTIEG5	10 10 10 9 10 9 9 10 9 10 9,75 9,33 9,50 9,60
149	1	DIFFPUNKTEHEIMGAST	10 10 10 10 10 10 10 10 10 10 10 10,00 10,00 10,00 10,00
150	1	DIFFPKTLS	10 10 10 10 10 10 10 10 10 10 10 10,00 10,00 10,00 10,00
151	22	DIFFGSTLS	9 10 10 10 10 10 10 10 10 10 10 9,75 10,00 10,00 9,90
152	144	DIFFKASSTLS	10 9 9 10 10 10 10 10 9 10 9 9,50 10,00 9,50 9,60
153	57	DIFFPOSLS	9 10 10 10 10 10 10 10 9 10 9,75 10,00 9,50 9,80
154	57	DIFFEWIGETABELLE	10 10 9 10 10 10 10 10 9 10 9,75 10,00 9,50 9,80
155	144	DIFFNAT	10 10 9 10 8 9 10 10 10 10 10 9,75 9,00 10,00 9,60
156	57	DIFNAT2	10 10 10 10 10 10 10 10 9 9 10,00 10,00 9,50 9,80
157	57	DIFFNATPOS	10 10 9 10 10 10 10 10 9 10 10 9,75 10,00 9,50 9,80
158	110	DIFNATPOS2	10 10 10 10 10 10 10 9 9 9 10,00 10,00 9,00 9,70
159	22	DIFFNATPOSMIT	10 10 10 10 10 9 10 10 10 10 10 10,00 9,67 10,00 9,90
160	162	DIFFSTADGR	10 10 9 9 10 10 9 9 10 9 9,50 9,67 9,50 9,50
161	57	DIFF10ABSTIEG	10 10 10 9 10 9 10 10 10 10 10 9,75 9,67 10,00 9,80
162	1	DIFFBRAZIL	10 10 10 10 10 10 10 10 10 10 10 10,00 10,00 10,00 10,00
163	57	DIFBRAZILN	9 10 10 10 10 10 10 10 9 10 9,75 10,00 9,50 9,80
164	22	DIFFGERMANYN	10 10 10 10 10 10 10 10 9 10 10 10,00 10,00 9,50 9,90
165	57	DIFFAUSTRIA	10 8 10 10 10 10 10 10 10 10 10 9,50 10,00 10,00 9,80
166	1	DIFFAUSTRIAN	10 10 10 10 10 10 10 10 10 10 10 10,00 10,00 10,00 10,00
167	57	DIFFUEFA1J	10 9 10 10 10 10 10 9 10 10 9,75 10,00 9,50 9,80
168	110	DIFFUEFA5J	10 10 10 9 10 10 10 9 9 10 9,75 10,00 9,00 9,70
169	1	DIFFHAEUFPLAZ	10 10 10 10 10 10 10 10 10 10 10 10,00 10,00 10,00 10,00

Gesamtergebnis Wrapper mit Klassifizierer: „Random Forest“ über alle 3 Suchmethoden

NR	Platz	Name	Gesamt	Prozent
1	138	HEIM	3,90	39,00
2	125	GAST	4,07	40,67
3	121	POSHEIM	4,10	41,00
4	154	PLUSMINUSHEIM	3,57	35,67
5	141	SIEGEHEIM	3,87	38,67
6	157	REMISHEIM	3,53	35,33
7	146	NIEDERLAGENHEIM	3,70	37,00
8	153	PUNKTEHEIM	3,63	36,33
9	81	TDHEIM	4,53	45,33
10	102	POSGAST	4,40	44,00
11	151	PLUSMINUSGAST	3,67	36,67
12	90	SIEGEGAST	4,47	44,67
13	144	REMISGAST	3,73	37,33
14	146	NIEDERLAGENGAST	3,70	37,00
15	116	PUNKTEGAST	4,13	41,33
16	85	TDGAST	4,50	45,00
17	154	MONAT	3,57	35,67
18	166	JAHRESZEIT	3,43	34,33
19	76	TABDIFF	4,57	45,67
20	137	TABTENDENZ	3,97	39,67
21	169	TABDIFFFORM	3,37	33,67
22	167	TABDIFFFORMVERG	3,40	34,00
23	85	SIEGDIFF	4,50	45,00
24	157	REMISDIFF	3,53	35,33
25	85	NIEDDIFF	4,50	45,00
26	58	PUNKTEDIFF	4,80	48,00
27	35	TDDIFF	5,03	50,33
28	135	TORDIFFFORM	4,00	40,00
29	11	TFHEIM	5,50	55,00
30	125	TFGAST	4,07	40,67
31	31	HEIMSTARK	5,13	51,33
32	161	GASTSTARK	3,50	35,00
33	167	SIEGEHEIM	3,40	34,00
34	112	VERSUS1HEIM	4,20	42,00
35	108	VERSUS1GAST	4,27	42,67
36	90	VERSUS1	4,47	44,67
37	138	VERSUS2HEIM	3,90	39,00
38	121	VERSUS2GAST	4,10	41,00
39	81	VERSUS2	4,53	45,33
40	151	VERSUS3HEIM	3,67	36,67
41	121	VERSUS3GAST	4,10	41,00
42	76	VERSUS3	4,57	45,67
43	144	VERSUS4HEIM	3,73	37,33
44	128	VERSUS4GAST	4,03	40,33
45	102	VERSUS4	4,40	44,00

46	146	VERSUS5HEIM	3,70	37,00
47	115	VERSUS5GAST	4,17	41,67
48	90	VERSUS5	4,47	44,67
49	143	TABGAST	3,77	37,67
50	102	TABVERGLEICH	4,40	44,00
51	128	FORM1HEIMHEIM	4,03	40,33
52	58	FORM1GASTAUSW	4,80	48,00
53	65	FORM1HEIM	4,73	47,33
54	47	FORM1GAST	4,93	49,33
55	11	DIFFFORM1	5,50	55,00
56	125	FORM2HEIMHEIM	4,07	40,67
57	35	FORM2GASTAUSW	5,03	50,33
58	63	FORM2HEIM	4,77	47,67
59	42	FORM2GAST	4,97	49,67
60	7	DIFFFORM2	5,53	55,33
61	138	FORM3HEIMHEIM	3,90	39,00
62	58	FORM3GASTAUSW	4,80	48,00
63	97	FORM3HEIM	4,43	44,33
64	53	FORM3GAST	4,83	48,33
65	2	DIFFFORM3	5,60	56,00
66	142	FORM4HEIMHEIM	3,83	38,33
67	31	FORM4GASTAUSW	5,13	51,33
68	74	FORM4HEIM	4,60	46,00
69	40	FORM4GAST	5,00	50,00
70	4	DIFFFORM4	5,57	55,67
71	135	FORM5HEIMHEIM	4,00	40,00
72	49	FORM5GASTAUSW	4,90	49,00
73	51	FORM5HEIM	4,87	48,67
74	31	FORM5GAST	5,13	51,33
75	17	DIFFFORM5	5,43	54,33
76	157	VIELETORE	3,53	35,33
77	76	AVG1PLATZHEIM	4,57	45,67
78	28	AVG1PLATZGAST	5,20	52,00
79	4	DIFFAVG1	5,57	55,67
80	90	AVG2PLATZHEIM	4,47	44,67
81	63	AVG2PLATZGAST	4,77	47,67
82	22	DIFFAVG2	5,27	52,67
83	67	AVG3PLATZHEIM	4,67	46,67
84	47	AVG3PLATZGAST	4,93	49,33
85	4	DIFFAVG3	5,57	55,67
86	102	AVG4PLATZHEIM	4,40	44,00
87	51	AVG4PLATZGAST	4,87	48,67
88	15	DIFFAVG4	5,47	54,67
89	90	AVG5PLATZHEIM	4,47	44,67
90	58	AVG5PLATZGAST	4,80	48,00
91	22	DIFFAVG5	5,27	52,67
92	116	HAEUFIGSTEPLAZIERUNGHEIM	4,13	41,33

93	108	HAEUFIGSTEPLAZIERUNGGAST	4,27	42,67
94	81	ADDNATIONALTOP9HEIM	4,53	45,33
95	35	ADDNATIONALTOP9GAST	5,03	50,33
96	49	ADDNATIONALLAENDERSP.TOP9H.	4,90	49,00
97	102	ADDNATIONALLAENDERSP.TOP9G.	4,40	44,00
98	90	ADDNATIONALTOP9POSHEIM	4,47	44,67
99	35	ADDNATIONALTOP9POSGAST	5,03	50,33
100	76	ADDNAT.LAENDERSP.TOPPOS9H.	4,57	45,67
101	116	ADDNAT.LAENDERSP.TOPPOS9G.	4,13	41,33
102	85	ADDMITTELNATLSPHEIM	4,50	45,00
103	97	ADDMITTELNATLSPGAST	4,43	44,33
104	74	ADDLSPKTHEIM	4,60	46,00
105	53	ADDLSPKTGAST	4,83	48,33
106	71	ADDGESCHOSSENETOREHEIM	4,63	46,33
107	112	ADDGESCHOSSENETOREGAST	4,20	42,00
108	128	ADDDURCHGELASSENETOREHEIM	4,03	40,33
109	121	ADDDURCHGELASSENETOREGAST	4,10	41,00
110	42	ADDPOSITIONHEIM	4,97	49,67
111	97	ADDPOSITIONGAST	4,43	44,33
112	128	ADDSTADIONGROESSEHEIM	4,03	40,33
113	112	ADDSTADIONGROESSEGAST	4,20	42,00
114	146	ADD10ABSTIEGHEIM	3,70	37,00
115	67	ADD10ABSTIEGGAST	4,67	46,67
116	146	ADDBRAZILHEIM	3,70	37,00
117	128	ADDBRAZILGAST	4,03	40,33
118	128	ADDBRAZILNATHEIM	4,03	40,33
119	128	ADDBRAZILNATGAST	4,03	40,33
120	71	ADDGERMANYNATHEIM	4,63	46,33
121	42	ADDGERMANYNATGAST	4,97	49,67
122	161	ADDAUSTRIAHEIM	3,50	35,00
123	163	ADDAUSTRIAGAST	3,47	34,67
124	157	ADDAUSTRIANATHEIM	3,53	35,33
125	154	ADDAUSTRIANATGAST	3,57	35,67
126	65	ADDEWIGETABELLEHEIM	4,73	47,33
127	22	ADDEWIGETABELLEGAST	5,27	52,67
128	107	ADDUEFA1JAHRHEIM	4,37	43,67
129	29	ADDUEFA1JAHRGAST	5,17	51,67
130	67	ADDUEFA5JAHREHEIM	4,67	46,67
131	53	ADDUEFA5JAHREGAST	4,83	48,33
132	108	ADDPUNKTEHEIMHEIM	4,27	42,67
133	58	ADDPUNKTEGASTGAST	4,80	48,00
134	81	ABSTIEG1HEIM	4,53	45,33
135	7	ABSTIEG1GAST	5,53	55,33
136	42	DIFFABSTIEG1	4,97	49,67
137	90	ABSTIEG2HEIM	4,47	44,67
138	7	ABSTIEG2GAST	5,53	55,33
139	42	DIFFABSTIEG2	4,97	49,67

140	76	ABSTIEG3HEIM	4,57	45,67
141	21	ABSTIEG3GAST	5,30	53,00
142	35	DIFFABSTIEG3	5,03	50,33
143	85	ABSTIEG4HEIM	4,50	45,00
144	22	ABSTIEG4GAST	5,27	52,67
145	53	DIFFABSTIEG4	4,83	48,33
146	97	ABSTIEG5HEIM	4,43	44,33
147	29	ABSTIEG5GAST	5,17	51,67
148	71	DIFFABSTIEG5	4,63	46,33
149	2	DIFFPUNKTEHEIMGAST	5,60	56,00
150	22	DIFFPKTLS	5,27	52,67
151	34	DIFFGSTLS	5,07	50,67
152	53	DIFFKASSTLS	4,83	48,33
153	15	DIFFPOSLS	5,47	54,67
154	67	DIFFEWIGETABELLE	4,67	46,67
155	18	DIFFNAT	5,37	53,67
156	18	DIFNAT2	5,37	53,67
157	7	DIFFNATPOS	5,53	55,33
158	22	DIFNATPOS2	5,27	52,67
159	11	DIFFNATPOSMIT	5,50	55,00
160	116	DIFFSTADGR	4,13	41,33
161	97	DIFF10ABSTIEG	4,43	44,33
162	116	DIFFBRAZIL	4,13	41,33
163	108	DIFBRAZILN	4,27	42,67
164	1	DIFFGERMANYN	7,77	77,67
165	163	DIFFAUSTRIA	3,47	34,67
166	163	DIFFAUSTRIAN	3,47	34,67
167	18	DIFFUEFA1J	5,37	53,67
168	11	DIFFUEFA5J	5,50	55,00
169	40	DIFFHAEUFPLAZ	5,00	50,00

InfoGain-Ergebnisse

NR	Platz	Name	24	35	46	57	25	36	47	26	37	27	3SD	4SD	5SD	Ges
1	2	HEIM	8,8	8,1	7,7	8,2	7,7	6,5	7,4	6,3	6,3	6,2	8,20	7,20	6,30	7,32
2	1	GAST	9,7	10,8	8,4	8,3	8,8	8,9	7,5	7,7	7,7	7,1	9,30	8,40	7,70	8,49
3	117	POSHEIM	1,5	2,2	2,8	2,6	1,6	2,5	2,5	2	2,4	1,9	2,28	2,20	2,20	2,20
4	154	PLUSMINUSHEIM											0,00	0,00	0,00	0,00
5	147	SIEGEHEIM		0,9									0,23	0,40	0,60	0,36
6	154	REMISHEIM											0,00	0,00	0,00	0,00
7	153	NIEDERLAGENHEIM									0,1		0,00	0,03	0,00	0,01
8	149	PUNKTEHEIM			0,6								0,15	0,27	0,15	0,19
9	110	TDHEIM	0,4	2,7	3,2	3,2	2	2,9	2,7	2,1	2,6	2	2,38	2,53	2,35	2,38
10	112	POSGAST	2,1	3,7	1,8	2,2	2,8	2,5	1,7	2,1	2,1	1,9	2,45	2,33	2,10	2,29
11	154	PLUSMINUSGAST											0,00	0,00	0,00	0,00
12	139	SIEGEGAST		2,3	0,4	0,3	1,6	1,7	0,7	1,4	1,3	1,1	0,75	1,33	1,35	1,08
13	154	REMISGAST											0,00	0,00	0,00	0,00
14	149	NIEDERLAGENGAST	0,5	0,2			0,8	0,1		0,2	0,1		0,18	0,30	0,15	0,19
15	141	PUNKTEGAST		1,9			1	1,5	0,3	0,9	1	1	0,48	0,93	0,95	0,76
16	108	TDGAST	0,8	3,6	1,8	3,2	2,6	2,7	2,4	2	2,7	2,2	2,35	2,57	2,35	2,40
17	154	MONAT											0,00	0,00	0,00	0,00
18	154	JAHRESZEIT											0,00	0,00	0,00	0,00
19	55	TABDIFF	3,4	4,7	3,4	3,6	4	3,9	3,2	3,5	3,5	2,7	3,78	3,70	3,50	3,59
20	36	TABTENDENZ	3,5	5,8	3,6	3,7	4,1	4,3	3,5	3,4	4	3,3	4,15	3,97	3,70	3,92
21	154	TABDIFFFORM											0,00	0,00	0,00	0,00
22	154	TABDIFFFORMVERG											0,00	0,00	0,00	0,00
23	66	SIEGDIFF	2,3	4,1	3,3	2,7	3	4,1	3,2	3,5	3,3	2,7	3,10	3,43	3,40	3,22
24	154	REMISDIFF											0,00	0,00	0,00	0,00
25	101	NIEDDIFF	2,5	3,8	2,2	2,5	2,5	2,7	2,6	2,1	3	2,7	2,75	2,60	2,55	2,66
26	40	PUNKTEDIFF	3	4,8	3,6	3,7	3,5	4,5	3,7	3,7	4,4	3,8	3,78	3,90	4,05	3,87
27	22	TDDIFF	5,3	5,5	4,3	3,8	5,2	4,6	4,4	4,8	4,5	4,4	4,73	4,73	4,65	4,68
28	64	TORDIFFFORM	1,3	4,3	3,9	3,6	2,8	4,1	3,3	3	3,6	2,8	3,28	3,40	3,30	3,27
29	136	TFHEIM	0,2	1,3	1,8	1,4	1,2	1,2	1,5	1,1	1,2	1,1	1,18	1,30	1,15	1,20
30	148	TFGAST	0,1		0,7	0,8	0,1		0,1	0,7	0,8		0,20	0,30	0,40	0,33
31	82	HEIMSTARK	2,3	2,9	3,4	3,5	2,4	2,9	3,6	2,5	3,2	2,8	3,03	2,97	2,85	2,95
32	152	GASTSTARK								0,3	0,1		0,00	0,10	0,05	0,04
33	151	SIEGEHEIM		0,5	0,5		0,6						0,25	0,20	0,00	0,16
34	137	VERSUS1HEIM	1	1,6	0,6	1,3	2,1	0,3	1,4	1,1	1	1,2	1,13	1,27	1,05	1,16
35	122	VERSUS1GAST	2,2	2,4	1,8	2,5	2	1,9	2,1	1,6	2,1	1,8	2,23	2,00	1,85	2,04
36	80	VERSUS1	1,8	3,8	3,3	3,6	3	3,1	3,3	2,4	3,1	2,6	3,13	3,13	2,75	3,00
37	143	VERSUS2HEIM	0,4	0,2	0,2	0,9	0,3	1	1	0,8	1	0,20	0,73	0,90	0,58	
38	119	VERSUS2GAST	2,2	2,5	2	2,6	2	2,2	2,1	1,7	2,1	1,8	2,33	2,10	1,90	2,12
39	87	VERSUS2	1,7	3,7	3,4	3,2	2,8	3,1	3,1	2,5	2,9	2,4	3,00	3,00	2,70	2,88
40	144	VERSUS3HEIM	0,2		0,2	0,7		1	0,8	0,6	1,1		0,10	0,57	0,70	0,46
41	121	VERSUS3GAST	2	2,6	1,9	2,2	2,1	2,2	1,9	1,7	2,1	1,8	2,18	2,07	1,90	2,05
42	96	VERSUS3	2,3	3,2	3	2,7	2,6	3,1	2,7	2,6	2,8	2,5	2,80	2,80	2,70	2,75
43	146	VERSUS4HEIM	0,2			0,8		0,4	0,8	0,8	1,1		0,05	0,40	0,80	0,41
44	125	VERSUS4GAST	2,1	2	1,5	2,3	2,1	2	2,1	1,7	2	1,8	1,98	2,07	1,85	1,96
45	99	VERSUS4	2,9	3,4	2,3	2,7	2,9	2,8	2,5	2,5	2,6	2,5	2,83	2,73	2,55	2,71

46	145	VERSUS5HEIM	0,3	0,2	0,9	0,4	0,9	0,5	1,1	0,13	0,43	0,70	0,43	
47	123	VERSUS5GAST	2	2,5	1,7	2	2,2	1,9	1,7	2,1	1,7	2,05	2,03	1,90
48	100	VERSUS5	2,6	3,3	2,6	2,4	2,6	3	2,6	2,5	2,8	2,4	2,73	2,73
49	142	TABGAST		2		1	1	0,6	0,4	1	0,5	0,50	0,87	0,70
50	93	TABVERGLEICH	2,1	3,6	2,6	3,1	2,6	3	2,8	2,3	3,1	2,5	2,85	2,80
51	131	FORM1HEIMHEIM	0,9	1,1	2	2,3	1,7	1,9	2	1,8	2	1,8	1,58	1,87
52	48	FORM1GASTAUSW	3,8	4,9	3,1	3,2	4,2	4,1	2,8	3,8	3,6	3,5	3,75	3,70
53	72	FORM1HEIM	2,6	3,1	3,8	3,6	2,7	3,2	3,5	2,8	3,1	2,7	3,28	3,13
54	60	FORM1GAST	3,4	4,8	3,2	3,1	4,1	3,6	2,8	3,3	3,2	3,1	3,63	3,50
55	4	DIFFFORM1	6,5	7,3	6,4	6,1	7,2	7,2	6,4	6,7	6,6	6,4	6,58	6,93
56	118	FORM2HEIMHEIM	1,6	1,8	2,5	2,7	2	2,3	2,4	2,2	2,2	2,1	2,15	2,23
57	45	FORM2GASTAUSW	3,4	4,7	3,6	3,5	4	4,3	3	3,9	3,6	3,5	3,80	3,77
58	75	FORM2HEIM	2,6	3	3,9	3,7	2,4	3,1	3,4	2,7	3,1	2,7	3,30	2,97
59	52	FORM2GAST	3,4	5,6	3,4	3,1	4,2	3,8	3	3,2	3,3	3,1	3,88	3,67
60	7	DIFFFORM2	5,7	7,1	6,1	5,8	7	7,1	6,3	6,6	7	6,4	6,18	6,80
61	134	FORM3HEIMHEIM	0,9	0,5	0,2	1,7	1,1	1,6	1,4	1,5	1,7	1,6	0,83	1,37
62	49	FORM3GASTAUSW	3,2	4,6	3,6	3,4	3,8	4,6	2,8	3,7	3,9	3,2	3,70	3,73
63	106	FORM3HEIM	2,3	2,5	2,8	3,1	2	2,5	2,7	2,1	2,4	2,1	2,68	2,40
64	63	FORM3GAST	2,6	4,5	3,3	3,7	3,4	3,7	2,9	3,1	3,3	2,8	3,53	3,33
65	10	DIFFFORM3	4,4	7,9	6,2	6,9	6,7	7,5	5,9	6,3	6,6	5,7	6,35	6,70
66	140	FORM4HEIMHEIM	0,4	0,7		1,7	1,2	1,3	1	1,4	1,5	1,4	0,70	1,17
67	29	FORM4GASTAUSW	3,4	4,7	4,3	4,4	4,3	4,8	3,5	4,5	4,1	3,9	4,20	4,20
68	102	FORM4HEIM	2,8	2,6	2,7	3,1	2,4	2,5	2,8	2,4	2,5	2,3	2,80	2,57
69	46	FORM4GAST	3	5,1	3,8	3,7	3,4	4,6	3,2	3,6	4,1	3	3,90	3,73
70	8	DIFFFORM4	5,3	8	6,3	6	6,8	7,4	5,6	6,4	6,7	5,9	6,40	6,60
71	124	FORM5HEIMHEIM	1,3	1,2	2,3	2,5	1,9	2,2	2,2	2	2,1	2	1,83	2,10
72	39	FORM5GASTAUSW	3,5	5	3,8	3,5	4,2	4,5	3	4	3,7	3,5	3,95	3,90
73	68	FORM5HEIM	2,6	3,2	4	3,6	2,6	3,3	3,5	2,9	3,2	2,8	3,35	3,13
74	44	FORM5GAST	3,7	5,4	3,8	3,2	4,4	3,9	3,2	3,4	3,5	3,2	4,03	3,83
75	11	DIFFFORM5	5,7	7,7	6,1	5,1	7,2	6,4	6,2	6,3	6,3	6,1	6,15	6,60
76	154	VIELETORE										0,00	0,00	0,00
77	70	AVG1PLATZHEIM	2,9	2,8	3,7	3,8	2,9	2,9	3,7	2,8	3,1	2,9	3,30	3,17
78	30	AVG1PLATZGAST	3,3	5,8	4,7	3,8	4	4,9	3,8	3,7	4,1	3,4	4,40	4,23
79	3	DIFFAVG1	5,4	8	8	5,7	5,9	7,4	7,2	6	7,1	6,2	6,78	6,83
80	83	AVG2PLATZHEIM	2,7	2,7	2,7	2,7	2,9	2,7	3,2	3	3,4	3,4	2,70	2,93
81	43	AVG2PLATZGAST	4,4	5,3	3,4	3	4	4,2	3,2	3,5	3,9	3,3	4,03	3,80
82	15	DIFFAVG2	5,4	6,8	4,4	3,9	5,7	6,3	4,5	5,6	5,4	5,2	5,13	5,50
83	81	AVG3PLATZHEIM	2,4	2,7	2,8	3,5	2,9	2,5	3,1	3,3	3,1	3,5	2,85	2,83
84	41	AVG3PLATZGAST	3,4	5,3	4,1	3,5	3,7	4,6	3,3	3,5	3,8	3,2	4,08	3,87
85	16	DIFFAVG3	4,2	6	5,4	5,5	4,7	5,3	5,2	5,2	5,3	5,3	5,28	5,07
86	90	AVG4PLATZHEIM	2,5	2,6	2,9	3,2	2,5	2,6	3,1	2,6	3,3	2,9	2,80	2,73
87	54	AVG4PLATZGAST	2,3	5	4,5	3,8	3,1	4,1	3,9	2,9	3,7	2,7	3,90	3,70
88	20	DIFFAVG4	4	5,2	5	5,1	4,4	4,8	5,1	4,6	5,2	5,4	4,83	4,77
89	91	AVG5PLATZHEIM	2,2	2,6	2,6	3,1	2,4	2,8	3,1	2,9	3,1	3,3	2,63	2,77
90	60	AVG5PLATZGAST	3,5	4,6	3,4	3	3,8	3,5	3,2	3,2	3,3	3,1	3,63	3,50
91	25	DIFFAVG5	4,9	4,9	4,5	4,2	4,7	4,4	4,5	4,3	4,6	4,4	4,63	4,53
92	114	HAEUF.PLAZIERUNGHEIM	1,6	2,4	2,5	2,5	2	2,2	2,8	1,8	2,4	2,1	2,25	2,33

