

NBA Game Prediction

Projektarbeit

B. Kühnis

Dezember 31, 2017

Advisors: Prof. Dr. Farhad D. Mehta

Abteilung Informatik, IFS, HSR

Zusammenfassung

Diese Arbeit beschäftigt sich mit der Frage, wie sehenswert ein NBA-Basketballspiel ist. Pro Tag gibt es, während der Saison, mehrere Spiele, welche ausgetragen werden. Da die Spiele in den USA gespielt werden, können wegen der Zeitverschiebung, oder durch soziale Verpflichtungen, die Spiele oft nicht live mitverfolgt werden. Wenn nun eine Person ein Spiel sehen möchte, hat diese mehrere Spiele zur Auswahl, wobei nicht klar ist, welches das sehenswerteste Spiel ist. Eine Wertung soll dem Benutzer helfen, das spannendste Spiel auszuwählen.

Dazu wurden verschiedene lineare Regressionsmodelle erstellt, wobei die Input-Werte die Statistiken über ein Spiel sind und die Resultat-Werte, welche die Sehenswürdigkeit beschreiben, unterschiedlich gewählt wurden.

NBA-Subreddit erstellt seit etwa zwei Jahren pro Spiel einen Thread, wobei die Benutzer diesen bewerten können. Dies hat zur Folge, dass jedes Spiel eine Wertung hat, welche abhängig von der Bewertung der Benutzer ist. Diese Wertung wird bei einem Modell als Resultat-Wert verwendet, mit der Idee, dass Threads und somit Spiele, mit einer hohen Wertung sehenswürdiger sind, als Spiele mit einer geringen Wertung.

Wikihoops ist eine Webseite die sich mit der gleichen Frage beschäftigt, wie diese Arbeit. Die Webseite bewertet Spiele mittels Game- und User-Rating. Game-Rating entsteht anhand einer Funktion. User-Ratings basieren auf den Game-Ratings, wobei Benutzer Wertungen zur Game-Rating abgeben können. Als Resultat-Werte für die Modelle wurden die Game-Rating und User-Ratings verwendet.

Die Arbeit fand keine eindeutige Antwort auf die Frage, wie sehenswert ein Spiel ist. Was damit zu begründen ist, dass die Sehenswürdigkeit suggestiv ist. Dies zeigt sich auch, wenn die Sehenswürdigkeit der Modelle mit den Resultat-Werten von Reddit und Wikihoops verglichen werden. Die Modelle ergeben eine unterschiedliche Sehenswürdigkeit für die gleichen Spiele.

Eine Möglichkeit, die Frage zu beantworten, ist der Weg, welchen Wikihoops eingeschlagen hat. Nämlich eine Funktion zu erstellen, welche ein Spiel bewertet und den Benutzer die Möglichkeit zu bieten, diese Wertung anzupassen. Was der Webseite jedoch fehlt, ist, die Wertungen der Benutzer in die Funktion einfließen zu lassen. Dies ist möglich, wie in dieser Arbeit gezeigt wird.

Inhaltsverzeichnis

Inhaltsverzeichnis	ii
1 Einleitung	1
1.1 Ausgangslage	1
1.2 Ziel der Arbeit	2
2 Problembeschreibung	3
2.1 Sehenswürdigkeit	3
2.2 Spoiler-Free	3
3 Lösungskonzept	5
3.1 Allgemeine Sehenswürdigkeit	5
3.1.1 Spielwertung durch die Benutzer	5
3.1.2 Sehenswürdigkeit Modell	6
3.1.3 Supervised Learning: Regression	6
3.2 Persönliche Sehenswürdigkeit	8
3.2.1 Unsupervised Learning: Kategorisierung	8
3.2.2 Persönliche Vorlieben	8
4 Umsetzung	9
4.1 Prototyp	9
4.1.1 Prototyp Sehenswürdigkeit Funktion	9
4.1.2 Resultat	10
4.2 Sehenswürdigkeit verbessern	12
4.2.1 Reddit-Ups Wertung	12
4.2.2 Wikihoops	14
4.2.3 Vergleich der Resultate	18
4.2.4 Schlussfolgerung	20
5 Ergebnis	22
5.1 Webseite	22
5.2 Offene Punkte	22
5.3 Ausblick	22

Kapitel 1

Einleitung

Das Projekt beschäftigt sich mit der Frage, wie sehenswert ein Basketballspiel ist. Dabei werden Spiele, die bereits beendet wurden, bewertet und deren Sehenswürdigkeit in einer Skala ausgegeben. Die Schwierigkeit der Fragestellung liegt in der Definition des Begriffs der Sehenswürdigkeit. Dieser ist subjektiv und somit gibt es für die Fragestellung keine allgemeingültige Antwort.

Die Arbeit beschäftigt sich neben der Analyse, wie sehenswert ein Basketball ist, auch mit der Frage, wie der Begriff der Sehenswürdigkeit in Bezug auf ein Basketballspiel überhaupt definiert werden kann.

1.1 Ausgangslage

Das Projekt wird von Grund auf neu entwickelt. Die Idee entstand durch das Überangebot an Spielen, welche an einem Tag durchgeführt werden. Zusätzlich kommt noch die geographische Komponente dazu. Diese hat zur Folge, dass wegen der Zeitverschiebung Spiele oft nicht live gesehen werden können. Zu einem späteren Zeitpunkt gibt es eine Auswahl an Spielen¹. Dabei ist es schwierig, das sehenswerteste Spiel auszuwählen. Die Auswahl kann einerseits durch die Vorlieben für ein oder mehrere Teams getroffen werden oder es werden die Resultate angeschaut, wobei dies die Spannung des Spieles mindert. Was also fehlt ist ein Richtwert, welcher die Sehenswürdigkeit des Spieles bestimmt².