93	105	HAEUF.PLAZIERUNGSGAST	2,2	3,4	2,3	2,4	2,7	2,7	2,2	2,3	2,2	2,2	2,58	2,53	2,25	2,46
94	51	ADDNATIONALTOP9HEIM	3,2	4,6	3,2	2,7	3,6	4	4,1	3,3	4	3,7	3,43	3,90	3,65	3,64
95	57	ADDNATIONALTOP9GAST	3,9	5,3	3,1	2,5	3,8	4,2	2,8	3,4	3,7	3,1	3,70	3,60	3,55	3,58
96	93	ADDNATLSPTOP9H.	2,6	2,8	3,5	3,2	2,4	2,4	3,3	2,2	2,7	2,6	3,03	2,70	2,45	2,77
97	102	ADDNATLSPTOP9G.	2,5	3,2	2,3	2,9	2,6	2,5	2,6	2,3	2,7	2,5	2,73	2,57	2,50	2,61
98	64	ADDNATTOP9POSHEIM	3,1	3,7	3,1	2,4	3,7	3,1	3,4	3,3	3,4	3,5	3,08	3,40	3,35	3,27
99	52	ADDNATTOP9POSGAST	4,2	5,3	3,1	2,4	3,9	4,2	2,8	3,4	3,7	3,1	3,75	3,63	3,55	3,61
100	88	ADDNATLSPTOPPOS9H.	2,6	2,7	3,1	2,8	2,6	2,4	3,5	2,6	3,2	3,1	2,80	2,83	2,90	2,86
101	104	ADDNATLSPTOPPOS9G.	2	3,1	2,1	2,6	2,4	2,5	2,4	2,2	2,8	2,9	2,45	2,43	2,50	2,50
102	68	ADDMITTELNATLSPHEIM	2,8	2,9	3,5	3,2	3,4	2,8	3,5	2,8	3,5	3,3	3,10	3,23	3,15	3,17
103	96	ADDMITTELNATLSPGAST	2,7	3,4	2,4	2,9	3	2,6	2,6	2,5	2,9	2,5	2,85	2,73	2,70	2,75
104	98	ADDLSPKTHEIM	2,5	3,1	2,7	3	2,5	2,3	3,1	2,5	2,6	3,1	2,83	2,63	2,55	2,74
105	59	ADDLSPKTGAST	2,8	4,7	3,5	3,8	3,8	3,7	3,3	3,2	3,4	3,1	3,70	3,60	3,30	3,53
106	116	ADDGESCHTOREHEIM	2	2,4	2,6	2,5	2,1	2,1	2,2	1,9	2,3	2	2,38	2,13	2,10	2,21
107	114	ADDGESCHTOREGAST	2,6	4,1	0,2	2,5	3	2,4	1,7	2	2,1	1,7	2,35	2,37	2,05	2,23
108	113	ADDDURCHGELTOREHEIM	2,3	2,3	2,5	2,4	2,2	2,4	2,4	2	2,3	2	2,38	2,33	2,15	2,28
109	127	ADDDURCHGELTOREGAST	2,1	2,5	2,9	1,9	1,8	2,3	1,7	1,8	1,7	1,7	1,88	2,00	1,75	1,87
110	74	ADDPOSITIONHEIM	2,5	3,5	3,2	3,3	2,7	3	3,4	2,6	3,2	3,4	3,13	3,03	2,90	3,08
111	84	ADDPOSITIONGAST	2,2	4	3,2	2,9	3,2	3,1	2,7	2,8	2,7	2,4	3,08	3,00	2,75	2,92
112	131	ADDSTADIONGROESSEHEIM	1,2	0,2	2,7	2,9	1,8	1	2,7	1,7	1,5	1,8	1,75	1,83	1,60	1,75
113	120	ADDSTADIONGROESSEGAST	2,5	2,8	2,6	0,9	2,4	2,1	2,1	2	1,9	1,8	2,20	2,20	1,95	2,11
114	135	ADD10ABSTIEGHEIM	0,6	1,2	1,1	1,6	1,1	0,8	1,9	0,9	1,6	1,3	1,13	1,27	1,25	1,21
115	79	ADD10ABSTIEGGAST	3	3,9	2,7	2,8	3,4	2,9	2,8	2,8	2,9	2,8	3,10	3,03	2,85	3,00
116	138	ADDBRAZILHEIM	2	1,9	0,9	0,2	1,7	1,4	1	1,3	0,9	1,20	1,10	1,15	1,13	
117	133	ADDBRAZILGAST	1,7	2,5	2,5	0,4	1,7	2,2	1,6	1,6	1,7	1,3	1,78	1,83	1,65	1,72
118	126	ADDBRAZILNATHEIM	2	2,3	2	1,9	2	1,7	2	1,6	1,7	1,6	2,05	1,90	1,65	1,88
119	130	ADDBRAZILNATGAST	2,3	2,7	1,6	0,1	2,1	2,1	1,4	1,9	1,8	1,7	1,68	1,87	1,85	1,77
120	86	ADDGERMANYNATHEIM	2,7	3,4	3,3	2,2	3	3,1	2,7	3,2	3	2,4	2,90	2,93	3,10	2,90
121	50	ADDGERMANYNATGAST	4,1	5,1	3,2	2,6	4	4,2	3	3,6	3,7	3,3	3,75	3,73	3,65	3,68
122	154	ADDAUSTRIAHEIM											0,00	0,00	0,00	0,00
123	154	ADDAUSTRIAGAST											0,00	0,00	0,00	0,00
124	154	ADDAUSTRIANATHEIM											0,00	0,00	0,00	0,00
125	154	ADDAUSTRIANATGAST											0,00	0,00	0,00	0,00
126	109	ADDEWIGETABELLEHEIM	3	2,3	2	2,4	2,6	1,8	2,5	2,1	2,5	2,7	2,43	2,30	2,30	2,39
127	95	ADDEWIGETABELLEGAST	1,3	3,7	3,3	3,3	2,5	2,8	2,9	2,5	2,9	2,4	2,90	2,73	2,70	2,76
128	107	ADDUEFA1JAHRHEIM	1,8	2,4	2,6	3,2	2	2,5	2,7	2,1	2,5	2,2	2,50	2,40	2,30	2,40
129	38	ADDUEFA1JAHRGAST	3,8	5,1	3,3	3,8	4	4,2	3,4	3,5	4,1	3,6	4,00	3,87	3,80	3,88
130	67	ADDUEFA5JAHRHEIM	2,7	3,2	3,4	3,5	3,3	2,7	3,5	3,2	3,2	3,1	3,20	3,17	3,20	3,18
131	57	ADDUEFA5JAHREGAST	3,6	4,6	4	3,1	3,7	3,8	3,3	3,4	3,3	3	3,83	3,60	3,35	3,58
132	128	ADDPUNKTEHEIMHEIM	1,9	2,3	2,7	1	1,9	2,6	1,6	2,2	1,9	1,73	1,83	1,90	1,81	
133	56	ADDPUNKTEGASTGAST	2,9	4,6	3,8	3,8	3,5	4	3,4	3,3	3,4	3,1	3,78	3,63	3,35	3,58
134	77	ABSTIEG1HEIM	3	3,2	3	3,1	2,9	3	3,2	2,8	3,1	2,9	3,08	3,03	2,95	3,02
135	24	ABSTIEG1GAST	3,9	6	4,6	4,2	4,7	5,2	4	4,4	4,6	4	4,68	4,63	4,50	4,56
136	37	DIFFABSTIEG1	3,4	4,1	4	4,2	3,8	3,9	3,9	3,9	3,8	3,8	3,93	3,87	3,90	3,89
137	78	ABSTIEG2HEIM	3	3,2	3	3	2,9	3	3,2	2,8	3,1	2,9	3,05	3,03	2,95	3,01
138	26	ABSTIEG2GAST	3,9	6	4,6	4,1	4,7	5,2	3,9	4,4	4,5	4	4,65	4,60	4,45	4,53
139	35	DIFFABSTIEG2	3,4	4,3	4,3	4,1	4	4,1	3,8	3,9	3,8	3,7	4,03	3,97	3,85	3,94

140	76	ABSTIEG3HEIM	3,1	3,2	3	3	3	3	3,2	2,9	3,1	3	3,08	3,07	3,00	3,05
141	27	ABSTIEG3GAST	3,6	6	4,6	4,1	4,5	5,2	3,9	4,3	4,5	3,9	4,58	4,53	4,40	4,46
142	31	DIFFABSTIEG3	3,6	4,6	4,4	4,2	4,2	4,3	3,9	4,1	3,9	3,8	4,20	4,13	4,00	4,10
143	73	ABSTIEG4HEIM	3,2	3,3	3	3	3,1	3	3,2	2,9	3,2	3	3,13	3,10	3,05	3,09
144	28	ABSTIEG4GAST	3,4	5,8	4,6	4,1	4,4	5,1	3,9	4,1	4,4	3,8	4,48	4,47	4,25	4,36
145	47	DIFFABSTIEG4	3,3	4,2	3,9	3,8	3,7	4	3,5	3,6	3,6	3,5	3,80	3,73	3,60	3,71
146	84	ABSTIEG5HEIM	3,1	3	2,7	3	2,9	2,8	3	2,8	3	2,9	2,95	2,90	2,90	2,92
147	32	ABSTIEG5GAST	3	5,4	4,2	4,1	4,1	4,7	3,6	3,9	4,2	3,6	4,18	4,13	4,05	4,08
148	62	DIFFABSTIEG5	3	3,8	3,6	3,7	3,3	3,5	3,3	3,5	3,4	3,3	3,53	3,37	3,45	3,44
149	18	DIFFPUNKTEHEIMGAST	3,7	5,8	5,2	5,9	4,6	5,4	4,6	4,5	5,2	4,8	5,15	4,87	4,85	4,97
150	17	DIFFPKTLS	3,3	6	5,3	5,8	5,2	5,1	5,3	4,7	5,7	5	5,10	5,20	5,20	5,14
151	34	DIFFGSTLS	3,9	5,8	3,4	3,5	4,6	4	3,6	3,5	3,9	3,5	4,15	4,07	3,70	3,97
152	33	DIFFKASSTLS	3,7	3,7	4,1	3,5	4,8	4,7	3,5	4,5	4,1	3,9	3,75	4,33	4,30	4,05
153	19	DIFFPOSLS	3,1	5,3	5,2	5,6	5,3	4,7	4,9	5	5,3	5,1	4,80	4,97	5,15	4,95
154	71	DIFFEWIGETABELLE	3,7	3,3	3,2	2,8	3,3	2,9	3,1	3	2,9	3	3,25	3,10	2,95	3,12
155	6	DIFFNAT	5,4	7,9	6	5,6	6,3	7	6,9	6,1	7,6	6,7	6,23	6,73	6,85	6,55
156	21	DIFNAT2	3,7	5,1	5	5,2	4,8	4,7	5,3	4,5	5,1	5,2	4,75	4,93	4,80	4,86
157	8	DIFFNATPOS	7,3	8,1	5,7	4,7	6,8	6,5	5,5	6,6	6,7	6,5	6,45	6,27	6,65	6,44
158	23	DIFNATPOS2	3,6	6,4	4,7	5	4,4	4,8	5	4,1	4,6	4,1	4,93	4,73	4,35	4,67
159	13	DIFFNATPOSMIT	4,7	7,4	5,8	4,7	5,8	6	5,5	5	5,6	5,4	5,65	5,77	5,30	5,59
160	111	DIFFSTADGR	3,4	1,2	3,1	2,3	2,7	2,1	2,7	2,2	2	2	2,50	2,50	2,10	2,37
161	92	DIFF10ABSTIEG	2,9	3,2	2,5	2,6	3,2	2,5	3	2,4	3	2,7	2,80	2,90	2,70	2,80
162	129	DIFFBRAZIL	1,6	2,8	2,8		1,7	2,7	1,5	1,8	1,6	1,2	1,80	1,97	1,70	1,77
163	89	DIFBRAZILN	2,8	3,8	2,6	2	3,4	2,8	2,3	2,8	2,9	2,9	2,80	2,83	2,85	2,83
164	5	DIFFGERMANYN	7,4	8,3	6,6	4,5	6,8	7,5	6	6,3	7,1	6,2	6,70	6,77	6,70	6,67
165	154	DIFFAUSTRIA											0,00	0,00	0,00	0,00
166	154	DIFFAUSTRIAN											0,00	0,00	0,00	0,00
167	14	DIFFUEFA1J	3,8	5,5	6,2	7	4,3	6	6	5	5,7	5,1	5,63	5,43	5,35	5,46
168	12	DIFFUEFA5J	5,7	6	5,5	4,9	6,6	5,7	6,3	5,6	6,5	5,7	5,53	6,20	6,05	5,85
169	42	DIFFHAEUFPLAZ	3,3	4,3	3,8	3,4	4	3,8	3,5	4,3	3,6	4,2	3,70	3,77	3,95	3,82

*Gesamtergebnisse auf Basis der oberen Tests mit Wrappern mit 5 Klassifizierern und InfoGain
(Angabe in Platzierungen, im Kapitel 4.4.7 bezeichnet als Ranking1)*

Ergänzungen zu Legende für die folgenden zwei Tabellen (in Anhang D):

NB – Naive Bayes

SVM – Support Vector Machines

RF – Random Forest

IG – InfoGain

NR	Platz	Name	NB	JRip	J48	SVM	RF	IG	Ges
1	96	HEIM	133	131	3	168	138	2	95,83333333
2	73	GAST	91	94	3	165	125	1	79,83333333
3	119	POSHEIM	143	117	51	101	121	117	108,333333
4	162	PLUSMINUSHEIM	167	153	79	126	154	154	138,833333
5	155	SIEGEHEIM	129	135	125	119	141	147	132,666667
6	148	REMISHEIM	162	135	29	140	157	154	129,5
7	155	NIEDERLAGENHEIM	154	143	114	86	146	153	132,666667
8	144	PUNKTEHEIM	109	131	96	126	153	149	127,333333
9	117	TDHEIM	136	74	149	94	81	110	107,333333
10	93	POSGAST	62	88	72	134	102	112	95
11	152	PLUSMINUSGAST	161	141	45	134	151	154	131
12	101	SIEGEGAST	136	85	40	94	90	139	97,3333333
13	145	REMISGAST	164	143	51	111	144	154	127,833333
14	143	NIEDERLAGENGAST	124	143	61	140	146	149	127,166667
15	128	PUNKTEGAST	142	92	109	75	116	141	112,5
16	60	TDGAST	59	74	51	57	85	108	72,3333333
17	164	MONAT	152	143	101	166	154	154	145
18	166	JAHRESZEIT	164	143	82	164	166	154	145,5
19	65	TABDIFF	64	35	72	147	76	54	74,6666667
20	93	TABTENDENZ	96	135	96	70	137	36	95
21	167	TABDIFFFORM	151	162	82	158	169	154	146
22	168	TABDIFFFORMVERG	155	162	109	158	167	154	150,833333
23	63	SIEGDIFF	77	88	29	101	85	66	74,3333333
24	151	REMISDIFF	139	164	18	152	157	154	130,666667
25	91	NIEDDIFF	88	88	61	140	85	101	93,8333333
26	50	PUNKTEDIFF	70	43	101	86	58	40	66,3333333
27	20	TDDIFF	24	23	35	30	35	21	28
28	146	TORDIFFFORM	145	141	131	152	135	64	128
29	53	TFHEIM	168	33	26	29	11	136	67,1666667
30	150	TFGAST	157	131	61	156	125	148	129,666667
31	77	HEIMSTARK	85	78	92	119	31	82	81,1666667
32	158	GASTSTARK	153	158	35	148	161	152	134,5
33	163	SIEGEHEIM	145	158	61	166	167	151	141,333333
34	109	VERSUS1HEIM	124	95	51	86	112	137	100,833333
35	112	VERSUS1GAST	91	95	82	111	108	122	101,5
36	83	VERSUS1	77	68	82	119	90	80	86
37	118	VERSUS2HEIM	113	143	38	70	138	143	107,5
38	105	VERSUS2GAST	68	95	82	111	121	119	99,3333333

39	98	VERSUS2	64	86	114	148	81	86	96,5
40	160	VERSUS3HEIM	133	166	96	126	151	144	136
41	137	VERSUS3GAST	77	113	153	119	121	121	117,333333
42	136	VERSUS3	113	103	163	152	76	96	117,166667
43	148	VERSUS4HEIM	149	168	51	119	144	146	129,5
44	133	VERSUS4GAST	88	117	101	134	128	125	115,5
45	106	VERSUS4	123	95	61	119	102	99	99,8333333
46	165	VERSUS5HEIM	136	153	143	148	146	145	145,166667
47	130	VERSUS5GAST	85	122	101	140	115	123	114,333333
48	129	VERSUS5	96	92	138	161	90	100	112,833333
49	138	TABGAST	116	153	13	161	143	142	121,333333
50	108	TABVERGLEICH	70	117	59	163	102	93	100,666667
51	161	FORM1HEIMHEIM	147	135	145	134	128	131	136,666667
52	42	FORM1GASTAUSW	26	58	96	75	58	47	60
53	64	FORM1HEIM	116	52	72	70	65	72	74,5
54	44	FORM1GAST	33	36	114	75	47	59	60,6666667
55	9	DIFFFORM1	33	8	3	8	11	4	11,1666667
56	135	FORM2HEIMHEIM	139	143	72	101	125	118	116,333333
57	43	FORM2GASTAUSW	17	46	109	111	35	44	60,3333333
58	62	FORM2HEIM	96	63	82	64	63	75	73,8333333
59	55	FORM2GAST	30	58	153	75	42	51	68,1666667
60	5	DIFFFORM2	11	6	10	5	7	7	7,66666667
61	158	FORM3HEIMHEIM	139	130	114	152	138	134	134,5
62	37	FORM3GASTAUSW	28	74	22	86	58	48	52,6666667
63	126	FORM3HEIM	73	110	149	126	97	106	110,166667
64	74	FORM3GAST	38	61	125	140	53	63	80
65	1	DIFFFORM3	9	1	2	5	2	10	4,83333333
66	140	FORM4HEIMHEIM	124	127	131	75	142	140	123,166667
67	22	FORM4GASTAUSW	5	23	72	19	31	29	29,8333333
68	100	FORM4HEIM	104	82	127	94	74	102	97,1666667
69	65	FORM4GAST	38	61	138	126	40	45	74,6666667
70	6	DIFFFORM4	11	3	18	3	4	8	7,83333333
71	153	FORM5HEIMHEIM	128	143	123	134	135	124	131,166667
72	31	FORM5GASTAUSW	20	40	51	64	49	39	43,8333333
73	70	FORM5HEIM	61	65	131	86	51	68	77
74	49	FORM5GAST	26	39	159	94	31	43	65,3333333
75	12	DIFFFORM5	11	9	13	15	17	11	12,6666667
76	169	VIELETORE	169	153	169	169	157	154	161,833333
77	115	AVG1PLATZHEIM	129	78	163	126	76	70	107
78	19	AVG1PLATZGAST	15	27	45	22	28	30	27,8333333
79	1	DIFFAVG1	1	2	18	1	4	3	4,83333333
80	90	AVG2PLATZHEIM	84	82	168	50	90	83	92,8333333
81	71	AVG2PLATZGAST	47	70	166	75	63	43	77,3333333
82	28	DIFFAVG2	8	9	167	18	22	15	39,8333333
83	65	AVG3PLATZHEIM	66	40	145	49	67	81	74,6666667
84	40	AVG3PLATZGAST	30	37	157	37	47	41	58,1666667
85	18	DIFFAVG3	38	11	72	9	4	15	24,8333333