¹Für gewisse Personen sicherlich nicht verständlich, warum ein Spiel, welches in der Vergangenheit liegt, noch angeschaut wird. Als Vergleich kann ein Basketballspiel mit einem Film verglichen werden: auch wenn der Film schon ein paar Tage alt ist, solange die Handlung nicht bekannt ist, kann dieser trotzdem spannend sein.

²Wenn das Basketballspiel wieder mit einem Film verglichen wird, so wäre der Richtwert wie die Kritik des Filmes.

1.2 Ziel der Arbeit

Das Ziel der Arbeit ist es, der Person, welche ein Basketballspiel schauen möchte, einen Richtwert zu geben, wie spannend dieses Spiel war, ohne dabei das Resultat zu verraten.

Kapitel 2

Problembeschreibung

NBA¹ Spiele werden in Nordamerika ausgetragen². Personen aus anderen Regionen können wegen der Zeitverschiebung die Spiele oft nicht live mitverfolgen. Wenn eine Person zu einem späteren Zeitpunkt ein Spiel sehen möchte, hat sie oft mehrere Spiele zur Auswahl. Das Ziel ist nun herauszufinden, welches das interessanteste Spiel war, ohne das Resultat vorher anzuschauen (Abschnitt 2.2). Eine Lösung dazu wäre, das Spiel des Lieblingsteams auszuwählen. Dabei ist jedoch nicht klar, ob dies das spannendste Spiel ist. Es muss also ein Richtwert erstellt werden, welcher die Sehenswürdigkeit des Spiels angibt, um bei der Auswahl des Spiels den Benutzer zu unterstützen. Die Frage ist nun, wie dieser Richtwert definiert wird.

2.1 Sehenswürdigkeit

Um den Begriff der Sehenswürdigkeit in Bezug auf ein Basketballspiel zu definieren, müssen die Daten zu einem Spiel analysiert werden. Am Ende ist die Sehenswürdigkeit jedoch subjektiv. Wenn Daten zum Benutzer vorhanden sind, soll eine genauere Sehenswürdigkeit berechnet werden (Abschnitt 3.2), als in den Fällen, in denen keine Daten zum Benutzer (Abschnitt 3.1) vorhanden sind. Zusätzlich müssen Kriterien gefunden werden, um die Sehenswürdigkeit zu quantifizieren.

2.2 Spoiler-Free

Die Wertung der Sehenswürdigkeit soll keine Rückschlüsse auf den Verlauf des Spieles ermöglichen und somit spoiler-free sein. Wenn der Algorithmus

¹National Basketball Association

²Es gibt eine kleine Anzahl von Exhibition-Spielen, welche an anderen Orten ausgetragen werden.

nur Spiele einer bestimmten Kategorie (z.B. Spiele, die in die Verlängerung gehen) als sehenswert bewertet, mindert dies die Spannung, da der Verlauf teilweise preisgegeben wird.

Kapitel 3

Lösungskonzept

Bei der Lösung werden zwei Use Cases beachtet. Einerseits die allgemeine Sehenswürdigkeit (Abschnitt 3.1) für anonyme Benutzer und die personalisierte Sehenswürdigkeit (Abschnitt 3.2) für Benutzer, zu welchen es zusätzliche Daten gibt. In den nächsten Kapiteln werden für jeden Use Case verschiedene Lösungen vorgestellt und beschrieben, wie diese umgesetzt werden könnten. Die Umsetzung der Lösungen wird im Kapitel 4 beschrieben, wobei nicht jedes Lösungskonzept umgesetzt wird.

3.1 Allgemeine Sehenswürdigkeit

Die allgemeine Sehenswürdigkeit zeigt für anonyme Benutzer auf, wie spannend ein Spiel ist. Dabei werden verschiedene Ansätze verfolgt. Die perfekte Lösung gibt es nicht, da, wie schon erwähnt, die Sehenswürdigkeit subjektiv ist.

3.1.1 Spielwertung durch die Benutzer

Wie oft auch auf anderen Webseiten verwendet¹, können Benutzer ein Spiel bewerten. Diese Wertung gibt dem anonymen Benutzer einen Richtwert, wie spannend das Spiel war.

Vorteile: Die Bewertung gibt genau den Durchschnitt der Meinungen über ein Spiel wieder.

Nachteile: Damit die Bewertung aussagekräftig ist, braucht es mehrere Wertungen, damit eine zuverlässige Aussage getroffen werden kann und statistische Ausbrüche vermieden werden.

¹Amazon, Digitec usw..

3.1.2 Sehenswürdigkeit Modell

Mittels den Daten zu einem Spiel wird ein Modell erstellt, welches Spiele nach deren Spannung bewertet. Dabei liegt die Schwierigkeit bei der Beurteilung, ob das Modell das richtige Resultat liefert. Indem der Benutzer ein Spiel mit einem '+' oder einem '-' bewerten kann (Unterabschnitt 3.1.1), ist es möglich, das Resultat des Modells zu beurteilen und das Modell mit der Zeit zu verfeinern.

Vorteile: Das Modell kann ohne Benutzerdaten auskommen und wenn Benutzerdaten vorhanden sind, welche das Modell bewerten, ist es möglich, dieses zu verbessern.

Nachteile: Die berechnete Sehenswürdigkeit ist stark vom Modell abhängig und es ist schwierig, ein perfektes Modell zu erstellen.

3.1.3 Supervised Learning: Regression

Supervised Learning eignet sich sehr gut, um ein Modell zu finden, welches zu den Daten passt. Dabei werden Statistiken von einem Spiel als Input-Werte verwendet. Zu jedem Spiel gibt es noch einen Resultat-Wert, der angibt, wie sehenswert das Spiel war. Mittels den Daten ist es möglich, eine Funktion zu approximieren, welche möglichst genau für alle Input-Werte eines Spieles die entsprechenden Resultatwerte ausgibt.

Die Schwierigkeit bei diesem Ansatz liegt darin, gute Resultatwerte zu finden. Die folgenden Kapitel zählen einige Ideen auf, wie solche Resultatwerte aussehen können.

Bild-/Video-/Ton-Analyse

Die Aufnahmen des Spiels werden analysiert. Dabei werden Bild, Video und Ton nach deren Sehenswürdigkeit bewertet. Aus diesen Daten wird die Sehenswürdigkeit eines Spieles bestimmt. Dieser Ansatz ist sehr aufwändig umzusetzen. Wenn die Analyse sehr gut ist, kann es zu einem exakten Resultat führen, da weit mehr Daten gesammelt werden können als nur die Statistiken von einem Spiel.

Vorteile: Neben den Statistiken über ein Spiel werden zusätzliche Daten beachtet.

Nachteile: Die berechnete Sehenswürdigkeit anhand von Bild/Video/Ton ist schwierig und aufwändig.

Zuschauerzahlen

Indem die Anzahl der Zuschauern pro Spiel betrachtet wird, kann eventuell eruiert werden, wie spannend ein Spiel ist. Die Idee dahinter ist, dass ein

Spiel mit mehr Zuschauern interessanter ist als eines mit wenigen. Die Zuschauerzahl kann entweder durch die Anzahl an Stadionbesuchern oder die Fernseheinschaltquote gemessen werden.

Vorteile: Die Sehenswürdigkeit ist abhängig von der Anzahl der Benutzer, woraus sich schliessen lässt, dass die Sehenswürdigkeit der Meinung der meisten Benutzer entsprechen wird.

Nachteile: Nicht in jeder Region hat es gleich viele Menschen, somit werden gewisse Teams immer mehr Zuschauer haben als andere. Zusätzlich kommt dazu, dass die Zuschauerzahlen nach dem Spiel sich nicht mehr ändern, was dazu führt, dass Methoden wie Social Media Echo, die auch Daten nach dem Spiel betrachten, eine genauere Aussage treffen können bezüglich der Sehenswürdigkeit eines Spiels.

Anzahl Highlights

Pro Tag werden über alle Spiele Highlights erstellt. Es wird davon ausgegangen, dass das Spiel mit den meisten Highlights das interessanteste war.

Vorteile: Highlights sind einfach zu quantifizieren.

Nachteile: Das Problem dabei ist, dass Highlights immer erstellt werden, auch wenn keine Spiele an einem Tag sehenswert sind.

Social Media Echo

Das Social Media Echo wird darüber bestimmt, wie viel über ein Spiel in den Social Media berichtet wird. Dabei wird die Annahme getroffen, dass wenn mehr über ein Spiel berichtet wird, dieses spannender ist.

Vorteile: Gleich wie bei der Anzahl der Zuschauer ist beim Social Media Echo das Resultat abhängig von der Benutzeranzahl.

Nachteile: Es ist unklar, ob das Social Media Echo die Sehenswürdigkeit reflektiert, da auch Kommentare darüber, wie langweilig ein Spiel ist, beachtet werden.

Reddit Score

Auf Reddit wird pro Spiel ein Thread erstellt, wobei Benutzer den Thread Up- oder Downvoten können. Daraus resultiert ein Score, welcher als Quantifizierung von Sehenswürdigkeit betrachtet werden kann.

Vorteile: Wie bei der Anzahl der Zuschauer ist der Reddit Score abhängig von der Anzahl Benutzer.

Nachteile: Die Relation von Score zu Sehenswürdigkeit stimmt eventuell nicht, da die Anzahl Reddit-Benutzer über die Zeit steigt. Somit erhalten

aktuelle Threads einen besseren Score als ältere Threads, da auf Reddit vor allem neue Threads bewertet werden.

3.2 Persönliche Sehenswürdigkeit

Um die persönliche Sehenswürdigkeit zu bestimmen, werden neben den Daten zu einem Spiel auch Daten vom Benutzer betrachtet. Dabei soll eine personalisierte und genauere Wertung der Sehenswürdigkeit der Spiele entstehen.

3.2.1 Unsupervised Learning: Kategorisierung

Indem sich der Benutzer registriert und Spiele nach deren Sehenswürdigkeit bewertet, können mittels Unsupervised Learning die Spiele in Kategorien aufgeteilt werden. Wenn der Benutzer nun oft Spiele einer gewissen Kategorie bewertet, können dem Benutzer Spiele einer Kategorie als spannend oder langweilig vorgeschlagen werden.

Vorteile: Durch die Kategorisierung können persönliche Vorschläge erstellt werden und eine genauere Wertung entsteht. Indem der Benutzer weitere Spiele bewertet, werden die Vorschläge laufend verbessert.

Nachteile: Die Kategorisierung der Spiele kann so aufgeteilt sein, dass für ein Benutzer keine Kategorie als spannend betrachtet wird. Dies führt dazu, dass die Kategorisierung verfeinert werden müsste, was mit Mehraufwand verbunden ist. Trotzdem kann nicht garantiert werden, dass eine zufriedenstellende Kategorisierung erstellt werden kann.

3.2.2 Persönliche Vorlieben

Eine andere Variante wäre, dass das allgemeine Modell (Unterabschnitt 3.1.2) gewichtet wird. Dabei muss der Benutzer angeben, für welche Werte des Modells, z.B. das Team, die Spieler, oder eine bestimmte Statistik, er sich interessiert, damit das Modell dementsprechend gewichtet werden kann.

Vorteile: Das Modell kann granulos eingestellt werden und für ältere Spiele kann die Sehenswürdigkeit einfach berechnet werden.

Nachteile: Das Resultat hängt davon ab, wie gut das Modell ist und durch die Gewichtung können dank dem Resultat Rückschlüsse gezogen werden, wie der Verlauf eines Spieles war.