86	114	AVG4PLATZHEIM	91	122	153	57	102	89	102,333333
87	51	AVG4PLATZGAST	51	58	123	64	51	53	66,6666667
88	26	DIFFAVG4	49	18	114	11	15	19	37,6666667
89	58	AVG5PLATZHEIM	51	70	82	38	90	90	70,1666667
90	56	AVG5PLATZGAST	46	78	138	33	58	59	68,6666667
91	14	DIFFAVG5	7	15	29	16	22	24	18,8333333
92	121	HAEUF.PLAZIERUNGHEIM	116	115	92	101	116	114	109
93	81	HAEUF.PLAZIERUNGGAST	104	117	29	50	108	105	85,5
94	68	ADDNATIONALTOP9HEIM	109	110	59	50	81	50	76,5
95	57	ADDNATIONALTOP9GAST	38	53	149	86	35	56	69,5
96	54	ADDNATLSPTOP9H.	73	37	131	25	49	93	68
97	104	ADDNATLSPTOP9G.	59	99	162	70	102	102	99
98	78	ADDNATTOP9POSHEIM	104	110	51	75	90	64	82,3333333
99	39	ADDNATTOP9POSGAST	44	50	92	61	35	51	55,5
100	89	ADDNATLSPTOPPOS9H.	104	74	114	101	76	87	92,6666667
101	97	ADDNATLSPTOPPOS9G.	96	103	82	75	116	104	96
102	87	ADDMITTELNATLSPHEIM	88	68	127	101	85	68	89,5
103	103	ADMITTELNATLSPGAST	70	103	165	57	97	96	98
104	82	ADDLSPKTHEIM	77	70	145	50	74	98	85,6666667
105	51	ADDLSPKTGAST	53	53	82	101	53	58	66,6666667
106	76	ADDGESCHTOREHEIM	68	48	149	33	71	116	80,8333333
107	88	ADDGESCHTOREGAST	73	115	40	94	112	114	91,3333333
108	123	ADDDURCHGELTOREHEIM	124	122	127	44	128	113	109,666667
109	123	ADDDURCHGELTOREGAST	116	103	153	38	121	127	109,666667
110	47	ADDPOSITIONHEIM	53	46	127	40	42	74	63,6666667
111	92	ADDPOSITIONGAST	77	113	131	64	97	83	94,1666667
112	142	ADDSTADIONGROESSEHEIM	150	135	157	46	128	131	124,5
113	131	ADDSTADIONGROESSEGAST	129	122	159	46	112	120	114,666667
114	134	ADD10ABSTIEGHEIM	147	158	61	50	146	135	116,166667
115	69	ADD10ABSTIEGGAST	77	103	109	25	67	79	76,6666667
116	125	ADDBRAZILHEIM	143	157	40	36	146	138	110
117	102	ADDBRAZILGAST	116	122	22	64	128	133	97,5
118	107	ADDBRAZILNATHEIM	135	131	26	57	128	126	100,5
119	115	ADDBRAZILNATGAST	109	127	22	126	128	130	107
120	41	ADDGERMANYNATHEIM	57	63	45	31	71	85	58,6666667
121	38	ADDGERMANYNATGAST	10	49	72	101	42	49	53,8333333
122	139	ADDAUSTRIAHEIM	159	158	35	70	161	154	122,833333
123	157	ADDAUSTRIAGAST	156	168	114	46	163	154	133,5
124	132	ADDAUSTRIANATHEIM	164	143	10	64	157	154	115,333333
125	127	ADDAUSTRIANATGAST	163	143	15	40	154	154	111,5
126	61	ADDEWIGETABELLEHEIM	109	67	61	28	65	109	73,1666667
127	31	ADDEWIGETABELLEGAST	33	18	51	44	22	95	43,8333333
128	72	ADDEUEFA1JAHRHEIM	91	100	9	61	107	107	79,1666667
129	30	ADDEUEFA1JAHRGAST	22	31	45	94	29	38	43,1666667
130	74	ADDEUEFA5JAHREHEIM	102	82	101	61	67	67	80
131	59	ADDEUEFA5JAHREGAST	44	53	138	86	53	56	71,6666667
132	85	ADDPUNKTEHEIMHEIM	77	100	61	50	108	128	87,3333333

133	46	ADDPUNKTEGASTGAST	43	53	61	101	58	55	61,8333333
134	110	ABSTIEG1HEIM	104	103	131	111	81	77	101,166667
135	16	ABSTIEG1GAST	38	15	45	12	7	23	23,3333333
136	48	DIFFABSTIEG1	57	44	92	119	42	37	65,1666667
137	95	ABSTIEG2HEIM	96	88	145	75	90	78	95,3333333
138	23	ABSTIEG2GAST	24	15	101	13	7	26	31
139	45	DIFFABSTIEG2	47	40	131	75	42	35	61,6666667
140	113	ABSTIEG3HEIM	96	81	143	140	76	76	102
141	24	ABSTIEG3GAST	22	18	79	22	21	27	31,5
142	35	DIFFABSTIEG3	30	27	101	75	35	31	49,8333333
143	110	ABSTIEG4HEIM	91	86	114	158	85	73	101,166667
144	25	ABSTIEG4GAST	19	25	79	19	22	28	32
145	79	DIFFABSTIEG4	62	53	138	156	53	46	84,6666667
146	120	ABSTIEG5HEIM	116	100	109	148	97	83	108,833333
147	27	ABSTIEG5GAST	20	29	82	40	29	32	38,6666667
148	86	DIFFABSTIEG5	85	65	114	134	71	62	88,5
149	8	DIFFPUNKTEHEIMGAST	14	13	8	4	2	17	9,66666667
150	21	DIFFPKTLS	49	26	40	25	22	16	29,6666667
151	33	DIFFGSTLS	53	34	26	86	34	34	44,5
152	34	DIFFKASSTLS	53	44	61	35	53	33	46,5
153	17	DIFFPOSLS	28	29	38	19	15	18	24,5
154	80	DIFFEWIGETABELLE	102	70	101	101	67	71	85,3333333
155	10	DIFFNAT	16	7	7	16	18	6	11,6666667
156	15	DIFNAT2	37	14	18	32	18	20	23,1666667
157	3	DIFFNATPOS	6	5	6	2	7	8	5,66666667
158	29	DIFNATPOS2	33	31	96	40	22	22	40,6666667
159	13	DIFFNATPOSMIT	18	11	15	24	11	13	15,3333333
160	141	DIFFSTADGR	113	135	159	111	116	111	124,166667
161	84	DIFF10ABSTIEG	73	103	29	126	97	92	86,6666667
162	122	DIFFBRAZIL	116	127	29	140	116	129	109,5
163	99	DIFBRAZILN	129	117	45	94	108	88	96,8333333
164	4	DIFFGERMANYN	2	22	1	9	1	5	6,66666667
165	154	DIFFAUSTRIA	158	166	40	111	163	154	132
166	147	DIFFAUSTRIAN	159	164	22	111	163	154	128,833333
167	11	DIFFUEFA1J	2	18	10	13	18	14	12,5
168	7	DIFFUEFA5J	4	4	15	7	11	12	8,83333333
169	36	DIFFHAEUFPLAZ	66	50	61	50	40	42	51,5

Gesamtergebnisse auf Basis der oberen Tests mit Wrappern mit 5 Klassifizierern (Angabe in Folds, im Kapitel 4.4.7 bezeichnet als Ranking2)

NR	Platz	Name	NB	JRip	J48	SVM	RF	Ges
1	133	HEIM	2,77	3,63	3,20	2,37	3,90	3,17
2	94	GAST	3,13	4,10	3,20	2,63	4,07	3,43
3	129	POSHEIM	2,50	3,83	2,47	3,23	4,10	3,23
4	165	PLUSMINUSHEIM	1,50	3,50	2,33	3,13	3,57	2,81
5	141	SIEGEHEIM	2,80	3,60	2,07	3,17	3,87	3,10
6	159	REMISHEIM	1,73	3,60	2,63	3,07	3,53	2,91
7	155	NIEDERLAGENHEIM	2,07	3,53	2,13	3,30	3,70	2,95
8	137	PUNKTEHEIM	3,00	3,63	2,23	3,13	3,63	3,13
9	106	TDHEIM	2,70	4,40	1,87	3,27	4,53	3,35
10	81	POSGAST	3,47	4,17	2,37	3,10	4,40	3,50
11	157	PLUSMINUSGAST	1,80	3,57	2,50	3,10	3,67	2,93
12	91	SIEGEGAST	2,70	4,23	2,53	3,27	4,47	3,44
13	160	REMISGAST	1,53	3,53	2,47	3,20	3,73	2,89
14	138	NIEDERLAGENGAST	2,87	3,53	2,40	3,07	3,70	3,11
15	126	PUNKTEGAST	2,53	4,13	2,17	3,33	4,13	3,26
16	66	TDGAST	3,53	4,40	2,47	3,47	4,50	3,67
17	165	MONAT	2,20	3,53	2,20	2,53	3,57	2,81
18	168	JAHRESZEIT	1,53	3,53	2,30	2,67	3,43	2,69
19	63	TABDIFF	3,43	5,17	2,37	3,03	4,57	3,71
20	127	TABTENDENZ	3,10	3,60	2,23	3,37	3,97	3,25
21	164	TABDIFFFORM	2,23	3,40	2,30	2,90	3,37	2,84
22	167	TABDIFFFORMVERG	2,00	3,40	2,17	2,90	3,40	2,77
23	74	SIEGDIFF	3,27	4,17	2,63	3,23	4,50	3,56
24	145	REMISDIFF	2,63	3,37	2,73	2,97	3,53	3,05
25	87	NIEDDIFF	3,17	4,17	2,40	3,07	4,50	3,46
26	62	PUNKTEDIFF	3,33	4,93	2,20	3,30	4,80	3,71
27	26	TDDIFF	4,23	5,70	2,60	4,03	5,03	4,32
28	149	TORDIFFFORM	2,47	3,57	2,00	2,97	4,00	3,00
29	56	TFHEIM	1,27	5,27	2,67	4,07	5,50	3,75
30	151	TFGAST	1,93	3,63	2,40	2,93	4,07	2,99
31	71	HEIMSTARK	3,20	4,33	2,27	3,17	5,13	3,62
32	156	GASTSTARK	2,17	3,43	2,60	3,00	3,50	2,94
33	163	SIEGEHEIM	2,47	3,43	2,40	2,53	3,40	2,85
34	102	VERSUS1HEIM	2,87	4,07	2,47	3,30	4,20	3,38
35	99	VERSUS1GAST	3,13	4,07	2,30	3,20	4,27	3,39
36	76	VERSUS1	3,27	4,47	2,30	3,17	4,47	3,53
37	123	VERSUS2HEIM	2,97	3,53	2,57	3,37	3,90	3,27
38	95	VERSUS2GAST	3,37	4,07	2,30	3,20	4,10	3,41
39	88	VERSUS2	3,43	4,20	2,13	3,00	4,53	3,46
40	146	VERSUS3HEIM	2,77	3,33	2,23	3,13	3,67	3,03
41	128	VERSUS3GAST	3,27	3,90	1,80	3,17	4,10	3,25
42	131	VERSUS3	2,97	3,97	1,53	2,97	4,57	3,20
43	149	VERSUS4HEIM	2,33	3,30	2,47	3,17	3,73	3,00

44	123	VERSUS4GAST	3,17	3,83	2,20	3,10	4,03	3,27
45	100	VERSUS4	2,90	4,07	2,40	3,17	4,40	3,39
46	153	VERSUS5HEIM	2,70	3,50	1,93	3,00	3,70	2,97
47	119	VERSUS5GAST	3,20	3,80	2,20	3,07	4,17	3,29
48	117	VERSUS5	3,10	4,13	1,97	2,83	4,47	3,30
49	134	TABGAST	2,93	3,50	2,83	2,83	3,77	3,17
50	106	TABVERGLEICH	3,33	3,83	2,43	2,77	4,40	3,35
51	148	FORM1HEIMHEIM	2,40	3,60	1,90	3,10	4,03	3,01
52	46	FORM1GASTAUSW	4,17	4,63	2,23	3,33	4,80	3,83
53	69	FORM1HEIM	2,93	4,73	2,37	3,37	4,73	3,63
54	39	FORM1GAST	4,07	5,13	2,13	3,33	4,93	3,92
55	10	DIFFFORM1	4,07	6,13	3,20	4,93	5,50	4,77
56	135	FORM2HEIMHEIM	2,63	3,53	2,37	3,23	4,07	3,17
57	37	FORM2GASTAUSW	4,40	4,87	2,17	3,20	5,03	3,93
58	69	FORM2HEIM	3,10	4,57	2,30	3,40	4,77	3,63
59	53	FORM2GAST	4,10	4,63	1,80	3,33	4,97	3,77
60	7	DIFFFORM2	4,57	6,30	2,87	5,20	5,53	4,89
61	143	FORM3HEIMHEIM	2,63	3,73	2,13	2,97	3,90	3,07
62	43	FORM3GASTAUSW	4,13	4,40	2,70	3,30	4,80	3,87
63	110	FORM3HEIM	3,30	3,93	1,87	3,13	4,43	3,33
64	63	FORM3GAST	4,00	4,60	2,07	3,07	4,83	3,71
65	4	DIFFFORM3	4,80	7,13	3,53	5,20	5,60	5,25
66	136	FORM4HEIMHEIM	2,87	3,77	2,00	3,33	3,83	3,16
67	16	FORM4GASTAUSW	5,07	5,70	2,37	4,37	5,13	4,53
68	92	FORM4HEIM	3,03	4,27	2,03	3,27	4,60	3,44
69	58	FORM4GAST	4,00	4,60	1,97	3,13	5,00	3,74
70	5	DIFFFORM4	4,57	6,73	2,73	5,50	5,57	5,02
71	138	FORM5HEIMHEIM	2,83	3,53	2,10	3,10	4,00	3,11
72	34	FORM5GASTAUSW	4,30	4,97	2,47	3,40	4,90	4,01
73	68	FORM5HEIM	3,50	4,53	2,00	3,30	4,87	3,64
74	44	FORM5GAST	4,17	5,03	1,67	3,27	5,13	3,85
75	12	DIFFFORM5	4,57	6,07	2,83	4,57	5,43	4,69
76	169	VIELETORE	0,20	3,50	0,00	0,00	3,53	1,45
77	122	AVG1PLATZHEIM	2,80	4,33	1,53	3,13	4,57	3,27
78	24	AVG1PLATZGAST	4,50	5,43	2,50	4,23	5,20	4,37
79	2	DIFFAVG1	5,57	6,77	2,73	6,60	5,57	5,45
80	121	AVG2PLATZHEIM	3,23	4,27	0,93	3,50	4,47	3,28
81	77	AVG2PLATZGAST	3,83	4,43	1,27	3,33	4,77	3,53
82	25	DIFFAVG2	4,83	6,07	1,20	4,40	5,27	4,35
83	65	AVG3PLATZHEIM	3,40	4,97	1,90	3,53	4,67	3,69
84	37	AVG3PLATZGAST	4,10	5,10	1,77	3,77	4,93	3,93
85	15	DIFFAVG3	4,00	6,03	2,37	4,90	5,57	4,57
86	113	AVG4PLATZHEIM	3,13	3,80	1,80	3,47	4,40	3,32
87	56	AVG4PLATZGAST	3,77	4,63	2,10	3,40	4,87	3,75
88	19	DIFFAVG4	3,80	5,80	2,13	4,87	5,47	4,41
89	60	AVG5PLATZHEIM	3,77	4,43	2,30	3,67	4,47	3,73
90	52	AVG5PLATZGAST	3,87	4,33	1,97	3,90	4,80	3,77

91	13	DIFFAVG5	4,97	5,87	2,63	4,50	5,27	4,65
92	119	HAEUF.PLAZIERUNGHEIM	2,93	3,87	2,27	3,23	4,13	3,29
93	89	HAEUF.PLAZIERUNGGAST	3,03	3,83	2,63	3,50	4,27	3,45
94	84	ADDNATIONALTOP9HEIM	3,00	3,93	2,43	3,50	4,53	3,48
95	51	ADDNATIONALTOP9GAST	4,00	4,70	1,87	3,30	5,03	3,78
96	40	ADDNATLSPTOP9H.	3,30	5,10	2,00	4,17	4,90	3,89
97	100	ADDNATLSPTOP9G.	3,53	4,03	1,60	3,37	4,40	3,39
98	90	ADDNATTOP9POSHEIM	3,03	3,93	2,47	3,33	4,47	3,45
99	41	ADDNATTOP9POSGAST	3,90	4,77	2,27	3,43	5,03	3,88
100	86	ADDNATLSPTOPPOS9H.	3,03	4,40	2,13	3,23	4,57	3,47
101	105	ADDNATLSPTOPPOS9G.	3,10	3,97	2,30	3,33	4,13	3,37
102	84	ADDMITTELNATLSPHEIM	3,17	4,47	2,03	3,23	4,50	3,48
103	109	ADDMITTELNATLSPGAST	3,33	3,97	1,47	3,47	4,43	3,33
104	75	ADDLSPKTHEIM	3,27	4,43	1,90	3,50	4,60	3,54
105	54	ADDLSPKTGAST	3,73	4,70	2,30	3,23	4,83	3,76
106	61	ADDGESCHTOREHEIM	3,37	4,83	1,87	3,90	4,63	3,72
107	93	ADDGESCHTOREGAST	3,30	3,87	2,53	3,27	4,20	3,43
108	123	ADDDURCHGELTOREHEIM	2,87	3,80	2,03	3,60	4,03	3,27
109	118	ADDDURCHGELTOREGAST	2,93	3,97	1,80	3,67	4,10	3,29
110	45	ADDPOSITIONHEIM	3,73	4,87	2,03	3,63	4,97	3,85
111	96	ADDPOSITIONGAST	3,27	3,90	2,00	3,40	4,43	3,40
112	144	ADDSTADIONGROESSEHEIM	2,30	3,60	1,77	3,57	4,03	3,05
113	130	ADDSTADIONGROESSEGAST	2,80	3,80	1,67	3,57	4,20	3,21
114	142	ADD10ABSTIEGHEIM	2,40	3,43	2,40	3,50	3,70	3,09
115	67	ADD10ABSTIEGGAST	3,27	3,97	2,17	4,17	4,67	3,65
116	131	ADDBRAZILHEIM	2,50	3,47	2,53	3,80	3,70	3,20
117	103	ADDBRAZILGAST	2,93	3,80	2,70	3,40	4,03	3,37
118	114	ADDBRAZILNATHEIM	2,73	3,63	2,67	3,47	4,03	3,31
119	112	ADDBRAZILNATGAST	3,00	3,77	2,70	3,13	4,03	3,33
120	42	ADDGERMANYNATHEIM	3,70	4,57	2,50	3,97	4,63	3,87
121	32	ADDGERMANYNATGAST	4,73	4,80	2,37	3,23	4,97	4,02
122	154	ADDAUSTRIAHEIM	1,83	3,43	2,60	3,37	3,50	2,95
123	161	ADDAUSTRIAGAST	1,97	3,30	2,13	3,57	3,47	2,89
124	152	ADDAUSTRIANATHEIM	1,53	3,53	2,87	3,40	3,53	2,97
125	146	ADDAUSTRIANATGAST	1,60	3,53	2,80	3,63	3,57	3,03
126	55	ADDEWIGETABELLEHEIM	3,00	4,50	2,40	4,13	4,73	3,75
127	28	ADDEWIGETABELLEGAST	4,07	5,80	2,47	3,60	5,27	4,24
128	73	ADDUEFA1JAHRHEIM	3,13	4,00	2,90	3,43	4,37	3,57
129	30	ADDUEFA1JAHRGAST	4,27	5,30	2,50	3,27	5,17	4,10
130	77	ADDUEFA5JAHREHEIM	3,07	4,27	2,20	3,43	4,67	3,53
131	58	ADDUEFA5JAHREGAST	3,90	4,70	1,97	3,30	4,83	3,74
132	83	ADDPUNKTEHEIMHEIM	3,27	4,00	2,40	3,50	4,27	3,49
133	48	ADDPUNKTEGASTGAST	3,93	4,70	2,40	3,23	4,80	3,81
134	108	ABSTIEG1HEIM	3,03	3,97	2,00	3,20	4,53	3,35
135	16	ABSTIEG1GAST	4,00	5,87	2,50	4,73	5,53	4,53
136	50	DIFFABSTIEG1	3,70	4,90	2,27	3,17	4,97	3,80
137	97	ABSTIEG2HEIM	3,10	4,17	1,90	3,33	4,47	3,39

138	18	ABSTIEG2GAST	4,23	5,87	2,20	4,60	5,53	4,49
139	47	DIFFABSTIEG2	3,83	4,97	2,00	3,33	4,97	3,82
140	97	ABSTIEG3HEIM	3,10	4,30	1,93	3,07	4,57	3,39
141	22	ABSTIEG3GAST	4,27	5,80	2,33	4,23	5,30	4,39
142	32	DIFFABSTIEG3	4,10	5,43	2,20	3,33	5,03	4,02
143	103	ABSTIEG4HEIM	3,13	4,20	2,13	2,90	4,50	3,37
144	22	ABSTIEG4GAST	4,33	5,63	2,33	4,37	5,27	4,39
145	72	DIFFABSTIEG4	3,47	4,70	1,97	2,93	4,83	3,58
146	114	ABSTIEG5HEIM	2,93	4,00	2,17	3,00	4,43	3,31
147	29	ABSTIEG5GAST	4,30	5,40	2,30	3,63	5,17	4,16
148	79	DIFFABSTIEG5	3,20	4,53	2,13	3,10	4,63	3,52
149	8	DIFFPUNKTEHEIMGAST	4,53	6,00	2,93	5,37	5,60	4,89
150	27	DIFFPKTLS	3,80	5,60	2,53	4,17	5,27	4,27
151	35	DIFFGSTLS	3,73	5,20	2,67	3,30	5,07	3,99
152	36	DIFFKASSTLS	3,73	4,90	2,40	3,83	4,83	3,94
153	21	DIFFPOSLS	4,13	5,40	2,57	4,37	5,47	4,39
154	79	DIFFEWIGETABELLE	3,07	4,43	2,20	3,23	4,67	3,52
155	11	DIFFNAT	4,43	6,23	3,00	4,50	5,37	4,71
156	20	DIFNAT2	4,03	5,90	2,73	3,93	5,37	4,39
157	3	DIFFNATPOS	5,03	6,37	3,17	6,20	5,53	5,26
158	30	DIFNATPOS2	4,07	5,30	2,23	3,63	5,27	4,10
159	14	DIFFNATPOSMIT	4,37	6,03	2,80	4,20	5,50	4,58
160	138	DIFFSTADGR	2,97	3,60	1,67	3,20	4,13	3,11
161	82	DIFF10ABSTIEG	3,30	3,97	2,63	3,13	4,43	3,49
162	114	DIFFBRAZIL	2,93	3,77	2,63	3,07	4,13	3,31
163	110	DIFBRAZILN	2,80	3,83	2,50	3,27	4,27	3,33
164	1	DIFFGERMANYN	5,33	5,77	5,00	4,90	7,77	5,75
165	162	DIFFAUSTRIA	1,87	3,33	2,53	3,20	3,47	2,88
166	158	DIFFAUSTRIAN	1,83	3,37	2,70	3,20	3,47	2,91
167	9	DIFFUEFA1J	5,33	5,80	2,87	4,60	5,37	4,79
168	6	DIFFUEFA5J	5,10	6,40	2,80	5,03	5,50	4,97
169	48	DIFFHAEUFPLAZ	3,40	4,77	2,40	3,50	5,00	3,81

Anhang E: Ergebnisse aus den „Nachher“-Tests im Kapitel 4.7

Legende für Anhang F:

Kom. – Ergebnisse des Komplettdatensatzes aus den Vorhertests

⬆ – Verbesserung gegenüber dem Komplettdatensatz

⬇ – Verschlechterung gegenüber dem Komplettdatensatz

● – Gegenüber dem Komplettdatensatz gleichgeblieben

Gesamtdurchschn. – Gesamtdurchschnittsergebnis aus allen 10 Trainingsset/Testsetkombinationen

Naive Bayes

Datensätze 1-6:

Datensatzname Trainingsset— Testset	Kom.	BestOf	Top40	Top40hg	Top85	IG40	IG85
2002/2004---2005	44,77	49,02 ⬆	48,04 ⬆	48,69 ⬆	48,37 ⬆	48,69 ⬆	49,02 ⬆
2003/2005---2006	44,44	42,81 ⬇	42,81 ⬇	43,46 ⬇	43,14 ⬇	42,81 ⬇	42,48 ⬇
2004/2006---2007	46,73	48,69 ⬆	48,37 ⬆	47,71 ⬆	49,34 ⬆	50,33 ⬆	49,02 ⬆
2005/2007---2008	47,14	47,39 ⬆	45,10 ⬇	45,10 ⬇	45,10 ⬇	46,73 ⬇	46,41 ⬇
2002/2005---2006	43,46	43,46 ●	44,12 ⬆	44,12 ⬆	44,44 ⬆	43,14 ⬇	43,14 ⬇
2003/2006---2007	49,35	49,35 ●	48,37 ⬇	48,37 ⬇	49,35 ●	49,35 ●	48,37 ⬇
2004/2007---2008	47,47	47,71 ⬆	46,08 ⬇	47,06 ⬇	45,42 ⬇	47,06 ⬇	45,42 ⬇
2002/2006---2007	48,69	49,35 ⬆	49,67 ⬆	49,35 ⬆	49,02 ⬆	50,65 ⬆	48,37 ⬇
2003/2007---2008	47,47	46,73 ⬇	46,08 ⬇	46,41 ⬇	45,75 ⬇	45,75 ⬇	45,75 ⬇
2002/2007---2008	47,81	47,39 ⬇	46,08 ⬇	46,41 ⬇	45,75 ⬇	46,73 ⬇	46,41 ⬇
Vergleich	----	5↑3↓2●	4↑6↓	4↑6↓	4↑5↓1●	3↑6↓1●	2↑8↓
Gesamtdurchschn.	46,73	47,19 ⬆	46,47 ⬇	46,67 ⬇	46,57 ⬇	47,12 ⬆	46,44 ⬇

Tabelle E.1a: Vergleich der Datensätze Naive Bayes, 1. Teil

Datensätze 7-11:

Datensatzname Trainingsset— Testset	Kom.	Fold40	Fold85	Diff	FStudio	Kicker
2002/2004---2005	44,77	49,02 ⬆	45,10 ⬆	47,71 ⬆	49,35 ⬆	46,41 ⬆
2003/2005---2006	44,44	43,79 ⬇	43,79 ⬇	43,14 ⬇	41,83 ⬇	42,48 ⬇
2004/2006---2007	46,73	48,37 ⬆	49,02 ⬆	49,02 ⬆	39,87 ⬇	48,04 ⬆
2005/2007---2008	47,14	45,75 ⬇	44,77 ⬇	46,08 ⬇	48,04 ⬆	45,10 ⬇
2002/2005---2006	43,46	44,12 ⬆	44,77 ⬇	43,79 ⬆	45,42 ⬆	42,48 ⬇
2003/2006---2007	49,35	48,04 ⬇	49,67 ⬆	49,35 ●	39,21 ⬇	49,35 ●
2004/2007---2008	47,47	45,42 ⬇	45,10 ⬇	46,73 ⬇	47,06 ⬇	45,75 ⬇
2002/2006---2007	48,69	48,69 ●	49,67 ⬆	48,69 ●	43,79 ⬇	49,35 ⬆
2003/2007---2008	47,47	46,08 ⬇	45,10 ⬇	45,10 ⬇	49,67 ⬆	47,39 ⬇
2002/2007---2008	47,81	46,73 ⬇	45,42 ⬇	46,08 ⬇	49,34 ⬆	47,39 ⬇
Vergleich	----	3↑6↓1●	4↑6↓	3↑5↓2●	5↑5↓	3↑6↓1●
Gesamtdurchschn.	46,73	46,60 ⬇	46,24 ⬇	46,57 ⬇	45,36 ⬇	46,37 ⬇

Tabelle E.1b: Vergleich der Datensätze Naive Bayes, 2. Teil

JRip

Datensätze 1-6:

Datensatzname Trainingsset— Testset	Kom.	BestOf	Top40	Top40hg	Top85	IG40	IG85
2002/2004---2005	46,73	42,48 ↓	44,77 ↓	48,37 ↑	46,73 ●	41,83 ↓	46,41 ↓
2003/2005---2006	47,06	46,08 ↓	42,81 ↓	44,77 ↓	45,10 ↓	47,38 ↑	47,39 ↑
2004/2006---2007	46,08	47,06 ↑	48,37 ↑	49,67 ↑	48,69 ↑	47,71 ↑	51,96 ↑
2005/2007---2008	48,48	45,10 ↓	44,12 ↓	51,31 ↑	48,69 ↑	46,41 ↓	47,06 ↓
2002/2005---2006	42,81	46,73 ↑	46,41 ↑	43,14 ↑	44,44 ↑	47,39 ↑	42,48 ↓
2003/2006---2007	50,00	50,00 ●	46,73 ↓	48,04 ↓	48,04 ↓	48,69 ↓	47,38 ↓
2004/2007---2008	48,82	45,42 ↓	46,08 ↓	46,73 ↓	50,33 ↑	49,67 ↑	51,96 ↑
2002/2006---2007	46,73	48,37 ↑	50,65 ↑	49,35 ↑	49,67 ↑	49,02 ↑	45,75 ↓
2003/2007---2008	47,14	47,06 ↓	49,02 ↑	48,04 ↑	48,04 ↑	51,96 ↑	48,37 ↑
2002/2007---2008	50,84	49,35 ↓	51,96 ↑	50,65 ↓	49,02 ↓	50,65 ↓	47,71 ↓
Vergleich	----	3↑6↓1●	5↑5↓	6↑4↓	6↑3↓1●	6↑4↓	4↑6↓
Gesamtdurchschn.	47,47	46,77 ↓	47,09 ↓	48,01 ↑	47,88 ↑	48,07 ↑	47,65 ↑

Tabelle E.2a: Vergleich der Datensätze JRip, 1. Teil

Datensätze 7-11:

Datensatzname Trainingsset— Testset	Kom.	Fold40	Fold85	Diff	FStudio	Kicker
2002/2004---2005	46,73	43,79 ↓	45,42 ↓	45,75 ↓	45,10 ↓	44,12 ↓
2003/2005---2006	47,06	45,10 ↓	44,44 ↓	43,79 ↓	44,44 ↓	46,41 ↓
2004/2006---2007	46,08	47,06 ↑	50,00 ↑	48,69 ↑	49,35 ↑	47,06 ↑
2005/2007---2008	48,48	50,00 ↑	46,41 ↓	49,35 ↑	50,00 ↑	44,44 ↓
2002/2005---2006	42,81	47,71 ↑	44,12 ↑	45,10 ↑	45,10 ↑	44,44 ↑
2003/2006---2007	50,00	50,33 ↑	51,30 ↑	51,63 ↑	49,02 ↓	49,67 ↓
2004/2007---2008	48,82	46,41 ↓	47,06 ↓	48,37 ↓	50,65 ↑	50,98 ↑
2002/2006---2007	46,73	50,65 ↑	50,00 ↑	46,73 ●	49,35 ↑	49,67 ↑
2003/2007---2008	47,14	44,12 ↓	49,67 ↑	47,71 ↑	49,67 ↑	48,04 ↑
2002/2007---2008	50,84	48,04 ↓	46,41 ↓	49,02 ↓	50,65 ↓	46,08 ↓
Vergleich	----	5↑5↓	5↑5↓	4↑5↓1●	6↑4↓	5↑5↓
Gesamtdurchschn.	47,47	47,32 ↓	47,48 ↑	47,61 ↑	48,33 ↑	47,09 ↓

Tabelle E.2b: Vergleich der Datensätze JRip, 2. Teil

Datensätze 1-6:

Datensatzname Trainingsset— Testset	Kom.	BestOf	Top40	Top40hg	Top85	IG40	IG85
2002/2004---2005	38,89	45,42 ↑	40,85 ↑	45,75 ↑	41,17 ↑	44,77 ↑	40,85 ↑
2003/2005---2006	39,87	44,12 ↑	44,77 ↑	45,42 ↑	40,52 ↑	46,73 ↑	44,77 ↑
2004/2006---2007	38,89	44,44 ↑	37,58 ↓	50,33 ↑	37,58 ↓	48,37 ↑	42,37 ↑
2005/2007---2008	36,36	40,52 ↑	36,60 ↑	40,52 ↑	37,58 ↑	42,16 ↑	37,25 ↑
2002/2005---2006	39,87	42,81 ↑	41,50 ↑	43,46 ↑	31,70 ↓	44,12 ↑	42,16 ↑
2003/2006---2007	43,79	43,79 ●	40,85 ↓	46,08 ↑	43,14 ↓	42,81 ↓	44,44 ↑
2004/2007---2008	38,38	44,44 ↑	42,81 ↑	42,48 ↑	38,89 ↑	40,52 ↑	36,60 ↓
2002/2006---2007	39,22	46,73 ↑	37,91 ↓	44,44 ↑	40,52 ↑	47,71 ↑	45,42 ↑
2003/2007---2008	39,73	40,20 ↑	38,56 ↓	39,87 ↑	39,54 ↓	40,20 ↑	41,18 ↑
2002/2007---2008	35,02	46,73 ↑	34,97 ↓	48,37 ↑	37,91 ↑	45,42 ↑	40,85 ↑
Vergleich	----	9↑1●	5↑5↓	10↑	6↑4↓	9↑1↓	9↑1↓
Gesamtdurchschn.	39,00	43,92 ↑	39,64 ↑	44,67 ↑	38,86 ↓	44,58 ↑	41,59 ↑

Tabelle E.3a: Vergleich der Datensätze J48, 1. Teil

Datensätze 7-11:

Datensatzname Trainingsset— Testset	Kom.	Fold40	Fold85	Diff	FStudio	Kicker
2002/2004---2005	38,89	36,27 ↓	34,64 ↓	38,89 ●	45,75 ↑	39,22 ↑
2003/2005---2006	39,87	40,20 ↑	42,48 ↑	33,99 ↓	33,89 ↓	40,52 ↑
2004/2006---2007	38,89	37,25 ↑	36,27 ↓	40,85 ↑	45,75 ↑	37,91 ↓
2005/2007---2008	36,36	36,27 ↓	38,24 ↑	35,62 ↓	45,10 ↑	46,73 ↑
2002/2005---2006	39,87	36,93 ↓	42,81 ↑	37,91 ↓	42,48 ↑	33,66 ↓
2003/2006---2007	43,79	34,64 ↓	40,52 ↓	38,56 ↓	45,10 ↑	41,18 ↓
2004/2007---2008	38,38	33,66 ↓	36,60 ↓	36,27 ↓	46,73 ↑	37,91 ↓
2002/2006---2007	39,22	39,22 ●	35,62 ↓	35,62 ↓	47,06 ↑	38,24 ↓
2003/2007---2008	39,73	38,89 ↓	39,22 ↓	39,22 ↓	47,06 ↑	48,69 ↑
2002/2007---2008	35,02	36,60 ↑	39,22 ↑	38,89 ↑	46,73 ↑	39,22 ↑
Vergleich	----	3↑6↓1●	4↑6↓	2↑7↓1●	9↑1↓	5↑5↓
Gesamtdurchschn.	39,00	36,99 ↓	38,56 ↓	37,58 ↓	44,57 ↑	40,33 ↑

Tabelle E.3b: Vergleich der Datensätze J48, 2. Teil

SVM

Datensätze 1-6:

Datensatzname Trainingsset— Testset	Kom.	BestOf	Top40	Top40hg	Top85	IG40	IG85
2002/2004---2005	48,69	45,10 ↓	47,39 ↓	46,73 ↓	46,73 ↓	44,77 ↓	46,08 ↓
2003/2005---2006	40,52	47,39 ↑	46,41 ↑	46,41 ↑	43,46 ↑	45,75 ↑	43,79 ↑
2004/2006---2007	44,44	46,08 ↑	50,00 ↑	48,37 ↑	49,35 ↑	47,39 ↑	47,71 ↑
2005/2007---2008	40,40	42,81 ↑	45,42 ↑	44,77 ↑	46,08 ↑	44,12 ↑	39,87 ↓
2002/2005---2006	42,81	49,02 ↑	48,37 ↑	46,41 ↑	44,77 ↑	49,02 ↑	43,14 ↑
2003/2006---2007	44,12	47,71 ↑	50,33 ↑	47,39 ↑	48,37 ↑	47,39 ↑	48,69 ↑
2004/2007---2008	41,08	44,77 ↑	45,75 ↑	45,10 ↑	42,16 ↑	43,79 ↑	42,81 ↑
2002/2006---2007	46,41	48,37 ↑	51,63 ↑	48,04 ↑	50,33 ↑	48,04 ↑	48,04 ↑
2003/2007---2008	46,13	44,77 ↓	45,10 ↓	44,12 ↓	46,08 ↓	44,12 ↓	45,75 ↓
2002/2007---2008	48,15	46,41 ↓	45,75 ↓	46,73 ↓	46,41 ↓	45,75 ↓	46,73 ↓
Vergleich	----	7↑3↓	7↑3↓	7↑3↓	7↑3↓	7↑3↓	6↑4↓
Gesamtdurchschn.	44,28	46,24 ↑	47,61 ↑	46,40 ↑	46,73 ↑	46,01 ↑	45,26 ↑

Tabelle E.4a: Vergleich der Datensätze SVM, 1. Teil

Datensätze 7-11:

Datensatzname Trainingsset— Testset	Kom.	Fold40	Fold85	Diff	FStudio	Kicker
2002/2004---2005	48,69	47,71 ↓	47,06 ↓	48,69 •	45,42 ↓	46,41 ↓
2003/2005---2006	40,52	47,71 ↑	46,41 ↑	47,39 ↑	46,08 ↑	44,77 ↑
2004/2006---2007	44,44	50,98 ↑	50,98 ↑	48,37 ↑	50,00 ↑	46,73 ↑
2005/2007---2008	40,40	45,10 ↑	46,08 ↑	47,06 ↑	44,77 ↑	46,41 ↑
2002/2005---2006	42,81	48,04 ↑	48,04 ↑	45,75 ↑	47,71 ↑	48,37 ↑
2003/2006---2007	44,12	50,33 ↑	51,96 ↑	51,96 ↑	48,37 ↑	47,71 ↑
2004/2007---2008	41,08	45,75 ↑	45,75 ↑	48,04 ↑	43,79 ↑	45,42 ↑
2002/2006---2007	46,41	50,65 ↑	50,33 ↑	51,31 ↑	49,02 ↑	49,67 ↑
2003/2007---2008	46,13	44,44 ↓	45,42 ↓	46,73 ↑	44,77 ↓	45,75 ↓
2002/2007---2008	48,15	45,42 ↓	47,39 ↓	47,06 ↓	46,73 ↓	44,77 ↓
Vergleich	----	7↑3↓	7↑3↓	8↑1↓1•	7↑3↓	7↑3↓
Gesamtdurchschn.	44,28	47,61 ↑	47,94 ↑	48,24 ↑	46,67 ↑	46,60 ↑

Tabelle E.4b: Vergleich der Datensätze SVM, 2. Teil

Random Forest

Datensätze 1-6:

Datensatzname Trainingsset— Testset	Kom.	BestOf	Top40	Top40hg	Top85	IG40	IG85
2002/2004---2005	41,83	44,77 ↑	42,48 ↑	41,50 ↓	42,48 ↑	46,41 ↑	44,12 ↑
2003/2005---2006	40,85	43,13 ↑	40,52 ↓	36,27 ↓	41,83 ↑	40,52 ↓	44,77 ↑
2004/2006---2007	38,89	45,10 ↑	42,81 ↑	42,48 ↑	42,16 ↑	42,48 ↑	42,16 ↑
2005/2007---2008	44,78	42,81 ↓	38,24 ↓	40,52 ↓	38,89 ↓	43,79 ↓	42,81 ↓
2002/2005---2006	41,50	37,91 ↓	41,83 ↑	36,60 ↓	41,17 ↓	38,56 ↓	36,60 ↓
2003/2006---2007	45,10	43,79 ↓	46,07 ↑	43,79 ↓	47,39 ↑	45,75 ↓	45,42 ↑
2004/2007---2008	42,42	45,42 ↑	45,42 ↑	45,42 ↑	43,79 ↑	39,87 ↓	44,77 ↑
2002/2006---2007	45,10	42,81 ↓	45,75 ↑	44,44 ↓	44,12 ↓	45,42 ↑	43,46 ↓
2003/2007---2008	43,10	41,50 ↓	46,08 ↑	39,54 ↓	44,44 ↑	42,81 ↓	45,42 ↑
2002/2007---2008	43,10	43,46 ↑	46,41 ↑	39,87 ↓	42,81 ↓	43,46 ↑	41,83 ↓
Vergleich	----	5↑5↓	8↑2↓	2↑8↓	6↑4↓	4↑6↓	6↑4↓
Gesamtdurchschn.	42,67	43,07 ↑	43,56 ↑	41,04 ↓	42,91 ↑	42,91 ↑	43,14 ↑

Tabelle E.5a: Vergleich der Datensätze Random Forest, 1. Teil

Datensätze 7-11:

Datensatzname Trainingsset— Testset	Kom.	Fold40	Fold85	Diff	FStudio	Kicker
2002/2004---2005	41,83	44,12 ↑	45,10 ↑	44,12 ↑	42,16 ↑	41,50 ↓
2003/2005---2006	40,85	42,16 ↑	45,10 ↑	37,91 ↓	38,56 ↓	39,54 ↓
2004/2006---2007	38,89	46,73 ↑	43,46 ↑	40,20 ↑	40,20 ↑	46,41 ↑
2005/2007---2008	44,78	37,25 ↓	42,48 ↓	44,44 ↓	40,20 ↓	38,89 ↓
2002/2005---2006	41,50	41,17 ↓	43,79 ↑	42,16 ↑	34,31 ↓	38,56 ↓
2003/2006---2007	45,10	46,41 ↑	45,42 ↑	47,39 ↑	43,14 ↓	46,41 ↑
2004/2007---2008	42,42	44,44 ↑	48,37 ↑	45,75 ↑	39,87 ↓	42,81 ↑
2002/2006---2007	45,10	43,79 ↓	44,77 ↓	43,79 ↓	39,22 ↓	39,87 ↓
2003/2007---2008	43,10	47,39 ↑	45,42 ↑	46,41 ↑	43,14 ↑	43,14 ↑
2002/2007---2008	43,10	39,53 ↓	46,41 ↑	38,56 ↓	41,18 ↓	42,48 ↓
Vergleich	----	6↑4↓	8↑2↓	6↑4↓	3↑7↓	4↑6↓
Gesamtdurchschn.	42,67	43,30 ↑	45,03 ↑	43,07 ↑	40,20 ↓	41,96 ↓

Tabelle E.5b: Vergleich der Datensätze Random Forest, 2. Teil

Anhang F Ergebnisse der Tests: Datensätze vs Tippstrategien (Kapitel 5.2)

Legende für Anhang F:

Name – Name des Datensatzes oder Tippstrategie

Punkte B1 – Punkte aus Benchmark1

Punkte LM – Punkte aus leichtem Miss

Punkte B2 – Punkte aus Benchmark2

Platz B1 – Platz im Benchmark1

Platz B2 – Platz im Benchmark2

Vorhersagenpunkte für die Saison 2005

Name	Platz B1	Platz B2	Punkte B1	Punkte LM	Punkte B2
TSHemsieg	52	67	131	96	358
TSTabellenplatz	50	62	133	96	362
TSTabellenplatz2	50	14	133	127	393
TSVersus	62	54	126	119	371
TSForm	56	32	129	124	382
TSForm5	13	19	146	96	388
TSLS	19	28	144	96	384
TS3LS	25	32	142	98	382
TSSiege	36	38	138	104	380
TSNied	62	62	126	110	362
TSTeams	13	19	146	96	388
Random	72	72	89	141	319
1. BestOf_3S_NaiveBayes	2	5	150	98	398
1. BestOf_3S_Jrip	53	68	130	96	356
1. BestOf_3S_J48	33	51	139	97	375
1. BestOf_3S_SVM	36	38	138	104	380
1. BestOf_3S_RandomForest	40	28	137	110	384
2. Top40_3S_NaiveBayes	12	12	147	100	394
2. Top40_3S_Jrip	40	55	137	96	370
2. Top40_3S_J48	65	49	125	127	377
2. Top40_3S_SVM	18	21	145	97	387
2. Top40_3S_RandomForest	53	45	130	118	378
3. Top40hg_3S_NaiveBayes	5	4	149	101	399
3. Top40hg_3S_Jrip	10	2	148	104	400
3. Top40hg_3S_J48	30	37	140	101	381
3. Top40hg_3S_SVM	21	16	143	104	390
3. Top40hg_3S_RandomForest	60	56	127	115	369
4. Top85_3S_NaiveBayes	10	5	148	102	398
4. Top85_3S_Jrip	21	32	143	96	382
4. Top85_3S_J48	62	53	126	120	372
4. Top85_3S_SVM	21	22	143	100	386
4. Top85_3S_RandomForest	53	45	130	118	378
5. Komplett_3S_NaiveBayes	40	17	137	115	389
5. Komplett_3S_Jrip	21	32	143	96	382
5. Komplett_3S_J48	68	65	119	122	360