Kapitel 4

Umsetzung

In diesem Kapitel wird die Umsetzung beschrieben, mit dem Ziel, dem Leser eine Übersicht zu geben, welche Schritte während der Arbeit durchgeführt wurden. Die Umsetzung wird in zwei Kapitel unterteilt. Zuerst wird der Prototyp beschrieben und anschliessend, wie dieser erweitert wurde, um am Schluss die Frage zu beantworten, wie sehenswert ein Basketballspiel ist.

4.1 Prototyp

Mit dem Prototyp verfolgte ich das Ziel, mich mit den Daten vertraut zu machen, sowie erstmals einen Supervised-Learning-Algorithmus anzuwenden. Die Daten wurden von [basketball-reference](http://basketball-reference.com)¹ geparsed und linear regression² wurde angewendet. Bei Supervised-Learning müssen Input-Werte x und ein Resultat-Wert y mitgegeben werden. Als x wurden die Statistiken über ein Spiel mitgegeben und als Resultatwerte wurde eine einfache Funktion definiert, welche unter 4.1.1 beschrieben wird.

4.1.1 Prototyp Sehenswürdigkeit Funktion

Um die Sehenswürdigkeit eines Spiels und somit die Resultatwerte y zu definieren, wurde eine Funktion erstellt. Diese ist so aufgebaut, dass jede Statistik s über das Spiel einen Punkt zu der Gesamtbewertung geben kann. Zusätzlich werden pq Punkte pro Viertel sowie die Punktedifferenz pro Viertel pd separat gewichtet. Es wird davon ausgegangen, dass Spiele mit einer geringen Punktedifferenz und einer hohen Punkteanzahl als sehenswert betrachtet werden können.

¹<http://basketball-reference.com>

²http://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LinearRegression.html

$$f(s_i) = \begin{cases} s_i > 90\% = 1 \\ s_i < 90\% = 0 \end{cases} \quad (4.1)$$

Die Funktion $f(s_i)$ gibt 1 zurück, wenn s_i zu den besten 90% gehört, ansonsten gibt sie 0 zurück. Somit entsteht ein *rating* pro Spiel, welches die Sehenswürdigkeit und somit die Resultatwerte y für den Prototyp liefert.

$$rating = 4 * (pq + pd) + \sum_{i=1}^{len(f)} f(s_i) \quad (4.2)$$

4.1.2 Resultat

Beim Plotten der erstellten (s. 4.1.1) und der vorhergesagten Resultatwerte wird ersichtlich, dass linear regression, sowie die Resultatwerte an sich, noch nicht optimal sind. Das Modell wäre dann perfekt, wenn alle vorausgesagten Resultatwerte den erstellten entsprechen würden. Dies wäre der Fall, wenn alle Daten sich auf der Gerade befinden würden.

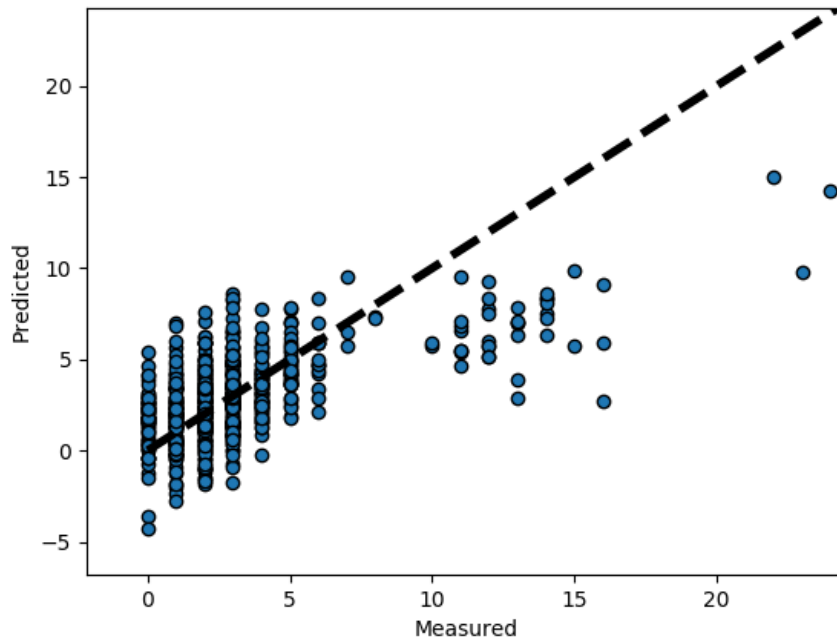


Abbildung 4.1: Erstellte und vorausgesagte Resultatwerte der Prototyp-Funktion.

# Daten: 3412		
Algorithmus	Training set score	Test set score
Linear Regression	0.44	0.45

Tabelle 4.1: Prototyp-Resultat des linear regression Modell.

Dies widerspiegelt sich auch in der mittleren quadratischen Abweichung³, welche 7.75 beträgt. Diese Abweichung wirkt sich auf den score (Bestimmtheitsmass⁴) aus, welcher mit 0.45 (s. Tabelle 4.1) zu tief ist, um mit dem Modell genaue Voraussagen zu machen⁵.