5. Komplett_3S_SVM	5	9	149	99	397
5. Komplett_3S_RandomForest	58	57	128	111	367
6. InfoGain40_3S_NaiveBayes	5	5	149	100	398
6. InfoGain40_3S_Jrip	58	70	128	96	352
6. InfoGain40_3S_J48	13	22	146	94	386
6. InfoGain40_3S_SVM	40	45	137	104	378
6. InfoGain40_3S_RandomForest	44	38	136	108	380
7. InfoGain85_3S_NaiveBayes	2	2	150	100	400
7. InfoGain85_3S_Jrip	25	38	142	96	380
7. InfoGain85_3S_J48	65	61	125	113	363
7. InfoGain85_3S_SVM	29	26	141	103	385
7. InfoGain85_3S_RandomForest	45	22	135	116	386
8. Fold40_3S_NaiveBayes	2	5	150	98	398
8. Fold40_3S_Jrip	49	60	134	96	364
8. Fold40_3S_J48	70	68	111	134	356
8. Fold40_3S_SVM	13	17	146	97	389
8. Fold40_3S_RandomForest	45	38	135	110	380
9. Fold85_3S_NaiveBayes	5	9	149	99	397
9. Fold85_3S_Jrip	33	52	139	96	374
9. Fold85_3S_J48	71	71	106	123	335
9. Fold85_3S_SVM	19	28	144	96	384
9. Fold85_3S_RandomForest	36	12	138	118	394
10. Differenz_3S_NaiveBayes	13	11	146	103	395
10. Differenz_3S_Jrip	30	49	140	97	377
10. Differenz_3S_J48	68	64	119	123	361
10. Differenz_3S_SVM	5	14	149	95	393
10. Differenz_3S_RandomForest	45	26	135	115	385
11. FStudio_3S_NaiveBayes	1	1	151	100	402
11. FStudio_3S_Jrip	36	44	138	103	379
11. FStudio_3S_J48	30	31	140	103	383
11. FStudio_3S_SVM	33	38	139	102	380
11. FStudio_3S_RandomForest	56	57	129	109	367
12. Kicker_3S_NaiveBayes	25	22	142	102	386
12. Kicker_3S_Jrip	45	59	135	96	366
12. Kicker_3S_J48	67	66	120	119	359
12. Kicker_3S_SVM	25	32	142	98	382
12. Kicker_3S_RandomForest	60	45	127	124	378

Vorhersagenpunkte für die Saison 2006

Name	Platz B1	Platz B2	Punkte B1	Punkte LM	Punkte B2
TSHirmsieg	58	115	134	79	347
TSTabellenplatz	69	120	132	79	343
TSTabellenplatz2	123	100	112	130	354
TSVersus	127	114	109	130	348
TSForm	99	66	126	111	363
TSForm5	45	66	136	91	363
TSLS	45	107	136	79	351
TS3LS	24	66	141	81	363
TSSiege	77	89	131	97	359
TSNied	119	120	115	113	343
TSTeams	24	77	141	79	361
Random	132	132	96	126	318
1. BestOf_3S_NaiveBayes	77	84	131	98	360
1. BestOf_3S_Jrip	24	77	141	79	361
1. BestOf_3S_J48	52	33	135	100	370
1. BestOf_3S_SVM	10	9	145	88	378
1. BestOf_3S_RandomForest	69	60	132	101	365
1. BestOf_4S_NaiveBayes	64	89	133	93	359
1. BestOf_4S_Jrip	16	60	143	79	365
1. BestOf_4S_J48	77	66	131	101	363
1. BestOf_4S_SVM	1	2	150	83	383
1. BestOf_4S_RandomForest	116	103	116	120	352
2. Top40_3S_NaiveBayes	77	64	131	102	364
2. Top40_3S_Jrip	77	125	131	79	341
2. Top40_3S_J48	38	1	137	114	388
2. Top40_3S_SVM	18	33	142	86	370
2. Top40_3S_RandomForest	103	60	124	117	365
2. Top40_4S_NaiveBayes	52	46	135	98	368
2. Top40_4S_Jrip	18	33	142	86	370
2. Top40_4S_J48	97	20	127	119	373
2. Top40_4S_SVM	3	14	148	79	375
2. Top40_4S_RandomForest	94	84	128	104	360
3. Top40hg_3S_NaiveBayes	64	56	133	100	366
3. Top40hg_3S_Jrip	38	101	137	79	353
3. Top40hg_3S_J48	30	16	139	96	374
3. Top40hg_3S_SVM	18	11	142	92	376
3. Top40hg_3S_RandomForest	126	129	111	110	332
3. Top40hg_4S_NaiveBayes	52	50	135	97	367
3. Top40hg_4S_Jrip	69	120	132	79	343
3. Top40hg_4S_J48	64	56	133	100	366
3. Top40hg_4S_SVM	18	41	142	85	369
3. Top40hg_4S_RandomForest	123	124	112	118	342
4. Top85_3S_NaiveBayes	69	74	132	98	362
4. Top85_3S_Jrip	33	97	138	79	355
4. Top85_3S_J48	103	20	124	125	373
4. Top85_3S_SVM	64	46	133	102	368

4. Top85_3S_RandomForest	94	84	128	104	360
4. Top85_4S_NaiveBayes	45	56	136	94	366
4. Top85_4S_Jrip	45	64	136	92	364
4. Top85_4S_J48	131	131	97	129	323
4. Top85_4S_SVM	38	16	137	100	374
4. Top85_4S_RandomForest	99	56	126	114	366
5. Komplett_3S_NaiveBayes	45	33	136	98	370
5. Komplett_3S_Jrip	15	50	144	79	367
5. Komplett_3S_J48	109	103	122	108	352
5. Komplett_3S_SVM	103	27	124	124	372
5. Komplett_3S_RandomForest	102	33	125	120	370
5. Komplett_4S_NaiveBayes	64	77	133	95	361
5. Komplett_4S_Jrip	77	125	131	79	341
5. Komplett_4S_J48	109	94	122	114	358
5. Komplett_4S_SVM	77	27	131	110	372
5. Komplett_4S_RandomForest	97	20	127	119	373
6. InfoGain40_3S_NaiveBayes	77	84	131	98	360
6. InfoGain40_3S_Jrip	10	41	145	79	369
6. InfoGain40_3S_J48	16	6	143	94	380
6. InfoGain40_3S_SVM	28	20	140	93	373
6. InfoGain40_3S_RandomForest	103	46	124	120	368
6. InfoGain40_4S_NaiveBayes	69	77	132	97	361
6. InfoGain40_4S_Jrip	10	16	145	84	374
6. InfoGain40_4S_J48	52	30	135	101	371
6. InfoGain40_4S_SVM	1	6	150	80	380
6. InfoGain40_4S_RandomForest	113	101	118	117	353
7. InfoGain85_3S_NaiveBayes	86	77	130	101	361
7. InfoGain85_3S_Jrip	10	41	145	79	369
7. InfoGain85_3S_J48	38	27	137	98	372
7. InfoGain85_3S_SVM	58	11	134	108	376
7. InfoGain85_3S_RandomForest	38	2	137	109	383
7. InfoGain85_4S_NaiveBayes	69	66	132	99	363
7. InfoGain85_4S_Jrip	86	128	130	79	339
7. InfoGain85_4S_J48	91	66	129	105	363
7. InfoGain85_4S_SVM	69	74	132	98	362
7. InfoGain85_4S_RandomForest	123	117	112	122	346
8. Fold40_3S_NaiveBayes	58	50	134	99	367
8. Fold40_3S_Jrip	33	89	138	83	359
8. Fold40_3S_J48	108	11	123	130	376
8. Fold40_3S_SVM	7	16	146	82	374
8. Fold40_3S_RandomForest	91	33	129	112	370
8. Fold40_4S_NaiveBayes	52	50	135	97	367
8. Fold40_4S_Jrip	7	8	146	87	379
8. Fold40_4S_J48	122	112	113	124	350
8. Fold40_4S_SVM	5	20	147	79	373
8. Fold40_4S_RandomForest	99	41	126	117	369
9. Fold85_3S_NaiveBayes	30	77	139	83	361

9. Fold85_3S_Jrip	33	97	138	79	355
9. Fold85_3S_J48	119	107	115	121	351
9. Fold85_3S_SVM	18	30	142	87	371
9. Fold85_3S_RandomForest	119	127	115	110	340
9. Fold85_4S_NaiveBayes	38	50	137	93	367
9. Fold85_4S_Jrip	52	113	135	79	349
9. Fold85_4S_J48	77	66	131	101	363
9. Fold85_4S_SVM	5	10	147	83	377
9. Fold85_4S_RandomForest	58	50	134	99	367
10. Differenz_3S_NaiveBayes	69	94	132	94	358
10. Differenz_3S_Jrip	58	115	134	79	347
10. Differenz_3S_J48	129	103	104	144	352
10. Differenz_3S_SVM	10	20	145	83	373
10. Differenz_3S_RandomForest	116	117	116	114	346
10. Differenz_4S_NaiveBayes	58	89	134	91	359
10. Differenz_4S_Jrip	33	60	138	89	365
10. Differenz_4S_J48	116	119	116	113	345
10. Differenz_4S_SVM	28	89	140	79	359
10. Differenz_4S_RandomForest	91	46	129	110	368
11. FStudio_3S_NaiveBayes	94	41	128	113	369
11. FStudio_3S_Jrip	45	107	136	79	351
11. FStudio_3S_J48	112	103	119	114	352
11. FStudio_3S_SVM	24	14	141	93	375
11. FStudio_3S_RandomForest	113	107	118	115	351
11. FStudio_4S_NaiveBayes	30	33	139	92	370
11. FStudio_4S_Jrip	33	97	138	79	355
11. FStudio_4S_J48	86	74	130	102	362
11. FStudio_4S_SVM	7	4	146	89	381
11. FStudio_4S_RandomForest	128	130	105	118	328
12. Kicker_3S_NaiveBayes	86	77	130	101	361
12. Kicker_3S_Jrip	18	66	142	79	363
12. Kicker_3S_J48	103	20	124	125	373
12. Kicker_3S_SVM	38	30	137	97	371
12. Kicker_3S_RandomForest	111	84	121	118	360
12. Kicker_4S_NaiveBayes	86	94	130	98	358
12. Kicker_4S_Jrip	45	107	136	79	351
12. Kicker_4S_J48	130	120	103	137	343
12. Kicker_4S_SVM	3	4	148	85	381
12. Kicker_4S_RandomForest	113	33	118	134	370

Vorhersagenpunkte für die Saison 2007

Name	Platz B1	Platz B2	Punkte B1	Punkte LM	Punkte B2
TSHirmsieg	103	160	143	78	364
TSTabellenplatz	112	171	142	78	362
TSTabellenplatz2	165	171	125	112	362
TSVersus	172	183	120	109	349
TSForm	155	120	131	113	375
TSForm5	172	191	120	92	332
TSLS	103	160	143	78	364
TS3LS	49	109	150	78	378
TSSiege	152	171	132	98	362
TSNied	132	156	137	93	367
TSTeams	3	13	158	78	394
Random	192	192	100	122	322
1. BestOf_3S_NaiveBayes	58	13	149	96	394
1. BestOf_3S_Jrip	99	104	144	91	379
1. BestOf_3S_J48	134	131	136	101	373
1. BestOf_3S_SVM	116	52	141	105	387
1. BestOf_3S_RandomForest	128	117	138	101	377
1. BestOf_4S_NaiveBayes	34	13	151	92	394
1. BestOf_4S_Jrip	19	70	153	78	384
1. BestOf_4S_J48	141	104	134	111	379
1. BestOf_4S_SVM	86	24	146	99	391
1. BestOf_4S_RandomForest	141	152	134	100	368
1. BestOf_5S_NaiveBayes	34	24	151	89	391
1. BestOf_5S_Jrip	66	124	148	78	374
1. BestOf_5S_J48	103	85	143	96	382
1. BestOf_5S_SVM	66	41	148	92	388
1. BestOf_5S_RandomForest	155	160	131	102	364
2. Top40_3S_NaiveBayes	66	79	148	87	383
2. Top40_3S_Jrip	66	124	148	78	374
2. Top40_3S_J48	186	189	115	112	342
2. Top40_3S_SVM	19	41	153	82	388
2. Top40_3S_RandomForest	155	160	131	102	364
2. Top40_4S_NaiveBayes	66	97	148	85	381
2. Top40_4S_Jrip	103	135	143	86	372
2. Top40_4S_J48	165	142	125	121	371
2. Top40_4S_SVM	14	52	154	79	387
2. Top40_4S_RandomForest	116	52	141	105	387
2. Top40_5S_NaiveBayes	26	52	152	83	387
2. Top40_5S_Jrip	10	41	155	78	388
2. Top40_5S_J48	184	177	116	124	356
2. Top40_5S_SVM	3	5	158	80	396
2. Top40_5S_RandomForest	120	24	140	111	391
3. Top40hg_3S_NaiveBayes	86	85	146	90	382
3. Top40hg_3S_Jrip	26	85	152	78	382
3. Top40hg_3S_J48	14	63	154	78	386
3. Top40hg_3S_SVM	66	1	148	102	398

3. Top40hg_3S_RandomForest	159	104	130	119	379
3. Top40hg_4S_NaiveBayes	66	70	148	88	384
3. Top40hg_4S_Jrip	79	135	147	78	372
3. Top40hg_4S_J48	116	70	141	102	384
3. Top40hg_4S_SVM	93	52	145	97	387
3. Top40hg_4S_RandomForest	141	109	134	110	378
3. Top40hg_5S_NaiveBayes	34	41	151	86	388
3. Top40hg_5S_Jrip	34	100	151	78	380
3. Top40hg_5S_J48	134	120	136	103	375
3. Top40hg_5S_SVM	79	79	147	89	383
3. Top40hg_5S_RandomForest	134	85	136	110	382
4. Top85_3S_NaiveBayes	34	31	151	88	390
4. Top85_3S_Jrip	58	109	149	80	378
4. Top85_3S_J48	186	185	115	117	347
4. Top85_3S_SVM	34	20	151	91	393
4. Top85_3S_RandomForest	162	152	129	110	368
4. Top85_4S_NaiveBayes	34	41	151	86	388
4. Top85_4S_Jrip	79	135	147	78	372
4. Top85_4S_J48	152	159	132	101	365
4. Top85_4S_SVM	66	41	148	92	388
4. Top85_4S_RandomForest	93	31	145	100	390
4. Top85_5S_NaiveBayes	49	52	150	87	387
4. Top85_5S_Jrip	26	85	152	78	382
4. Top85_5S_J48	167	169	124	115	363
4. Top85_5S_SVM	14	31	154	82	390
4. Top85_5S_RandomForest	139	85	135	112	382
5. Komplett_3S_NaiveBayes	103	52	143	101	387
5. Komplett_3S_Jrip	116	174	141	78	360
5. Komplett_3S_J48	179	181	119	112	350
5. Komplett_3S_SVM	134	109	136	106	378
5. Komplett_3S_RandomForest	179	185	119	109	347
5. Komplett_4S_NaiveBayes	34	5	151	94	396
5. Komplett_4S_Jrip	19	70	153	78	384
5. Komplett_4S_J48	141	135	134	104	372
5. Komplett_4S_SVM	139	124	135	104	374
5. Komplett_4S_RandomForest	128	66	138	109	385
5. Komplett_5S_NaiveBayes	58	20	149	95	393
5. Komplett_5S_Jrip	103	160	143	78	364
5. Komplett_5S_J48	172	176	120	118	358
5. Komplett_5S_SVM	112	131	142	89	373
5. Komplett_5S_RandomForest	128	124	138	98	374
6. InfoGain40_3S_NaiveBayes	14	38	154	81	389
6. InfoGain40_3S_Jrip	86	147	146	78	370
6. InfoGain40_3S_J48	66	13	148	98	394
6. InfoGain40_3S_SVM	93	20	145	103	393
6. InfoGain40_3S_RandomForest	159	150	130	109	369
6. InfoGain40_4S_NaiveBayes	34	38	151	87	389

6. InfoGain40_4S_Jrip	26	85	152	78	382
6. InfoGain40_4S_J48	155	124	131	112	374
6. InfoGain40_4S_SVM	93	31	145	100	390
6. InfoGain40_4S_RandomForest	120	117	140	97	377
6. InfoGain40_5S_NaiveBayes	10	5	155	86	396
6. InfoGain40_5S_Jrip	49	109	150	78	378
6. InfoGain40_5S_J48	86	66	146	93	385
6. InfoGain40_5S_SVM	79	41	147	94	388
6. InfoGain40_5S_RandomForest	124	70	139	106	384
7. InfoGain85_3S_NaiveBayes	49	52	150	87	387
7. InfoGain85_3S_Jrip	1	1	159	80	398
7. InfoGain85_3S_J48	159	169	130	103	363
7. InfoGain85_3S_SVM	86	10	146	103	395
7. InfoGain85_3S_RandomForest	162	135	129	114	372
7. InfoGain85_4S_NaiveBayes	66	63	148	90	386
7. InfoGain85_4S_Jrip	93	152	145	78	368
7. InfoGain85_4S_J48	134	131	136	101	373
7. InfoGain85_4S_SVM	58	10	149	97	395
7. InfoGain85_4S_RandomForest	124	100	139	102	380
7. InfoGain85_5S_NaiveBayes	66	38	148	93	389
7. InfoGain85_5S_Jrip	120	142	140	91	371
7. InfoGain85_5S_J48	124	70	139	106	384
7. InfoGain85_5S_SVM	79	31	147	96	390
7. InfoGain85_5S_RandomForest	148	97	133	115	381
8. Fold40_3S_NaiveBayes	66	70	148	88	384
8. Fold40_3S_Jrip	99	157	144	78	366
8. Fold40_3S_J48	188	183	114	121	349
8. Fold40_3S_SVM	8	20	156	81	393
8. Fold40_3S_RandomForest	103	100	143	94	380
8. Fold40_4S_NaiveBayes	58	66	149	87	385
8. Fold40_4S_Jrip	58	85	149	84	382
8. Fold40_4S_J48	182	177	117	122	356
8. Fold40_4S_SVM	49	79	150	83	383
8. Fold40_4S_RandomForest	124	104	139	101	379
8. Fold40_5S_NaiveBayes	58	70	149	86	384
8. Fold40_5S_Jrip	10	41	155	78	388
8. Fold40_5S_J48	172	177	120	116	356
8. Fold40_5S_SVM	10	31	155	80	390
8. Fold40_5S_RandomForest	141	135	134	104	372
9. Fold85_3S_NaiveBayes	49	41	150	88	388
9. Fold85_3S_Jrip	19	24	153	85	391
9. Fold85_3S_J48	189	190	111	113	335
9. Fold85_3S_SVM	8	5	156	84	396
9. Fold85_3S_RandomForest	148	142	133	105	371
9. Fold85_4S_NaiveBayes	6	5	157	82	396
9. Fold85_4S_Jrip	79	135	147	78	372
9. Fold85_4S_J48	168	160	123	118	364

9. Fold85_4S_SVM	34	79	151	81	383
9. Fold85_4S_RandomForest	148	104	133	113	379
9. Fold85_5S_NaiveBayes	26	41	152	84	388
9. Fold85_5S_Jrip	19	70	153	78	384
9. Fold85_5S_J48	190	180	109	134	352
9. Fold85_5S_SVM	14	41	154	80	388
9. Fold85_5S_RandomForest	132	120	137	101	375
10. Differenz_3S_NaiveBayes	49	24	150	91	391
10. Differenz_3S_Jrip	86	147	146	78	370
10. Differenz_3S_J48	172	174	120	120	360
10. Differenz_3S_SVM	19	52	153	81	387
10. Differenz_3S_RandomForest	148	120	133	109	375
10. Differenz_4S_NaiveBayes	34	52	151	85	387
10. Differenz_4S_Jrip	3	13	158	78	394
10. Differenz_4S_J48	181	142	118	135	371
10. Differenz_4S_SVM	1	3	159	79	397
10. Differenz_4S_RandomForest	93	10	145	105	395
10. Differenz_5S_NaiveBayes	58	63	149	88	386
10. Differenz_5S_Jrip	103	160	143	78	364
10. Differenz_5S_J48	190	188	109	125	343
10. Differenz_5S_SVM	6	3	157	83	397
10. Differenz_5S_RandomForest	141	124	134	106	374
11. FStudio_3S_NaiveBayes	170	97	122	137	381
11. FStudio_3S_Jrip	34	79	151	81	383
11. FStudio_3S_J48	120	109	140	98	378
11. FStudio_3S_SVM	19	13	153	88	394
11. FStudio_3S_RandomForest	168	160	123	118	364
11. FStudio_4S_NaiveBayes	172	109	120	138	378
11. FStudio_4S_Jrip	49	85	150	82	382
11. FStudio_4S_J48	128	142	138	95	371
11. FStudio_4S_SVM	66	52	148	91	387
11. FStudio_4S_RandomForest	152	117	132	113	377
11. FStudio_5S_NaiveBayes	141	109	134	110	378
11. FStudio_5S_Jrip	34	100	151	78	380
11. FStudio_5S_J48	99	124	144	86	374
11. FStudio_5S_SVM	49	31	150	90	390
11. FStudio_5S_RandomForest	172	181	120	110	350
12. Kicker_3S_NaiveBayes	79	24	147	97	391
12. Kicker_3S_Jrip	99	131	144	85	373
12. Kicker_3S_J48	184	185	116	115	347
12. Kicker_3S_SVM	103	160	143	78	364
12. Kicker_3S_RandomForest	112	66	142	101	385
12. Kicker_4S_NaiveBayes	34	24	151	89	391
12. Kicker_4S_Jrip	26	85	152	78	382
12. Kicker_4S_J48	164	152	126	116	368
12. Kicker_4S_SVM	86	147	146	78	370
12. Kicker_4S_RandomForest	112	79	142	99	383

12. Kicker_5S_NaiveBayes	34	13	151	92	394
12. Kicker_5S_Jrip	26	85	152	78	382
12. Kicker_5S_J48	182	157	117	132	366
12. Kicker_5S_SVM	26	85	152	78	382
12. Kicker_5S_RandomForest	170	150	122	125	369

Vorhersagenpunkte für die Saison 2008

Name	Platz B1	Platz B2	Punkte B1	Punkte LM	Punkte B2
TSHeimsieg	31	89	147	74	368
TSTabellenplatz	43	123	145	75	365
TSTabellenplatz2	121	14	139	110	388
TSVersus	202	144	125	112	362
TSForm	43	20	145	95	385
TSForm5	173	226	134	77	345
TSLS	139	200	138	74	350
TS3LS	160	213	136	76	348
TSSiege	121	116	139	88	366
TSNied	94	54	141	92	374
TSTeams	38	116	146	74	366
Random	248	240	110	120	340
1. BestOf_3S_NaiveBayes	43	20	145	95	385
1. BestOf_3S_Jrip	139	134	138	88	364
1. BestOf_3S_J48	205	66	124	124	372
1. BestOf_3S_SVM	182	235	131	81	343
1. BestOf_3S_RandomForest	182	154	131	99	361
1. BestOf_4S_NaiveBayes	38	28	146	88	380
1. BestOf_4S_Jrip	121	194	139	74	352
1. BestOf_4S_J48	160	44	136	105	377
1. BestOf_4S_SVM	152	200	137	76	350
1. BestOf_4S_RandomForest	121	73	139	92	370
1. BestOf_5S_NaiveBayes	64	46	143	90	376
1. BestOf_5S_Jrip	52	89	144	80	368
1. BestOf_5S_J48	209	175	123	111	357
1. BestOf_5S_SVM	152	200	137	76	350
1. BestOf_5S_RandomForest	199	154	127	107	361
1. BestOf_6S_NaiveBayes	43	33	145	88	378
1. BestOf_6S_Jrip	18	46	151	74	376
1. BestOf_6S_J48	64	18	143	100	386
1. BestOf_6S_SVM	79	140	142	79	363
1. BestOf_6S_RandomForest	178	103	133	101	367
2. Top40_3S_NaiveBayes	139	123	138	89	365
2. Top40_3S_Jrip	168	188	135	83	353
2. Top40_3S_J48	241	243	112	114	338
2. Top40_3S_SVM	121	196	139	73	351
2. Top40_3S_RandomForest	234	213	117	114	348
2. Top40_4S_NaiveBayes	94	83	141	87	369
2. Top40_4S_Jrip	94	134	141	82	364

2. Top40_4S_J48	182	123	131	103	365
2. Top40_4S_SVM	107	186	140	74	354
2. Top40_4S_RandomForest	121	49	139	97	375
2. Top40_5S_NaiveBayes	94	89	141	86	368
2. Top40_5S_Jrip	21	30	150	79	379
2. Top40_5S_J48	231	181	118	120	356
2. Top40_5S_SVM	139	188	138	77	353
2. Top40_5S_RandomForest	94	49	141	93	375
2. Top40_6S_NaiveBayes	94	103	141	85	367
2. Top40_6S_Jrip	1	4	159	77	395
2. Top40_6S_J48	250	252	107	-107	107
2. Top40_6S_SVM	107	251	140	-140	140
2. Top40_6S_RandomForest	79	250	142	-142	142
3. Top40hg_3S_NaiveBayes	139	116	138	90	366
3. Top40hg_3S_Jrip	4	4	157	81	395
3. Top40hg_3S_J48	205	226	124	97	345
3. Top40hg_3S_SVM	152	163	137	85	359
3. Top40hg_3S_RandomForest	205	200	124	102	350
3. Top40hg_4S_NaiveBayes	52	54	144	86	374
3. Top40hg_4S_Jrip	64	161	143	74	360
3. Top40hg_4S_J48	190	163	130	99	359
3. Top40hg_4S_SVM	139	188	138	77	353
3. Top40hg_4S_RandomForest	121	61	139	95	373
3. Top40hg_5S_NaiveBayes	79	73	142	86	370
3. Top40hg_5S_Jrip	31	89	147	74	368
3. Top40hg_5S_J48	213	226	122	101	345
3. Top40hg_5S_SVM	168	218	135	77	347
3. Top40hg_5S_RandomForest	218	123	121	123	365
3. Top40hg_6S_NaiveBayes	79	73	142	86	370
3. Top40hg_6S_Jrip	6	22	155	74	384
3. Top40hg_6S_J48	27	1	148	101	397
3. Top40hg_6S_SVM	64	103	143	81	367
3. Top40hg_6S_RandomForest	213	163	122	115	359
4. Top85_3S_NaiveBayes	139	123	138	89	365
4. Top85_3S_Jrip	24	66	149	74	372
4. Top85_3S_J48	238	231	115	114	344
4. Top85_3S_SVM	94	169	141	76	358
4. Top85_3S_RandomForest	225	222	119	108	346
4. Top85_4S_NaiveBayes	121	103	139	89	367
4. Top85_4S_Jrip	10	25	154	74	382
4. Top85_4S_J48	225	235	119	105	343
4. Top85_4S_SVM	195	239	129	83	341
4. Top85_4S_RandomForest	173	123	134	97	365
4. Top85_5S_NaiveBayes	107	89	140	88	368
4. Top85_5S_Jrip	31	89	147	74	368
4. Top85_5S_J48	218	175	121	115	357
4. Top85_5S_SVM	94	154	141	79	361

4. Top85_5S_RandomForest	160	44	136	105	377
4. Top85_6S_NaiveBayes	107	123	140	85	365
4. Top85_6S_Jrip	21	54	150	74	374
4. Top85_6S_J48	236	218	116	115	347
4. Top85_6S_SVM	79	169	142	74	358
4. Top85_6S_RandomForest	182	140	131	101	363
5. Komplett_3S_NaiveBayes	79	30	142	95	379
5. Komplett_3S_Jrip	24	66	149	74	372
5. Komplett_3S_J48	241	196	112	127	351
5. Komplett_3S_SVM	209	196	123	105	351
5. Komplett_3S_RandomForest	160	103	136	95	367
5. Komplett_4S_NaiveBayes	52	33	144	90	378
5. Komplett_4S_Jrip	38	123	146	73	365
5. Komplett_4S_J48	231	240	118	104	340
5. Komplett_4S_SVM	202	247	125	85	335
5. Komplett_4S_RandomForest	195	161	129	102	360
5. Komplett_5S_NaiveBayes	64	33	143	92	378
5. Komplett_5S_Jrip	52	144	144	74	362
5. Komplett_5S_J48	225	243	119	100	338
5. Komplett_5S_SVM	121	175	139	79	357
5. Komplett_5S_RandomForest	190	49	130	115	375
5. Komplett_6S_NaiveBayes	52	30	144	91	379
5. Komplett_6S_Jrip	10	25	154	74	382
5. Komplett_6S_J48	250	246	107	122	336
5. Komplett_6S_SVM	43	103	145	77	367
5. Komplett_6S_RandomForest	190	61	130	113	373
6. InfoGain40_3S_NaiveBayes	64	73	143	84	370
6. InfoGain40_3S_Jrip	79	83	142	85	369
6. InfoGain40_3S_J48	195	208	129	91	349
6. InfoGain40_3S_SVM	168	188	135	83	353
6. InfoGain40_3S_RandomForest	173	49	134	107	375
6. InfoGain40_4S_NaiveBayes	52	66	144	84	372
6. InfoGain40_4S_Jrip	14	11	152	87	391
6. InfoGain40_4S_J48	205	181	124	108	356
6. InfoGain40_4S_SVM	173	231	134	76	344
6. InfoGain40_4S_RandomForest	213	144	122	118	362
6. InfoGain40_5S_NaiveBayes	107	123	140	85	365
6. InfoGain40_5S_Jrip	1	8	159	74	392
6. InfoGain40_5S_J48	209	226	123	99	345
6. InfoGain40_5S_SVM	168	208	135	79	349
6. InfoGain40_5S_RandomForest	182	103	131	105	367
6. InfoGain40_6S_NaiveBayes	64	89	143	82	368
6. InfoGain40_6S_Jrip	6	14	155	78	388
6. InfoGain40_6S_J48	121	33	139	100	378
6. InfoGain40_6S_SVM	107	169	140	78	358
6. InfoGain40_6S_RandomForest	178	61	133	107	373
7. InfoGain85_3S_NaiveBayes	79	89	142	84	368

7. InfoGain85_3S_Jrip	52	144	144	74	362
7. InfoGain85_3S_J48	239	243	114	110	338
7. InfoGain85_3S_SVM	213	231	122	100	344
7. InfoGain85_3S_RandomForest	182	169	131	96	358
7. InfoGain85_4S_NaiveBayes	121	103	139	89	367
7. InfoGain85_4S_Jrip	1	8	159	74	392
7. InfoGain85_4S_J48	241	249	112	108	332
7. InfoGain85_4S_SVM	182	235	131	81	343
7. InfoGain85_4S_RandomForest	152	140	137	89	363
7. InfoGain85_5S_NaiveBayes	107	116	140	86	366
7. InfoGain85_5S_Jrip	27	73	148	74	370
7. InfoGain85_5S_J48	200	154	126	109	361
7. InfoGain85_5S_SVM	107	144	140	82	362
7. InfoGain85_5S_RandomForest	121	73	139	92	370
7. InfoGain85_6S_NaiveBayes	79	116	142	82	366
7. InfoGain85_6S_Jrip	38	73	146	78	370
7. InfoGain85_6S_J48	202	200	125	100	350
7. InfoGain85_6S_SVM	64	116	143	80	366
7. InfoGain85_6S_RandomForest	198	181	128	100	356
8. Fold40_3S_NaiveBayes	107	83	140	89	369
8. Fold40_3S_Jrip	12	28	153	74	380
8. Fold40_3S_J48	246	222	111	124	346
8. Fold40_3S_SVM	139	200	138	74	350
8. Fold40_3S_RandomForest	239	222	114	118	346
8. Fold40_4S_NaiveBayes	121	123	139	87	365
8. Fold40_4S_Jrip	79	144	142	78	362
8. Fold40_4S_J48	252	242	103	133	339
8. Fold40_4S_SVM	107	186	140	74	354
8. Fold40_4S_RandomForest	160	33	136	106	378
8. Fold40_5S_NaiveBayes	94	103	141	85	367
8. Fold40_5S_Jrip	168	213	135	78	348
8. Fold40_5S_J48	225	154	119	123	361
8. Fold40_5S_SVM	160	222	136	74	346
8. Fold40_5S_RandomForest	43	27	145	91	381
8. Fold40_6S_NaiveBayes	64	103	143	81	367
8. Fold40_6S_Jrip	31	89	147	74	368
8. Fold40_6S_J48	241	238	112	118	342
8. Fold40_6S_SVM	121	188	139	75	353
8. Fold40_6S_RandomForest	218	196	121	109	351
9. Fold85_3S_NaiveBayes	152	134	137	90	364
9. Fold85_3S_Jrip	79	61	142	89	373
9. Fold85_3S_J48	234	208	117	115	349
9. Fold85_3S_SVM	94	163	141	77	359
9. Fold85_3S_RandomForest	190	49	130	115	375
9. Fold85_4S_NaiveBayes	139	103	138	91	367
9. Fold85_4S_Jrip	52	144	144	74	362
9. Fold85_4S_J48	241	247	112	111	335

9. Fold85_4S_SVM	107	175	140	77	357
9. Fold85_4S_RandomForest	27	33	148	82	378
9. Fold85_5S_NaiveBayes	139	123	138	89	365
9. Fold85_5S_Jrip	14	33	152	74	378
9. Fold85_5S_J48	221	194	120	112	352
9. Fold85_5S_SVM	121	163	139	81	359
9. Fold85_5S_RandomForest	121	54	139	96	374
9. Fold85_6S_NaiveBayes	121	116	139	88	366
9. Fold85_6S_Jrip	79	169	142	74	358
9. Fold85_6S_J48	221	163	120	119	359
9. Fold85_6S_SVM	43	89	145	78	368
9. Fold85_6S_RandomForest	79	24	142	99	383
10. Differenz_3S_NaiveBayes	94	54	141	92	374
10. Differenz_3S_Jrip	18	46	151	74	376
10. Differenz_3S_J48	249	200	109	132	350
10. Differenz_3S_SVM	52	134	144	76	364
10. Differenz_3S_RandomForest	160	33	136	106	378
10. Differenz_4S_NaiveBayes	64	54	143	88	374
10. Differenz_4S_Jrip	27	73	148	74	370
10. Differenz_4S_J48	246	231	111	122	344
10. Differenz_4S_SVM	31	83	147	75	369
10. Differenz_4S_RandomForest	107	33	140	98	378
10. Differenz_5S_NaiveBayes	139	73	138	94	370
10. Differenz_5S_Jrip	38	73	146	78	370
10. Differenz_5S_J48	221	208	120	109	349
10. Differenz_5S_SVM	64	154	143	75	361
10. Differenz_5S_RandomForest	79	13	142	105	389
10. Differenz_6S_NaiveBayes	94	71	141	89	371
10. Differenz_6S_Jrip	21	54	150	74	374
10. Differenz_6S_J48	225	213	119	110	348
10. Differenz_6S_SVM	52	144	144	74	362
10. Differenz_6S_RandomForest	231	218	118	111	347
11. FStudio_3S_NaiveBayes	31	1	147	103	397
11. FStudio_3S_Jrip	12	12	153	84	390
11. FStudio_3S_J48	139	154	138	85	361
11. FStudio_3S_SVM	152	169	137	84	358
11. FStudio_3S_RandomForest	209	134	123	118	364
11. FStudio_4S_NaiveBayes	52	6	144	106	394
11. FStudio_4S_Jrip	6	8	155	82	392
11. FStudio_4S_J48	64	61	143	87	373
11. FStudio_4S_SVM	173	213	134	80	348
11. FStudio_4S_RandomForest	213	185	122	111	355
11. FStudio_5S_NaiveBayes	14	1	152	93	397
11. FStudio_5S_Jrip	14	33	152	74	378
11. FStudio_5S_J48	52	33	144	90	378
11. FStudio_5S_SVM	152	188	137	79	353
11. FStudio_5S_RandomForest	180	66	132	108	372

11. FStudio_6S_NaiveBayes	18	14	151	86	388
11. FStudio_6S_Jrip	6	22	155	74	384
11. FStudio_6S_J48	64	89	143	82	368
11. FStudio_6S_SVM	64	134	143	78	364
11. FStudio_6S_RandomForest	200	103	126	115	367
12. Kicker_3S_NaiveBayes	139	103	138	91	367
12. Kicker_3S_Jrip	160	140	136	91	363
12. Kicker_3S_J48	64	17	143	101	387
12. Kicker_3S_SVM	79	144	142	78	362
12. Kicker_3S_RandomForest	225	226	119	107	345
12. Kicker_4S_NaiveBayes	107	83	140	89	369
12. Kicker_4S_Jrip	5	18	156	74	386
12. Kicker_4S_J48	236	218	116	115	347
12. Kicker_4S_SVM	121	175	139	79	357
12. Kicker_4S_RandomForest	182	71	131	109	371
12. Kicker_5S_NaiveBayes	43	89	145	78	368
12. Kicker_5S_Jrip	31	89	147	74	368
12. Kicker_5S_J48	24	7	149	95	393
12. Kicker_5S_SVM	107	175	140	77	357
12. Kicker_5S_RandomForest	180	89	132	104	368
12. Kicker_6S_NaiveBayes	43	83	145	79	369
12. Kicker_6S_Jrip	94	181	141	74	356
12. Kicker_6S_J48	221	208	120	109	349
12. Kicker_6S_SVM	152	200	137	76	350
12. Kicker_6S_RandomForest	190	144	130	102	362

Anhang G: Komplette Tabellen aus den Tabellenvorhersagen im Kapitel 5.3

Legende

Dunkelgrün – Meister

Olivgrün – Int.Wettbewerbe (Champions League)

Hellgrün – Int.Wettbewerbe (Europa League)

Rot – Absteiger

(M) – der aktuelle Meister

(N) – der aktuelle Aufsteiger

Platz – der vorhergesagte Platz vom jeweiligen Klassifizierer für diese Saison

Vereinsname – entsprechender Verein der 1.Bundesliga

Punkte – die vorhergesagte Punktzahl vom jeweiligen Klassifizierer für diese Saison

Saison 2005

FStudio_3S_NaiveBayes (1.)

Platz	Vereinsname	Punkte
1.	Bayern München (M)	102
2.	Werder Bremen	96
3.	Hamburger SV	82
4.	FC Schalke 04	81
5.	VFB Stuttgart	71
6.	Borussia Dortmund	69
7.	Hertha BSC Berlin	61
8.	Bayer Leverkusen	56
9.	Hannover 96	39
10.	Mönchengladbach	35
11.	FSV Mainz 05	31
12.	VFL Wolfsburg	30
13.	Arminia Bielefeld	28
14.	1.FC Köln (N)	24
15.	1.FC Nürnberg (N)	24
16.	MSV Duisburg	23
17.	1.FC Kaiserslautern	22
18.	Eintracht Frankfurt (N)	18

Top40hg_3S_JRip (10.)

Platz	Vereinsname	Punkte
1.	Bayern München (M)	87
2.	Hamburger SV	70
3.	FC Schalke 04	69
4.	Werder Bremen	63
5.	Borussia Dortmund	60
6.	Bayer Leverkusen	59
7.	VFB Stuttgart	57
8.	Hertha BSC Berlin	55
9.	VFL Wolfsburg	48
10.	1.FC Kaiserslautern	47
11.	Hannover 96	45
12.	Mönchengladbach	44
13.	1.FC Nürnberg	42
14.	FSV Mainz 05	38
15.	Arminia Bielefeld	33
16.	MSV Duisburg (N)	31
17.	1.FC Köln (N)	23
18.	Eintracht Frankfurt (N)	23

InfoGain40_3S_J48 (13.)

Platz	Vereinsname	Punkte
1.	Werder Bremen (M)	87
2.	Bayern München	81
3.	Bayer Leverkusen	78
4.	FC Schalke 04	78
5.	VFB Stuttgart	76
6.	Borussia Dortmund	73
7.	1.FC Kaiserslautern	55
8.	Arminia Bielefeld	48
9.	Hannover 96	47
10.	Hertha BSC Berlin	46
11.	Mönchengladbach	41
12.	Hamburger SV	39
13.	1.FC Köln (N)	33
14.	VFL Wolfsburg	31
15.	FSV Mainz 05	27
16.	Eintracht Frankfurt (N)	26
17.	1.FC Nürnberg	23
18.	MSV Duisburg (N)	21

Komplett_3S_SVM (5.)

Platz	Vereinsname	Punkte
1.	Werder Bremen	81
2.	Bayern München (M)	78
3.	VFB Stuttgart	66
4.	Bayer Leverkusen	63
5.	Mönchengladbach	55
6.	MSV Duisburg	52
7.	Borussia Dortmund	52
8.	FC Schalke 04	52
9.	1.FC Kaiserslautern	52
10.	Hamburger SV	48
11.	VFL Wolfsburg	48
12.	Hertha BSC Berlin	45
13.	FSV Mainz 05	43
14.	Arminia Bielefeld	40
15.	Hannover 96	40
16.	1.FC Nürnberg (N)	30
17.	Eintracht Frankfurt (N)	24
18.	1.FC Köln (N)	20

Fold85_3S_RandomForest (35.)

Platz	Vereinsname	Punkte
1.	Borussia Dortmund	71
2.	Werder Bremen	71
3.	Hamburger SV	71
4.	Hertha BSC Berlin	68
5.	Bayer Leverkusen	62
6.	FC Schalke 04	61
7.	Bayern München (M)	56
8.	VFB Stuttgart	55
9.	VFL Wolfsburg	48
10.	1.FC Kaiserslautern	44
11.	Mönchengladbach	41
12.	Hannover 96	39
13.	1.FC Köln (N)	36
14.	Arminia Bielefeld	33
15.	FSV Mainz 05	30
16.	Eintracht Frankfurt (N)	30
17.	MSV Duisburg	27
18.	1.FC Nürnberg (N)	25

Saison 2006

Fold85_3S_NaiveBayes (28.)

Platz	Vereinsname	Punkte
1.	Bayern München (M)	100
2.	Werder Bremen	95
3.	FC Schalke 04	91
4.	Bayer Leverkusen	79
5.	Hamburger SV	76
6.	VFB Stuttgart	73
7.	Borussia Dortmund	70
8.	Hertha BSC Berlin	60
9.	Hannover 96	45
10.	Mönchengladbach	42
11.	1.FC Nürnberg	40
12.	VFL Wolfsburg	39
13.	Eintracht Frankfurt	25
14.	FSV Mainz 05	21
15.	Arminia Bielefeld	20
16.	VFL Bochum (N)	19
17.	Energie Cottbus (N)	7
18.	Alemannia Aachen (N)	4

Fold40_4S_JRip (8.)

Platz	Vereinsname	Punkte
1.	Hamburger SV	95
2.	Werder Bremen	95
3.	Bayern München (M)	88
4.	FC Schalke 04	60
5.	Bayer Leverkusen	57
6.	Hertha BSC Berlin	51
7.	VFB Stuttgart	49
8.	1.FC Nürnberg	42
9.	Borussia Dortmund	42
10.	Mönchengladbach	41
11.	Arminia Bielefeld	40
12.	Hannover 96	38
13.	VFL Wolfsburg	37
14.	Eintracht Frankfurt	36
15.	Energie Cottbus (N)	33
16.	FSV Mainz 05	33
17.	VFL Bochum (N)	33
18.	Alemannia Aachen (N)	30

InfoGain40_3S_J48 (14.)

Platz	Vereinsname	Punkte
1.	FC Schalke 04	91
2.	Werder Bremen	83
3.	Bayern München (M)	71
4.	VFB Stuttgart	68
5.	Bayer Leverkusen	62
6.	Borussia Dortmund	59
7.	Arminia Bielefeld	54
8.	Eintracht Frankfurt	51
9.	Hannover 96	50
10.	FSV Mainz 05	45
11.	Hamburger SV	44
12.	Mönchengladbach	36
13.	1.FC Nürnberg	31
14.	Hertha BSC Berlin	31
15.	VFL Wolfsburg	26
16.	VFL Bochum (N)	26
17.	Alemannia Aachen (N)	26
18.	Energie Cottbus (N)	25

BestOf_4S_SVM (1.)

Platz	Vereinsname	Punkte
1.	Werder Bremen	85
2.	Bayern München (M)	84
3.	VFB Stuttgart	83
4.	FC Schalke 04	76
5.	Bayer Leverkusen	71
6.	VFL Bochum (N)	60
7.	Borussia Dortmund	58
8.	Hertha BSC Berlin	54
9.	Mönchengladbach	43
10.	Hamburger SV	42
11.	VFL Wolfsburg	42
12.	Arminia Bielefeld	38
13.	FSV Mainz 05	36
14.	Alemannia Aachen (N)	36
15.	Hannover 96	33
16.	1.FC Nürnberg	30
17.	Eintracht Frankfurt	21
18.	Energie Cottbus (N)	14

InfoGain85_4S_RandomForest (37.)

Platz	Vereinsname	Punkte
1.	Bayern München (M)	79
2.	Werder Bremen	79
3.	Hamburger SV	74
4.	FC Schalke 04	72
5.	Bayer Leverkusen	58
6.	VFB Stuttgart	52
7.	VFL Wolfsburg	48
8.	Hertha BSC Berlin	48
9.	Borussia Dortmund	45
10.	Mönchengladbach	44
11.	FSV Mainz 05	43
12.	Arminia Bielefeld	36
13.	1.FC Nürnberg	34
14.	Eintracht Frankfurt	31
15.	Hannover 96	30
16.	Alemannia Aachen (N)	30
17.	VFL Bochum (N)	28
18.	Energie Cottbus (N)	22

Saison 2007

Fold85_4S_NaiveBayes (6.)