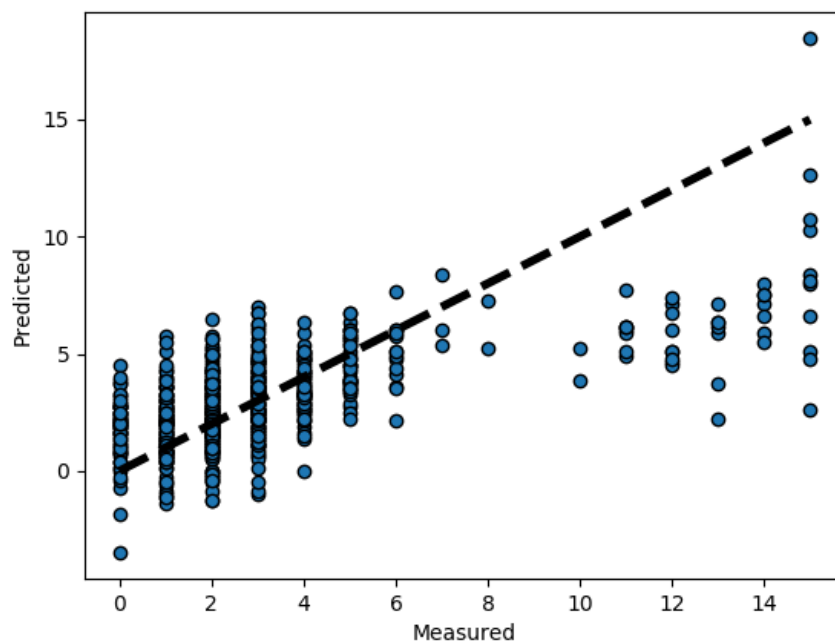


Abbildung 4.2: Erstellte und vorausgesagte Resultatwerte der Prototyp-Funktion, wobei der Rückgabewert auf die Werte von 0-15 begrenzt ist.

Wenn die Funktion so angepasst wird, dass der Rückgabewert nur zwischen 0-15 liegen kann (s. Abbildung 4.2), wird das Problem gut ersichtlich. Die Spiele, welche eine hohe Wertung erhalten und daher als sehenswert betrachtet werden, können besonders schlecht vorausgesagt werden.

³https://de.wikipedia.org/wiki/Mittlere_quadratische_Abweichung

⁴<https://de.wikipedia.org/wiki/Bestimmtheitsma%C3%9F>

⁵Das beste Resultat des Bestimmtheitsmasses wäre 1.0. Dies würde bedeuten, dass alle Punkte sich auf der Geraden befinden (s. http://scikit-learn.org/stable/modules/model_evaluation.html#r2-score)

Also kann davon ausgegangen werden, dass die verwendete Resultatwerte y (Funktion 4.1.1) oder die Input-Werte x nicht genügen, um ein Modell aufzustellen, um die Sehenswürdigkeit zu bestimmen.

4.2 Sehenswürdigkeit verbessern

Beim Prototyp 4.1 wurde eine Funktion 4.1.1 verwendet, um die Resultatwerte y zu bestimmen. Mit dem resultierenden Modell kann die Sehenswürdigkeit eines Spiels nur schlecht bestimmt werden. In diesem Kapitel werden andere Ideen verfolgt, um die Resultatwerte y zu bestimmen. Dabei werden die Idee der Reddit-Ups 3.1.3, sowie die Daten von der Seite Wikihoops verwendet 4.2.2.

Die folgenden Abschnitte beschreiben zuerst die einzelnen verwendeten Resultatwerte y , wobei die Werte genauer betrachtet werden und erklärt wird, warum und wie diese konsolidiert wurden. Zu jedem Abschnitt werden die Resultate mittels dem Ridge Modell⁶ und den Random-Forest Modell⁷ erstellt. Dabei werden die verwendeten Resultatwerte y , sowie die vorausgesagten Resultatwerte y geplottet, um zu sehen, wie gut das Modell die Sehenswürdigkeit bestimmt⁸. Die Resultate des score (Bestimmtheitsmass) der Trainingsdaten, sowie die Testdaten werden als Tabelle dargestellt. Zum Schluss werden die Resultate verglichen und Schlüsse gezogen, wie die Sehenswürdigkeit bestimmt werden kann.

4.2.1 Reddit-Ups Wertung

Reddit hat eine sehr grosse NBA-Fangemeinschaft. Diese erstellt seit etwa zwei Jahren für jedes Spiel einen Reddit-Thread, wobei jeder Benutzer eine Stimme abgeben kann, ob der Thread spannend ist oder nicht. Die Anzahl an Reddit-Ups kann als Indikator für die Sehenswürdigkeit des Spiels betrachtet werden 3.1.3. Wenn alle Daten dargestellt werden, ist ersichtlich, dass es erst ab Mitte 2015 genügend Daten gibt (s. Abbildung 4.3).

Wenn die Daten zwischen 2015 und 2017 betrachtet werden, gibt es einen Unterschied in der Höhe der Wertungen. Dies ist darin zu erklären, dass die Anzahl Abonnenten des NBA-Subreddits sich in dieser Zeit verdreifacht hat⁹. Ein weiterer Punkt, welcher zu beachten ist, sind Threads mit einer sehr

⁶http://scikit-learn.org/stable/modules/generated/sklearn.linear_model.Ridge.html

⁷<http://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestRegressor.html>

⁸Ideal wäre, wenn die gemessenen Resultatwerte gleich den vorausgesagten Resultatwerten sind. Dies wäre der Fall, wenn sich alle Punkte auf der Linie befinden würden.

⁹Mitte 2015 hatte der NBA-Subreddit 300k Abonnenten und anfangs Dezember 2017 waren es 900k (<http://redditmetrics.com/r/nba>).