Platz	Vereinsname	Punkte
1.	Bayern München	98
2.	Werder Bremen	96
3.	VFB Stuttgart (M)	81
4.	FC Schalke 04	80
5.	Bayer Leverkusen	80
6.	Hamburger SV	75
7.	Borussia Dortmund	66
8.	Hertha BSC Berlin	56
9.	Hannover 96	54
10.	VFL Wolfsburg	42
11.	1.FC Nürnberg	34
12.	VFL Bochum	33
13.	Eintracht Frankfurt	25
14.	Arminia Bielefeld	23
15.	Karlsruher SC (N)	14
16.	FC Hansa Rostock (N)	14
17.	Energie Cottbus	12
18.	MSV Duisburg (N)	5

InfoGain85_3S_JRip (1.)

Platz	Vereinsname	Punkte
1.	Bayern München	83
2.	FC Schalke 04	83
3.	Hamburger SV	75
4.	Werder Bremen	74
5.	Bayer Leverkusen	70
6.	VFB Stuttgart (M)	56
7.	VFL Wolfsburg	45
8.	Borussia Dortmund	45
9.	VFL Bochum	45
10.	Hannover 96	43
11.	1.FC Nürnberg	40
12.	Hertha BSC Berlin	39
13.	Eintracht Frankfurt	39
14.	Karlsruher SC (N)	39
15.	MSV Duisburg (N)	36
16.	Arminia Bielefeld	33
17.	FC Hansa Rostock (N)	33
18.	Energie Cottbus	30

Top40hg_3S_J48 (14.)

Platz	Vereinsname	Punkte
1.	Bayern München	81
2.	VFB Stuttgart (M)	81
3.	FC Schalke 04	81
4.	Werder Bremen	78
5.	Bayer Leverkusen	75
6.	Hamburger SV	69
7.	Borussia Dortmund	69
8.	Hertha BSC Berlin	65
9.	Hannover 96	57
10.	Arminia Bielefeld	42
11.	1.FC Nürnberg	39
12.	VFL Wolfsburg	36
13.	VFL Bochum	28
14.	Energie Cottbus	25
15.	Eintracht Frankfurt	24
16.	FC Hansa Rostock (N)	24
17.	Karlsruher SC (N)	21
18.	MSV Duisburg (N)	21

Differenz_4S_SVM (1.)

Platz	Vereinsname	Punkte
1.	Werder Bremen	97
2.	Bayern München	93
3.	Bayer Leverkusen	90
4.	VFB Stuttgart (M)	88
5.	FC Schalke 04	72
6.	Hannover 96	69
7.	Hamburger SV	63
8.	VFL Bochum	40
9.	Eintracht Frankfurt	34
10.	Hertha BSC Berlin	33
11.	Borussia Dortmund	33
12.	FC Hansa Rostock (N)	30
13.	MSV Duisburg (N)	30
14.	1.FC Nürnberg	30
15.	Arminia Bielefeld	30
16.	Karlsruher SC (N)	30
17.	VFL Wolfsburg	28
18.	Energie Cottbus	25

Top85_4S_RandomForest (92.)

Platz	Vereinsname	Punkte
1.	Werder Bremen	91
2.	Bayern München	86
3.	VFB Stuttgart (M)	75
4.	Bayer Leverkusen	62
5.	Hannover 96	61
6.	Hamburger SV	56
7.	FC Schalke 04	56
8.	1.FC Nürnberg	46
9.	VFL Wolfsburg	39
10.	Borussia Dortmund	39
11.	VFL Bochum	38
12.	Eintracht Frankfurt	36
13.	Arminia Bielefeld	35
14.	Energie Cottbus	32
15.	Hertha BSC Berlin	30
16.	MSV Duisburg (N)	26
17.	FC Hansa Rostock (N)	25
18.	Karlsruher SC (N)	19

Saison 2008

FStudio_5S_NaiveBayes (14.)

Platz	Vereinsname	Punkte
1.	Bayern München (M)	93
2.	Bayer Leverkusen	77
3.	Hamburger SV	76
4.	1899 Hoffenheim (N)	73
5.	Hertha BSC Berlin	71
6.	Werder Bremen	67
7.	VFL Wolfsburg	67
8.	FC Schalke 04	67
9.	VFB Stuttgart	65
10.	Borussia Dortmund	51
11.	VFL Bochum	26
12.	Arminia Bielefeld	24
13.	1.FC Köln (N)	22
14.	Eintracht Frankfurt	21
15.	Hannover 96	20
16.	Energie Cottbus	20
17.	Karlsruher SC	17
18.	Mönchengladbach (N)	10

Top40_6S_JRip (1.)

Platz	Vereinsname	Punkte
1.	Bayern München (M)	81
2.	FC Schalke 04	73
3.	VFL Wolfsburg	63
4.	Hamburger SV	62
5.	Bayer Leverkusen	59
6.	Werder Bremen	58
7.	Borussia Dortmund	57
8.	Hannover 96	54
9.	Hertha BSC Berlin	51
10.	VFL Bochum	46
11.	Karlsruher SC	44
12.	VFB Stuttgart	43
13.	Mönchengladbach (N)	39
14.	Arminia Bielefeld	39
15.	Eintracht Frankfurt	38
16.	Energie Cottbus	36
17.	1.FC Köln (N)	36
18.	1899 Hoffenheim (N)	33

Kicker_5S_J48 (24.)

Platz	Vereinsname	Punkte
1.	VFB Stuttgart	86
2.	Bayern München (M)	72
3.	Bayer Leverkusen	69
4.	Werder Bremen	66
5.	FC Schalke 04	66
6.	Hamburger SV	58
7.	Hannover 96	50
8.	1.FC Köln (N)	42
9.	VFL Wolfsburg	42
10.	Hertha BSC Berlin	42
11.	VFL Bochum	39
12.	Mönchengladbach (N)	38
13.	Energie Cottbus	38
14.	Karlsruher SC	38
15.	1899 Hoffenheim (N)	36
16.	Eintracht Frankfurt	30
17.	Arminia Bielefeld	26
18.	Borussia Dortmund	23

Differenz_4S_SVM (31.)

Platz	Vereinsname	Punkte
1.	Bayern München (M)	99
2.	FC Schalke 04	90
3.	Werder Bremen	89
4.	Bayer Leverkusen	82
5.	Hamburger SV	72
6.	VFB Stuttgart	67
7.	Hannover 96	65
8.	Borussia Dortmund	54
9.	1899 Hoffenheim (N)	40
10.	Hertha BSC Berlin	37
11.	Mönchengladbach (N)	30
12.	Arminia Bielefeld	30
13.	Eintracht Frankfurt	30
14.	VFL Wolfsburg	30
15.	VFL Bochum	27
16.	Energie Cottbus	24
17.	1.FC Köln (N)	24
18.	Karlsruher SC	23

Fold85_4S_RandomForest (27.)

Platz	Vereinsname	Punkte
1.	Bayern München (M)	94
2.	FC Schalke 04	83
3.	Werder Bremen	82
4.	VFB Stuttgart	75
5.	Bayer Leverkusen	75
6.	Hamburger SV	71
7.	VFL Wolfsburg	48
8.	Borussia Dortmund	45
9.	Hertha BSC Berlin	43
10.	1899 Hoffenheim (N)	35
11.	Karlsruher SC	35
12.	Mönchengladbach (N)	33
13.	Arminia Bielefeld	32
14.	Energie Cottbus	29
15.	VFL Bochum	28
16.	Hannover 96	27
17.	Eintracht Frankfurt	24
18.	1.FC Köln (N)	21

Anhang H: Komplette Tabellen aus den Tabellenvorhersagen im Kapitel 5.4 (Große Trainingssets)

(Legende wie in Anhang F)

BestOF1963_NaiveBayes

Platz	Vereinsname	Punkte
1.	Bayern München (M)	102
2.	Hamburger SV	86
3.	FC Schalke 04	74
4.	VFB Stuttgart	72
5.	Bayer Leverkusen	72
6.	VFL Wolfsburg	69
7.	Werder Bremen	67
8.	Hertha BSC Berlin	66
9.	Borussia Dortmund	62
10.	1899 Hoffenheim (N)	53
11.	Eintracht Frankfurt	34
12.	Hannover 96	30
13.	1.FC Köln (N)	30
14.	Arminia Bielefeld	18
15.	VFL Bochum	18
16.	Mönchengladbach (N)	15
17.	Karlsruher SC	14
18.	Energie Cottbus	4

Komplett1963_JRip

Platz	Vereinsname	Punkte
1.	Bayern München (M)	78
2.	Werder Bremen	60
3.	FC Schalke 04	60
4.	VFB Stuttgart	57
5.	Hamburger SV	51
6.	VFL Wolfsburg	51
7.	Hertha BSC Berlin	51
8.	Bayer Leverkusen	51
9.	Borussia Dortmund	51
10.	Hannover 96	48
11.	Eintracht Frankfurt	48
12.	VFL Bochum	48
13.	Energie Cottbus	45
14.	Arminia Bielefeld	45
15.	1.FC Köln (N)	45
16.	Karlsruher SC	45
17.	Mönchengladbach (N)	42
18.	1899 Hoffenheim (N)	42

BestOF1963_J48

Platz	Vereinsname	Punkte
1.	FC Schalke 04	69
2.	Bayern München (M)	68
3.	Hamburger SV	63
4.	Borussia Dortmund	61
5.	Bayer Leverkusen	60
6.	VFL Wolfsburg	58
7.	Werder Bremen	56
8.	VFB Stuttgart	55
9.	Hertha BSC Berlin	54
10.	Mönchengladbach (N)	47
11.	1899 Hoffenheim (N)	45
12.	Arminia Bielefeld	42
13.	Hannover 96	40
14.	Karlsruher SC	40
15.	VFL Bochum	37
16.	Energie Cottbus	32
17.	1.FC Köln (N)	32
18.	Eintracht Frankfurt	31

BestOF1963_SVM

Bayern München (M) 102 Punkte, restlichen Vereine 48 Punkte

Komplett1963_SVM

Alle Vereine – 51 Punkte

Komplett1963_RandomForest

Platz	Vereinsname	Punkte
1.	Bayern München (M)	84
2.	FC Schalke 04	66
3.	Bayer Leverkusen	64
4.	VFB Stuttgart	63
5.	Werder Bremen	60
6.	Hamburger SV	55
7.	VFL Wolfsburg	51
8.	Borussia Dortmund	50
9.	Hertha BSC Berlin	44
10.	VFL Bochum	44
11.	Mönchengladbach (N)	42
12.	1899 Hoffenheim (N)	39
13.	1.FC Köln (N)	37
14.	Eintracht Frankfurt	37
15.	Hannover 96	36
16.	Arminia Bielefeld	35
17.	Karlsruher SC	33
18.	Energie Cottbus	31

Anhang I: Kapitel 5.6 - Vorhersage der Saison 2009/2010

Hier finden Sie die Vorhersage des *BestOf_3S_Naive Bayes* Datensatzes, bei den ersten 7 Spieltagen sind die Ergebnisse bereits eingetragen, die Übereinstimmungen in den Vorhersagen und Ergebnissen fettgedruckt und in einer Zeile Ergebnis abschließend gezählt. Bei den restlichen Spieltagen wird dem Leser die Möglichkeit zum Selbsteintragen gegeben.

Prozentverteilung VH (=Vorhersage) bedeutet wie die Prozentverteilung unter den drei Möglichkeiten der Tendenzvorhersage sich verteilt hat (0-Unentschieden, 1-Heimsieg, 2-Auswärtssieg; dabei ist die ausgewählte Variante fettgedruckt, das abweichende Ergebnis, bis zum inklusive 7.Spieltag, unterstrichen).

Spieltag 1 Saison 2009/2010 (Datum 7.8.2009-9.8.2009)				Prozentverteilung VH		
Nr.	Spiel	Vorhersage	Ergebnis	1	0	2
1	VfL Wolfsburg (M) - VFB Stuttgart	1	1	88,5	11,3	0,2
2	Borussia Dortmund - 1. FC Köln	1	1	99,9	0,1	0
3	1. FC Nürnberg (N) - FC Schalke 04	2	2	0	0	100
4	Werder Bremen - Eintracht Frankfurt	1	2	100	0	<u>0</u>
5	Hertha BSC Berlin - Hannover 96	1	1	100	0	0
6	FSV Mainz 05 (N) - Bayer Leverkusen	2	0	0	<u>0</u>	100
7	1899 Hoffenheim - Bayern München	2	0	0	<u>0,2</u>	99,8
8	VfL Bochum - Mönchengladbach	0	0	47,8	48,7	3,5
9	SC Freiburg (N) - Hamburger SV	2	0	0	<u>0</u>	100
Ergebnis: 5/9						
Spieltag 2 Saison 2009/2010 (Datum 15.8.2009-16.8.2009)				Prozentverteilung VH		
Nr.	Spiel	Vorhersage	Ergebnis	1	0	2
10	Bayern München - Werder Bremen	1	0	91,5	<u>8,5</u>	0
11	VFB Stuttgart - SC Freiburg (N)	1	1	100	0	0
12	Hamburger SV - Borussia Dortmund	1	1	82,2	17	0,8
13	Bayer Leverkusen - 1899 Hoffenheim	1	1	56,5	43,1	0,4
14	Hannover 96 - FSV Mainz 05 (N)	1	0	67,8	<u>32,2</u>	0
15	Eintracht Frankfurt - 1. FC Nürnberg(N)	1	0	69,1	<u>30,2</u>	0,7
16	1. FC Köln - VfL Wolfsburg (M)	2	2	0	1,1	98,9
17	Mönchengladbach - Hertha BSC Berlin	2	1	<u>0</u>	6,8	93,2
18	FC Schalke 04 - VfL Bochum	1	1	99,9	0,1	0
Ergebnis: 5/9						
Spieltag 3 Saison 2009/2010 (Datum 21.8.2009-23.8.2009)				Prozentverteilung VH		
Nr.	Spiel	Vorhersage	Ergebnis	1	0	2
19	1899 Hoffenheim - FC Schalke 04	2	0	0	<u>5,4</u>	94,6
20	Borussia Dortmund - VFB Stuttgart	2	0	0,6	<u>24,9</u>	74,6
21	1. FC Köln - Eintracht Frankfurt	0	0	16,7	78,7	4,6
22	SC Freiburg (N) - Bayer Leverkusen	2	2	0	0,2	99,8
23	FSV Mainz 05 (N) - Bayern München	2	1	<u>0</u>	0	100
24	1. FC Nürnberg (N) - Hannover 96	0	2	1,1	54,3	<u>44,6</u>
25	VfL Bochum - Hertha BSC Berlin	2	1	<u>0</u>	6,4	93,5
26	VfL Wolfsburg (M) - Hamburger SV	1	2	75,3	24,5	<u>0,2</u>
27	Werder Bremen - Mönchengladbach	1	1	100	0	0
Ergebnis 3/9						

Spieltag 4 Saison 2009/2010 (Datum 28.8.2009-30.8.2009)				Prozentverteilung VH		
Nr.	Spiel	VH	Ergebnis	1	0	2
28	Mönchengladbach - FSV Mainz 05 (N)	0	1	<u>18,7</u>	80	1,3
29	VFB Stuttgart - 1. FC Nürnberg (N)	1	0	100	<u>0</u>	0
30	FC Schalke 04 - SC Freiburg (N)	1	2	100	0	<u>0</u>
31	Bayer Leverkusen - VFL Bochum	1	1	99,8	0,2	0
32	Hannover 96 - 1899 Hoffenheim	0	2	13,4	81,4	5,2
33	Eintracht Frankfurt - Borussia Dortmund	2	0	0	<u>3,8</u>	96,2
34	Bayern München - VFL Wolfsburg (M)	1	1	91,7	8,2	0,1
35	Hertha BSC Berlin - Werder Bremen	2	2	0,1	5,2	94,7
36	Hamburger SV - 1. FC Köln	1	1	100	0	0
Ergebnis 4/9						
Spieltag 5 Saison 2009/2010 (Datum 12.9.2009-13.9.2009)				Prozentverteilung VH		
Nr.	Spiel	VH	Ergebnis	1	0	2
37	VFL Wolfsburg (M) - Bayer Leverkusen	1	2	81,3	15,6	<u>3,2</u>
38	Borussia Dortmund - Bayern München	2	2	0	1,3	98,7
39	1899 Hoffenheim - VFL Bochum	1	1	82,6	17,3	0,1
40	SC Freiburg (N) - Eintracht Frankfurt	0	2	0,4	52,7	<u>46,9</u>
41	FSV Mainz 05 (N) - Hertha BSC Berlin	2	1	<u>0</u>	5,6	94,4
42	1. FC Nürnberg (N) - Mönchengladbach	0	1	<u>2,8</u>	54,8	42,4
43	Hamburger SV - VFB Stuttgart	1	1	96,1	3,8	0,1
44	Werder Bremen - Hannover 96	1	0	100	<u>0</u>	0
45	1. FC Köln - FC Schalke 04	2	2	0	1,1	98,9
Ergebnis 4/9						
Spieltag 6 Saison 2009/2010 (Datum 18.9.2009-20.9.2009)				Prozentverteilung VH		
Nr.	Spiel	VH	Ergebnis	1	0	2
46	Bayern München - 1. FC Nürnberg (N)	1	1	100	0	0
47	VFB Stuttgart - 1. FC Köln	1	2	100	0	<u>0</u>
48	Hertha BSC Berlin - SC Freiburg (N)	1	2	99,7	0,3	<u>0</u>
49	FC Schalke 04 - VFL Wolfsburg (M)	0	2	41,5	52,3	<u>6,3</u>
50	Bayer Leverkusen - Werder Bremen	2	0	2,7	<u>39,5</u>	57,8
51	Hannover 96 - Borussia Dortmund	2	0	0,4	<u>32,7</u>	66,9
52	Eintracht Frankfurt - Hamburger SV	2	0	0	<u>0,2</u>	99,8
53	VFL Bochum - FSV Mainz 05 (N)	0	2	28,9	70,2	<u>0,8</u>
54	Mönchengladbach - 1899 Hoffenheim	0	2	1,1	64,2	<u>34,7</u>
Ergebnis 1/9						
Spieltag 7 Saison 2009/2010 (Datum 25.9.2009-27.9.2009)				Prozentverteilung VH		
Nr.	Spiel	VH	Ergebnis	1	0	2
55	VFL Wolfsburg (M) - Hannover 96	1	1	99,9	0,1	0
56	Hamburger SV - Bayern München	0	1	<u>26,1</u>	49,9	24
57	Borussia Dortmund - FC Schalke 04	2	2	0,3	27,2	72,5
58	1899 Hoffenheim - Hertha BSC Berlin	0	1	<u>17,8</u>	63,9	18,3
59	Werder Bremen - FSV Mainz 05 (N)	1	1	0	100	0
60	1. FC Köln - Bayer Leverkusen	2	2	0,6	0	99,4
61	Eintracht Frankfurt - VFB Stuttgart	2	2	0,5	0	99,5
62	SC Freiburg (N) - Mönchengladbach	2	1	<u>0,7</u>	27,2	72,2
63	1. FC Nürnberg (N) - VFL Bochum	0	2	1,7	77,1	21,3
Ergebnis 5/9						

Spieltag 8 Saison 2009/2010 (Datum 2.10.2009-4.10.2009)				Prozentverteilung VH		
Nr.	Spiel	VH	Ergebnis	1	0	2
64	Bayern München - 1. FC Köln	1		100	0	0
65	VFB Stuttgart - Werder Bremen	0		29,9	49,7	20,4
66	Hertha BSC Berlin - Hamburger SV	2		0	2,9	97,1
67	FC Schalke 04 - Eintracht Frankfurt	1		99,9	0,1	0
68	Bayer Leverkusen - 1. FC Nürnberg (N)	1		99,9	0,1	0
69	Hannover 96 - SC Freiburg (N)	1		85,1	14,9	0
70	VFL Bochum - VFL Wolfsburg (M)	2		0	0,7	99,3
71	Mönchengladbach - Borussia Dortmund	2		0	5,4	94,6
72	FSV Mainz 05 (N) - 1899 Hoffenheim	2		0,1	13,3	86,6
Ergebnis						
Spieltag 9 Saison 2009/2010 (Datum 17.10.2009-18.10.2009)				Prozentverteilung VH		
Nr.	Spiel	VH	Ergebnis	1	0	2
73	VFL Wolfsburg (M) - Mönchengladbach	1		100	0	0
74	VFB Stuttgart - FC Schalke 04	1		48,5	43,2	8,3
75	Hamburger SV - Bayer Leverkusen	1		98,8	1,2	0
76	Borussia Dortmund - VFL Bochum	1		96,3	3,7	0
77	Werder Bremen - 1899 Hoffenheim	1		99,8	0,2	0
78	1. FC Köln - FSV Mainz 05 (N)	1		51,4	48,3	0,3
79	Eintracht Frankfurt - Hannover 96	2		4,6	31,5	63,9
80	SC Freiburg (N) - Bayern München	2		0	0	100
81	1. FC Nürnberg (N) - Hertha BSC Berlin	2		0	9,4	90,5
Ergebnis						
Spieltag 10 Saison 2009/2010 (Datum 23.10.2009-25.10.2009)				Prozentverteilung VH		
Nr.	Spiel	VH	Ergebnis	1	0	2
82	Bayern München - Eintracht Frankfurt	1		100	0	0
83	Hertha BSC Berlin - VFL Wolfsburg (M)	2		0,4	12,2	87,5
84	1899 Hoffenheim - 1. FC Nürnberg (N)	1		98,4	1,7	0
85	FC Schalke 04 - Hamburger SV	0		7,1	50,4	42,5
86	Bayer Leverkusen - Borussia Dortmund	1		63,8	31,7	4,6
87	Hannover 96 - VFB Stuttgart	2		0	4	96
88	VFL Bochum - Werder Bremen	2		0	0,1	99,9
89	Mönchengladbach - 1. FC Köln	0		26,1	38,9	35
90	FSV Mainz 05 (N) - SC Freiburg (N)	0		18,9	76,8	4,3
Ergebnis						
Spieltag 11 Saison 2009/2010 (Datum ca 31.10.2009)				Prozentverteilung VH		
Nr.	Spiel	VH	Ergebnis	1	0	2
91	VFL Wolfsburg (M) - FSV Mainz 05 (N)	1		100	0	0
92	VFB Stuttgart - Bayern München	2		0,3	9,3	90,3
93	Hamburger SV - Mönchengladbach	1		100	0	0
94	Borussia Dortmund - Hertha BSC Berlin	0		31,9	65,6	2,5
95	FC Schalke 04 - Bayer Leverkusen	1		62,8	23,3	13,9
96	1. FC Köln - Hannover 96	0		5,5	52,8	41,7
97	Eintracht Frankfurt - VFL Bochum	0		17,7	65,2	17
98	SC Freiburg (N) - 1899 Hoffenheim	2		0	10,1	89,9
99	1. FC Nürnberg (N) - Werder Bremen	2		0	0,1	99,9
Ergebnis						

Spieltag 12 Saison 2009/2010 (Datum ca 07.11.2009)				Prozentverteilung VH		
Nr.	Spiel	VH	Ergebnis	1	0	2
100	Bayern München - FC Schalke 04	1		96,6	3,4	0
101	Hertha BSC Berlin - 1. FC Köln	1		95,4	4,3	0,2
102	1899 Hoffenheim - VFL Wolfsburg (M)	0		0,1	18,6	81,3
103	Bayer Leverkusen - Eintracht Frankfurt	1		98,3	1,7	0
104	Werder Bremen - Borussia Dortmund	1		99	0,9	0
105	Hannover 96 - Hamburger SV	2		0	2,8	97,2
106	VFL Bochum - SC Freiburg (N)	0		49,3	50,3	0,4
107	Mönchengladbach - VFB Stuttgart	2		0	0,6	99,4
108	FSV Mainz 05 (N) - 1. FC Nürnberg (N)	0		18,6	71,3	10,1
Ergebnis						
Spieltag 13 Saison 2009/2010 (Datum ca 21.11.2009)				Prozentverteilung VH		
Nr.	Spiel	VH	Ergebnis	1	0	2
109	VFL Wolfsburg (M) - 1. FC Nürnberg (N)	1		100	0	0
110	Bayern München - Bayer Leverkusen	1		99,7	0,3	0
111	VFB Stuttgart - Hertha BSC Berlin	1		97,5	2,5	0
112	Hamburger SV - VFL Bochum	1		100	0	0
113	Borussia Dortmund - FSV Mainz 05 (N)	1		99,1	0,9	0
114	FC Schalke 04 - Hannover 96	1		99,9	0,1	0
115	1. FC Köln - 1899 Hoffenheim	0		0,9	52,6	46,5
116	Eintracht Frankfurt - Mönchengladbach	0		15,8	61	23,2
117	SC Freiburg (N) - Werder Bremen	2		0	0	100
Ergebnis						
Spieltag 14 Saison 2009/2010 (Datum ca 28.11.2009)				Prozentverteilung VH		
Nr.	Spiel	VH	Ergebnis	1	0	2
118	Hertha BSC Berlin - Eintracht Frankfurt	1		95,7	4,2	0
119	1899 Hoffenheim - Borussia Dortmund	0		3,6	64,4	32
120	Bayer Leverkusen - VFB Stuttgart	2		10,3	31,1	58,5
121	Werder Bremen - VFL Wolfsburg (M)	1		80	18,4	1,6
122	Hannover 96 - Bayern München	2		0	0,1	99,9
123	VFL Bochum - 1. FC Köln	2		5,5	43	51,5
124	Mönchengladbach - FC Schalke 04	2		0	1,1	98,9
125	FSV Mainz 05 (N) - Hamburger SV	2		0	0	100
126	1. FC Nürnberg (N) - SC Freiburg (N)	0		12,7	85,3	2
Ergebnis						
Spieltag 15 Saison 2009/2010 (Datum ca 5.12.2009)				Prozentverteilung VH		
Nr.	Spiel	VH	Ergebnis	1	0	2
127	VFL Wolfsburg (M) - SC Freiburg (N)	1		100	0	0
128	Bayern München - Mönchengladbach	1		100	0	0
129	VFB Stuttgart - VFL Bochum	1		100	0	0
130	Hamburger SV - 1899 Hoffenheim	1		99,9	0,1	0
131	Borussia Dortmund - 1. FC Nürnberg (N)	1		99,6	0,4	0
132	FC Schalke 04 - Hertha BSC Berlin	1		98	2	0
133	Hannover 96 - Bayer Leverkusen	2		0	4,7	95,3
134	1. FC Köln - Werder Bremen	2		0	0,4	99,6
135	Eintracht Frankfurt - FSV Mainz 05 (N)	0		43,4	56,2	0,4
Ergebnis						

Spieltag 16 Saison 2009/2010 (Datum ca 12.12.2009)				Prozentverteilung VH		
Nr.	Spiel	VH	Ergebnis	1	0	2
136	VFL Wolfsburg (M) - Borussia Dortmund	1		84,4	15,4	0,2
137	Hertha BSC Berlin - Bayer Leverkusen	2		0,9	5,1	94,1
138	1899 Hoffenheim - Eintracht Frankfurt	1		70,6	29,3	0,1
139	Werder Bremen - FC Schalke 04	1		82,7	15,4	1,8
140	VFL Bochum - Bayern München	2		0	0	100
141	Mönchengladbach - Hannover 96	0		12,5	49,4	38,1
142	SC Freiburg (N) - 1. FC Köln	2		0,4	19,1	80,5
143	FSV Mainz 05 (N) - VFB Stuttgart	2		0	0,1	99,9
144	1. FC Nürnberg (N) - Hamburger SV	2		0	0,1	99,9
Ergebnis						
Spieltag 17 Saison 2009/2010 (Datum ca 19.12.2009)				Prozentverteilung VH		
Nr.	Spiel	VH	Ergebnis	1	0	2
145	Bayern München - Hertha BSC Berlin	1		99,9	0,1	0
146	VFB Stuttgart - 1899 Hoffenheim	1		97,7	2,3	0
147	Hamburger SV - Werder Bremen	1		75,9	23,9	0,2
148	Borussia Dortmund - SC Freiburg (N)	1		99,7	0,3	0
149	FC Schalke 04 - FSV Mainz 05 (N)	1		100	0	0
150	Bayer Leverkusen - Mönchengladbach	1		99,5	0,5	0
151	Hannover 96 - VFL Bochum	0		42,7	56,8	0,5
152	1. FC Köln - 1. FC Nürnberg (N)	1		54,4	45,2	0,4
153	Eintracht Frankfurt - VFL Wolfsburg (M)	2		0	1	99
Ergebnis						

Ende der Hinrunde

Spieltag 18 Saison 2009/2010 (Datum ca 16.1.2010)				Prozentverteilung VH		
Nr.	Spiel	VH	Ergebnis	1	0	2
154	VFB Stuttgart - VFL Wolfsburg (M)	0		44	50,7	5,3
155	1. FC Köln - Borussia Dortmund	2		0	4,6	95,4
156	FC Schalke 04 - 1. FC Nürnberg (N)	1		100	0	0
157	Hamburger SV - SC Freiburg (N)	1		100	0	0
158	Eintracht Frankfurt - Werder Bremen	2		0	0,2	99,8
159	Hannover 96 - Hertha BSC Berlin	0		3	72,6	24,4
160	Bayer Leverkusen - FSV Mainz 05 (N)	1		99,8	0,2	0
161	Bayern München - 1899 Hoffenheim	1		99,9	0,1	0
162	Mönchengladbach - VFL Bochum	0		22,3	68,7	9
Ergebnis						
Spieltag 19 Saison 2009/2010 (Datum ca 23.1.2010)				Prozentverteilung VH		
Nr.	Spiel	VH	Ergebnis	1	0	2
163	Werder Bremen - Bayern München	2		5,9	25,4	68,7
164	SC Freiburg (N) - VFB Stuttgart	2		0	0	100
165	Borussia Dortmund - Hamburger SV	2		0,3	23,4	76,3
166	VFL Bochum - FC Schalke 04	2		0	0,4	99,6
167	1899 Hoffenheim - Bayer Leverkusen	2		0,3	8,9	90,8
168	FSV Mainz 05 (N) - Hannover 96	2		0,3	15,6	84,1
169	VFL Wolfsburg (M) - 1. FC Köln	1		100	0	0
170	1. FC Nürnberg (N) - Eintracht Frankfurt	0		0,8	73,5	25,6
171	Hertha BSC Berlin - Mönchengladbach	1		97,5	2,4	0,1
Ergebnis						

Spieltag 20 Saison 2009/2010 (Datum ca 30.1.2010)				Prozentverteilung VH		
Nr.	Spiel	VH	Ergebnis	1	0	2
172	Hamburger SV - VFL Wolfsburg (M)	1		92,8	7,1	0
173	VFB Stuttgart - Borussia Dortmund	1		92,4	7,5	0,1
174	FC Schalke 04 - 1899 Hoffenheim	1		98	2	0
175	Mönchengladbach - Werder Bremen	2		0	0,5	99,5
176	Eintracht Frankfurt - 1. FC Köln	2		16,4	29,7	53,9
177	Hertha BSC Berlin - VFL Bochum	1		92,8	7,1	0,1
178	Bayer Leverkusen - SC Freiburg (N)	1		99,9	0,1	0
179	Bayern München - FSV Mainz 05 (N)	1		100	0	0
180	Hannover 96 - 1. FC Nürnberg (N)	1		79	21	0
Ergebnis						
Spieltag 21 Saison 2009/2010 (Datum ca 6.2.2010)				Prozentverteilung VH		
Nr.	Spiel	VH	Ergebnis	1	0	2
181	VFL Wolfsburg (M) - Bayern München	2		0,7	17,7	81,5
182	1. FC Nürnberg (N) - VFB Stuttgart	2		0	0,2	99,8
183	Werder Bremen - Hertha BSC Berlin	1		99,9	0,1	0
184	1. FC Köln - Hamburger SV	2		0	0,3	99,7
185	SC Freiburg (N) - FC Schalke 04	2		0	0,1	99,9
186	VFL Bochum - Bayer Leverkusen	2		0	0,4	99,6
187	1899 Hoffenheim - Hannover 96	1		65,3	33,4	1,3
188	Borussia Dortmund - Eintracht Frankfurt	1		93,7	6,3	0
189	FSV Mainz 05 (N) - Mönchengladbach	2		1,7	27,4	70,9
Ergebnis						
Spieltag 22 Saison 2009/2010 (Datum ca 13.2.2010)				Prozentverteilung VH		
Nr.	Spiel	VH	Ergebnis	1	0	2
190	Bayer Leverkusen - VFL Wolfsburg (M)	0		6,9	61,3	31,8
191	VFB Stuttgart - Hamburger SV	0		26	52	22
192	Bayern München - Borussia Dortmund	1		99,7	0,3	0
193	VFL Bochum - 1899 Hoffenheim	0		0,7	51,4	47,8
194	Hannover 96 - Werder Bremen	2		0	0,2	0,98
195	FC Schalke 04 - 1. FC Köln	1		99,9	0,1	0
196	Eintracht Frankfurt - SC Freiburg (N)	1		63,3	35,9	0,8
197	Hertha BSC Berlin - FSV Mainz 05 (N)	1		98,8	1,2	0
198	Mönchengladbach - 1. FC Nürnberg (N)	1		69	30,6	0,5
Ergebnis						
Spieltag 23 Saison 2009/2010 (Datum ca 20.2.2010)				Prozentverteilung VH		
Nr.	Spiel	VH	Ergebnis	1	0	2
199	1. FC Nürnberg (N) - Bayern München	2		0	0	100
200	1. FC Köln - VFB Stuttgart	2		0	0,5	99,5
201	SC Freiburg (N) - Hertha BSC Berlin	2		0	5,1	94,9
202	VFL Wolfsburg (M) - FC Schalke 04	1		50,9	42,3	6,8
203	Werder Bremen - Bayer Leverkusen	1		97,4	2,1	0,4
204	Borussia Dortmund - Hannover 96	1		90,4	9,5	0,1
205	Hamburger SV - Eintracht Frankfurt	1		100	0	0
206	FSV Mainz 05 (N) - VFL Bochum	0		2,9	63	34,1
207	1899 Hoffenheim - Mönchengladbach	1		85,9	13,8	0,2
Ergebnis						

Spieltag 24 Saison 2009/2010 (Datum ca 27.2.2010)				Prozentverteilung VH		
Nr.	Spiel	VH	Ergebnis	1	0	2
208	Hannover 96 - VFL Wolfsburg (M)	2		0	7,2	92,8
209	Bayern München - Hamburger SV	1		84,4	15,3	0,3
210	FC Schalke 04 - Borussia Dortmund	1		92	7,8	0,2
211	Hertha BSC Berlin - 1899 Hoffenheim	0		38,3	55,6	6,1
212	FSV Mainz 05 (N) - Werder Bremen	2		0	0,1	99,9
213	Bayer Leverkusen - 1. FC Köln	1		99,7	0,3	0
214	VFB Stuttgart - Eintracht Frankfurt	1		99,9	0,1	0
215	Mönchengladbach - SC Freiburg (N)	1		72,2	27,5	0,3
216	VFL Bochum - 1. FC Nürnberg (N)	0		45,2	54,2	0,6
Ergebnis						
Spieltag 25 Saison 2009/2010 (Datum ca 6.3.2010)				Prozentverteilung VH		
Nr.	Spiel	VH	Ergebnis	1	0	2
217	1. FC Köln - Bayern München	2		0	0	100
218	Werder Bremen - VFB Stuttgart	1		89,8	7,5	2,8
219	Hamburger SV - Hertha BSC Berlin	1		99,9	0,1	0
220	Eintracht Frankfurt - FC Schalke 04	2		0	0,5	99,5
221	1. FC Nürnberg (N) - Bayer Leverkusen	2		0	0,2	99,8
222	SC Freiburg (N) - Hannover 96	2		0,1	9,6	90,3
223	VFL Wolfsburg (M) - VFL Bochum	1		100	0	0
224	Borussia Dortmund - Mönchengladbach	1		96,9	3,1	0
225	1899 Hoffenheim - FSV Mainz 05 (N)	1		95,1	4,9	0
Ergebnis						
Spieltag 26 Saison 2009/2010 (Datum ca 13.3.2010)				Prozentverteilung VH		
Nr.	Spiel	VH	Ergebnis	1	0	2
226	Mönchengladbach - VFL Wolfsburg (M)	2		0	2,8	97,2
227	FC Schalke 04 - VFB Stuttgart	0		40	41,1	18,9
228	Bayer Leverkusen - Hamburger SV	2		1,2	29,5	69,2
229	VFL Bochum - Borussia Dortmund	2		0	4,9	95,1
230	1899 Hoffenheim - Werder Bremen	2		0	5,9	94,1
231	FSV Mainz 05 (N) - 1. FC Köln	2		2,1	13,6	84,4
232	Hannover 96 - Eintracht Frankfurt	0		32,5	66,8	0,7
233	Bayern München - SC Freiburg (N)	1		100	0	0
234	Hertha BSC Berlin - 1. FC Nürnberg (N)	1		99,2	0,8	0
Ergebnis						
Spieltag 27 Saison 2009/2010 (Datum ca 20.3.2010)				Prozentverteilung VH		
Nr.	Spiel	VH	Ergebnis	1	0	2
235	Eintracht Frankfurt - Bayern München	2		0	0	100
236	VFL Wolfsburg (M) - Hertha BSC Berlin	1		98,2	1,8	0
237	1. FC Nürnberg (N) - 1899 Hoffenheim	2		0,1	36,7	63,1
238	Hamburger SV - FC Schalke 04	1		91,1	8,9	0
239	Borussia Dortmund - Bayer Leverkusen	2		3,3	26,5	70,1
240	VFB Stuttgart - Hannover 96	1		99,9	0,1	0
241	Werder Bremen - VFL Bochum	1		100	0	0
242	1. FC Köln - Mönchengladbach	0		15,5	59,4	25,1
243	SC Freiburg (N) - FSV Mainz 05 (N)	0		3,4	90,1	6,5
Ergebnis						

Spieltag 28 Saison 2009/2010 (Datum ca 27.3.2010)				Prozentverteilung VH		
Nr.	Spiel	VH	Ergebnis	1	0	2
244	FSV Mainz 05 (N) - VFL Wolfsburg (M)	2		0	0,2	99,8
245	Bayern München - VFB Stuttgart	1		96,7	3,3	0
246	Mönchengladbach - Hamburger SV	2		0	0,6	99,4
247	Hertha BSC Berlin - Borussia Dortmund	2		4,4	36,5	59,1
248	Bayer Leverkusen - FC Schalke 04	0		8,7	59,6	31,7
249	Hannover 96 - 1. FC Köln	1		56,8	40,8	2,4
250	VFL Bochum - Eintracht Frankfurt	0		7,9	82,5	9,6
251	1899 Hoffenheim - SC Freiburg (N)	1		97,1	2,9	0
252	Werder Bremen - 1. FC Nürnberg (N)	1		100	0	0
Ergebnis						
Spieltag 29 Saison 2009/2010 (Datum ca 3.4.2010)				Prozentverteilung VH		
Nr.	Spiel	VH	Ergebnis	1	0	2
253	FC Schalke 04 - Bayern München	2		0,9	13,2	85,9
254	1. FC Köln - Hertha BSC Berlin	2		0,6	36,1	63,3
255	VFL Wolfsburg (M) - 1899 Hoffenheim	1		97	3	0
256	Eintracht Frankfurt - Bayer Leverkusen	2		0	0,6	99,4
257	Borussia Dortmund - Werder Bremen	2		0,5	24	75,5
258	Hamburger SV - Hannover 96	1		100	0	0
259	SC Freiburg (N) - VFL Bochum	2		0,6	48,1	51,3
260	VFB Stuttgart - Mönchengladbach	1		100	0	0
261	1. FC Nürnberg (N) - FSV Mainz 05 (N)	0		8,1	90,1	1,8
Ergebnis						
Spieltag 30 Saison 2009/2010 (Datum ca 10.4.2010)				Prozentverteilung VH		
Nr.	Spiel	VH	Ergebnis	1	0	2
262	1. FC Nürnberg (N) - VFL Wolfsburg (M)	2		0	0,7	99,3
263	Bayer Leverkusen - Bayern München	2		0	1,6	98,4
264	Hertha BSC Berlin - VFB Stuttgart	2		0,2	5	94,8
265	VFL Bochum - Hamburger SV	2		0	0,2	99,8
266	FSV Mainz 05 (N) - Borussia Dortmund	2		0	1,8	98,2
267	Hannover 96 - FC Schalke 04	2		0	4,3	95,7
268	1899 Hoffenheim - 1. FC Köln	1		91,2	8,3	0,5
269	Mönchengladbach - Eintracht Frankfurt	0		11,1	86	2,9
270	Werder Bremen - SC Freiburg (N)	1		100	0	0
Ergebnis						
Spieltag 31 Saison 2009/2010 (Datum ca 10.4.2010)				Prozentverteilung VH		
Nr.	Spiel	VH	Ergebnis	1	0	2
271	Eintracht Frankfurt - Hertha BSC Berlin	2		0,2	9	90,8
272	Borussia Dortmund - 1899 Hoffenheim	0		34,1	64,7	1,1
273	VFB Stuttgart - Bayer Leverkusen	1		74,8	17,6	7,6
274	VFL Wolfsburg (M) - Werder Bremen	0		24,9	60,1	15,1
275	Bayern München - Hannover 96	1		100	0	0
276	1. FC Köln - VFL Bochum	0		20	75,3	4,7
277	FC Schalke 04 - Mönchengladbach	1		99,9	0,1	0
278	Hamburger SV - FSV Mainz 05 (N)	1		100	0	0
279	SC Freiburg (N) - 1. FC Nürnberg (N)	0		5,2	88,2	6,6
Ergebnis						

Spieltag 32 Saison 2009/2010 (Datum ca 24.4.2010)				Prozentverteilung VH		
Nr.	Spiel	VH	Ergebnis	1	0	2
280	SC Freiburg (N) - VFL Wolfsburg (M)	2		0	0,2	99,8
281	Mönchengladbach - Bayern München	2		0	0,1	99,9
282	VFL Bochum - VFB Stuttgart	2		0	0,3	99,7
283	1899 Hoffenheim - Hamburger SV	2		0	5	95
284	1. FC Nürnberg (N) - Borussia Dortmund	2		0	2,9	97,1
285	Hertha BSC Berlin - FC Schalke 04	2		0,1	5,4	94,5
286	Bayer Leverkusen - Hannover 96	1		98,5	1,5	0
287	Werder Bremen - 1. FC Köln	1		100	0	0
288	FSV Mainz 05 (N) - Eintracht Frankfurt	0		1,3	61,1	37,6
Ergebnis						
Spieltag 33 Saison 2009/2010 (Datum ca 1/2.5.2010)				Prozentverteilung VH		
Nr.	Spiel	VH	Ergebnis	1	0	2
289	Borussia Dortmund - VFL Wolfsburg (M)	0		1	57	41,9
290	Bayer Leverkusen - Hertha BSC Berlin	1		90,7	8,9	0,4
291	Eintracht Frankfurt - 1899 Hoffenheim	0		1,1	57,4	41,5
292	FC Schalke 04 - Werder Bremen	0		12,6	47,9	39,5
293	Bayern München - VFL Bochum	1		100	0	0
294	Hannover 96 - Mönchengladbach	0		46	51,1	2,9
296	1. FC Köln - SC Freiburg (N)	1		53,1	46,6	0,3
296	VFB Stuttgart - FSV Mainz 05 (N)	1		100	0	0
297	Hamburger SV - 1. FC Nürnberg (N)	1		100	0	0
Ergebnis						
Spieltag 34 Saison 2009/2010 (Datum ca 8.5.2010)				Prozentverteilung VH		
Nr.	Spiel	VH	Ergebnis	1	0	2
298	Hertha BSC Berlin - Bayern München	2		0	0,2	99,8
299	1899 Hoffenheim - VFB Stuttgart	2		0,1	8,7	91,2
300	Werder Bremen - Hamburger SV	1		52,5	37,9	9,7
301	SC Freiburg (N) - Borussia Dortmund	2		0	0,9	99,1
302	FSV Mainz 05 (N) - FC Schalke 04	2		0	0,1	99,9
303	Mönchengladbach - Bayer Leverkusen	2		0	0,9	99,1
304	VFL Bochum - Hannover 96	2		2,5	39,3	58,2
305	1. FC Nürnberg (N) - 1. FC Köln	0		2,2	49,1	48,7
306	VFL Wolfsburg (M) - Eintracht Frankfurt	1		99,9	0,1	0
Ergebnis						

Ende der Saison

Endtabelle Saisonvorhersage 2009/2010

Platz	Vereinsname	Punkte
1.	FC Bayern München	100
2.	Hamburger SV	87
-.	Werder Bremen	87
4.	VFL Wolfsburg (M)	80
5.	VFB Stuttgart	79
6.	Schalke 04	74
7.	Bayer Leverkusen	68
8.	Borussia Dortmund	55
9.	Hertha BSC Berlin	49
10.	1899 Hoffenheim	42
11.	Hannover 96	32
12.	1. FC Köln	28
13.	Mönchengladbach	23
14.	Eintracht Frankfurt	17
15.	VFL Bochum	15
16.	1.FC Nürnberg (N)	10
17.	FSV Mainz 05 (N)	9
18.	SC Freiburg (N)	6

Tabelle I.1: Endtabelle Saisonvorhersage 2009/2010