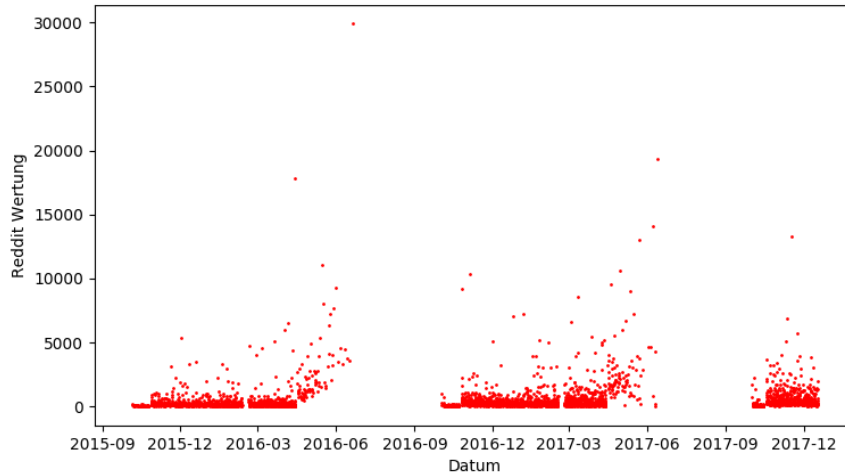


Abbildung 4.3: Reddit-Ups Wertung zwischen 2015-2017.

hohen Wertung. Dies ist mit der Reddit-Architektur zu begründen. Sobald ein Thread während einer kurzen Zeit eine hohe Wertung erreicht, wird dieser auf der Reddit-Startseite angezeigt. Dies hat zur Folge, dass der Thread von mehr Benutzern gesehen und bewertet wird und somit eine höhere Wertung erreicht. Zwischen April und Juni sind die Wertungen auch höher. Dies hat damit zu tun, dass während dieser Zeit die Playoffs stattfinden und in dieser Periode mehr Benutzer die NBA verfolgen¹⁰.

Reddit-Wertung konsolidieren

Um die Wertung zu konsolidieren wurden mehrere Methoden versucht. Die erfolgreichste Konsolidierungsmethode bestand darin, die Spiele mit einer Wertung von über 2500 auf 2500 zu reduzieren. Dies ist sinnvoll, da Spiele, die eine Wertung von 2500 oder höher erreichen, als sehenswert betrachtet werden können und es keinen Unterschied macht, ob die Wertung höher ist.

Resultat

Die Resultate sind etwa gleich wie beim Prototypen. Das Modell mit den konsolidierten Wertungen 4.2.1 ist etwas besser als mit den originalen Wertungen (s. Tabelle 4.2), jedoch können die normalisierten Wertungen sehr schlecht vorhergesagt werden (s. Abbildung 4.4).

¹⁰https://en.wikipedia.org/wiki/2017_NBA_Playoffs

4.2. Sehenswürdigkeit verbessern

# Daten: 4190	Originale Daten		Normalisierte Daten	
Algorithmus	Training set score	Test set score	Training set score	Test set score
Riged	0.26	0.19	0.40	0.36
Random Forest	0.86	0.30	0.90	0.40

Tabelle 4.2: Reddit Modell Resultate

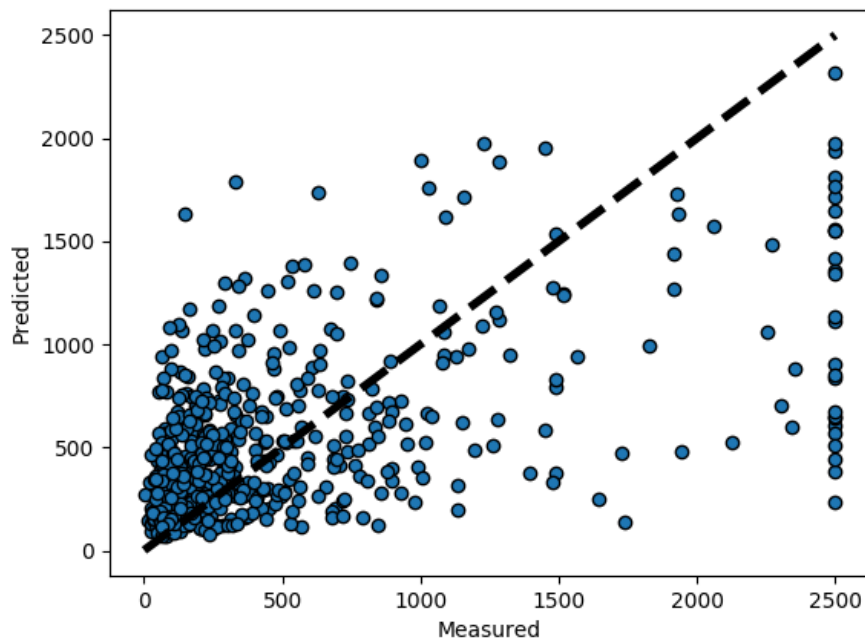


Abbildung 4.4: Random Forest Reddit-Ups normalisierte Resultatwerte.

Wenn die Daten nicht normalisiert werden, ist die Grafik (s. Abbildung 4.5) weniger aussagekräftig. Dies ist auch in der Tabelle 4.2 ersichtlich. Spannend ist, dass Ridge zwar eine schlechtere Trainingswertung ergibt, jedoch die Testwertung und die Trainingswertung recht nahe beieinander sind. Dies ist auch in der Abbildung 4.6 ersichtlich.

4.2.2 Wikihoops

Wikihoops¹¹ ist eine Webseite, die das gleiche Ziel verfolgt, wie diese Arbeit, also die Sehenswürdigkeit eines NBA-Spiels zu bestimmen, ohne dabei das Resultat zu verraten. Die Sehenswürdigkeit wird dabei auf zwei Arten präsentiert: Game- und User-Rating. Das Game-Rating ist das Resultat eines Algorithmus, welcher dem Spiel anhand der Statistiken eine Wertung zwischen 1 und 10 gibt. Das User-Rating erlaubt es den registrierten Benutzern,

¹¹<http://wikihoops.com>

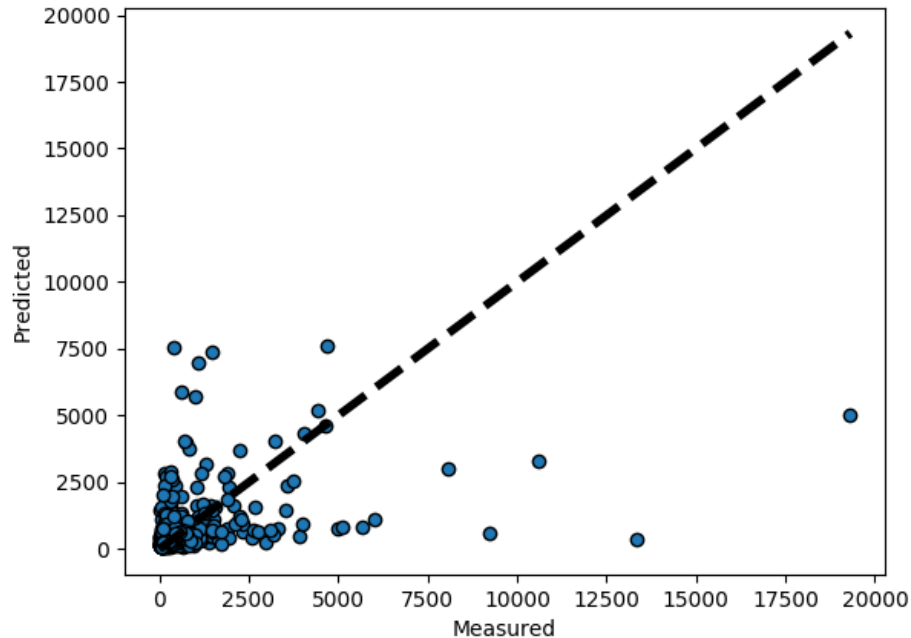


Abbildung 4.5: Random Forest Reddit-Ups originale Resultatwerte.

für jedes Spiel zu entscheiden, ob es sehenswert war oder nicht. Diese Wertung wird dem Game-Rating addiert oder subtrahiert und ergibt so das User-Rating.

Wenn das Game-Rating geplottet wird, ergibt dieses eine gleichmässige Verteilung, da die Werte zwischen 1 und 10 begrenzt sind (s. Abbildung 4.7).

Beim User-Rating sieht es ganz anders aus (s. Abbildung 4.8): die Verteilung der Wertung ist ungleichmässig, mit einzelnen Spielen, welche klar ins Positive gehen. Es ist auch ersichtlich, dass es nicht viele Wertungen pro Spiel gibt, da die Wertungen oft den Wert zwischen 1 und 10 übersteigen. Dabei ist zu beachten, dass eine negative Wertung mit einer positiven Wertung aufgehoben wird.

Zu jedem User-Rating ergibt sich zusätzlich noch eine Prozentzahl, welche ausweist, wie enig sich die Benutzer sind. Dieser Wert zeigt ein Problem der Frage auf. In nur etwa 40% der Spiele sind sich die Benutzer zu 75% oder mehr enig, was die Sehenswürdigkeit eines Spiels betrifft.

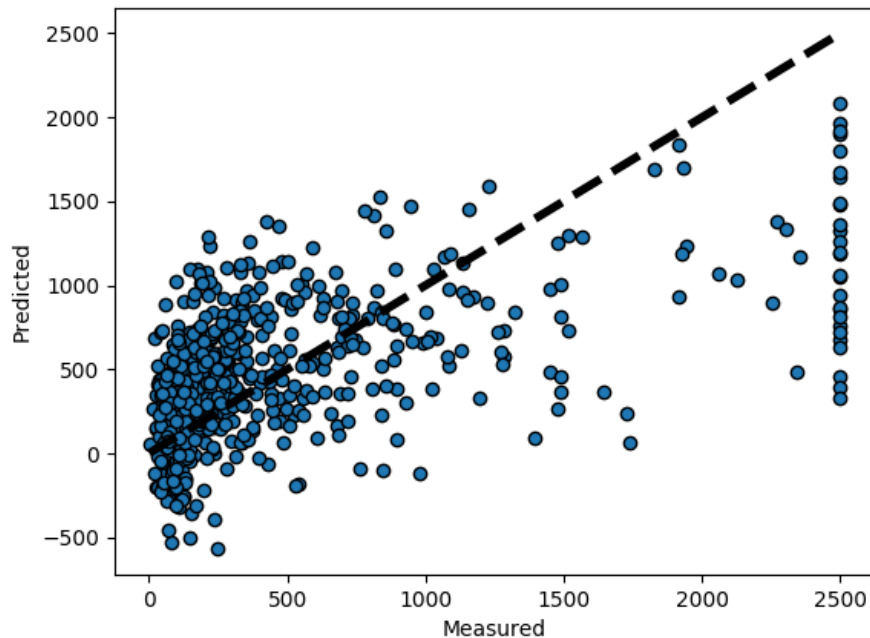


Abbildung 4.6: Ridge Reddit-Ups normalisierte Resultatwerte.

Wikihoops-Wertung konsolidieren

Die Game-Rating Werte müssen nicht konsolidiert werden, da sich diese nur in einem Bereich von 1-10 bewegen (s. Abbildung 4.7). Bei den User-Ratings wurden alle Wertungen, welche grösser als 15 sind, auf 15 reduziert und alle Werte, welche kleiner als -5 sind, auf -5 erhöht. Der Grund dafür ist, dass Spiele, welche eine höhere Wertung als 15 aufweisen, alle als sehenswert betrachtet werden können und dabei die Wertung keine Rolle mehr spielt. Das Gleiche gilt für Spiele mit einer geringen Wertung, welche als nicht sehenswert betrachtet werden.

Resultate

Es ist ersichtlich, dass die Werte des Game-Rating recht gut bestimmt werden können (s. Tabelle 4.4). Dies hat damit zu tun, dass das Modell eine Annäherung der Funktion findet, welche verwendet wurde, um das Game-Rating zu berechnen. Beim User-Rating sieht es etwas schlechter aus. Zwar basiert dieses auf dem Game-Rating, jedoch variieren die Resultatwerte je nach den Bewertungen der Benutzer. Eine weitere Schwierigkeit ist, dass sich die Benutzer nicht immer einig sind, wie sehenswert ein Spiel ist (s. Abbildung 4.10).

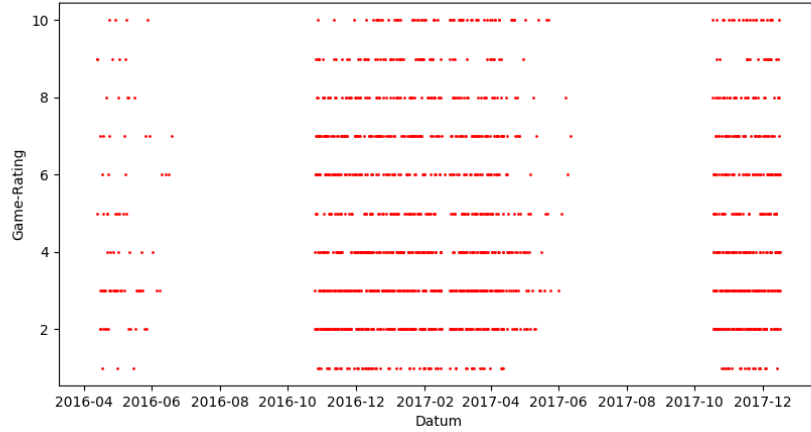


Abbildung 4.7: Wikihoops Game-Rating über die Zeit.

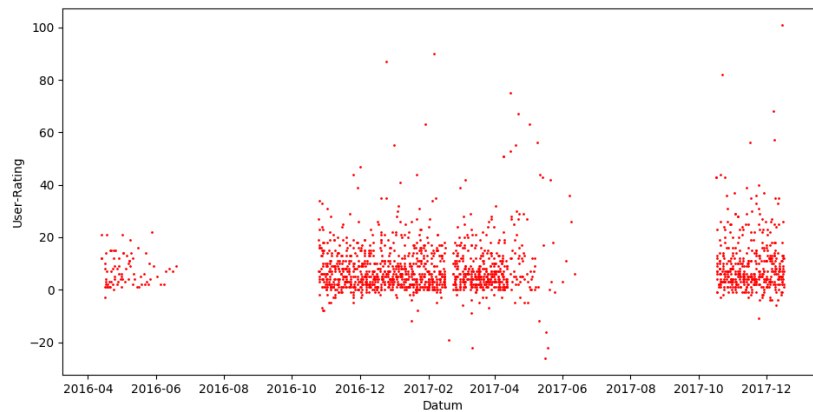


Abbildung 4.8: Wikihoops User-Rating zwischen 2016 und 2017.

# Daten: 1734	Originale Daten		Normalisierte Daten	
Algorithmus	Training set score	Test set score	Training set score	Test set score
Riged	0.66	0.66	0.50	0.43
Random Forest	0.92	0.58	0.96	0.78

Tabelle 4.3: Wikihoops User-Rating Resultat.

Wenn nun nur Daten betrachtet werden, wo sich die Benutzer zu 75% einig sind, was die Sehenswürdigkeit eines Spieles angeht, sehen die Zahlen ganz anders aus (s. Tabelle 4.5). Der score pro Methode sinkt massiv. Dabei sehe ich das Hauptproblem darin, dass es mit 339 Werten zu wenig Informationen gibt, um das Modell sinnvoll zu trainieren. Dies ist vor allem beim der Veränderung der Game-Rating scores gut ersichtlich(s. Tabelle 4.6). Die-

4.2. Sehenswürdigkeit verbessern

# Daten: 1734		
Algorithmus	Training set score	Test set score
Riged	0.44	0.33
Random Forest	0.96	0.80

Tabelle 4.4: Wikihoops Game-Rating Resultate.

se Funktion zu finden, sollte für die Modelle nicht allzu schwierig sein. Da jedoch zu wenig Daten vorhanden sind, sank der test-score auch dort um 40%.

# Daten: 339	Originale Daten		Normalisierte Daten	
Algorithmus	Training set score	Test set score	Training set score	Test set score
Riged	0.53	0.29	0.59	0.18
Random Forest	0.91	0.45	0.91	0.33

Tabelle 4.5: Wikihoops User-Rating mit Daten in denen sich die Benutzer eing sind.

# Daten: 339		
Algorithmus	Training set score	Test set score
Riged	0.37	0.07
Random Forest	0.90	0.38

Tabelle 4.6: Wikihoops Game-Rating mit Daten in denen sich die Benutzer eing sind.

Das Resultat spiegelt sich auch beim Plotten der Daten (s. Abbildungen 4.9 und 4.10) wider. Spannend dabei ist, dass jene Spiele, bei denen sich die Benutzer eing sind, auch die sehenswürdigen Spiele sind.

4.2.3 Vergleich der Resultate

Bei den drei verwendeten Resultatwerten, Reddit-Ups, Wikihoops Game-Rating und Wikihoops User-Rating, waren die Resultate unterschiedlich. Das beste Modell wurde mit den Game-Ratings von Wikihoops erreicht, was, wie schon erwähnt, damit zu begründen ist, dass das Modell die Funktion findet, welche zur Berechnung des Game-Ratings verwendet wurde. Spannend ist es jedoch, den Unterschied zu betrachten zwischen dem Wikihoops User-Rating und den Reddit-Ups, wo nicht nur eine Funktion hinter der Wertung steckt, sondern auch Benutzermeinungen einen Einfluss haben.

Wenn für die beiden Wertungen die konsolidierte Version (s. 4.2.1, 4.2.2) nach deren Sehenswürdigkeit betrachtet wird, ist es möglich, diese zu vergleichen. Dabei wird die Differenz der Sehenswürdigkeit pro Spiel betrachtet. Spannend wäre, wenn beide Wertungen die gleichen Spiele als sehenswürdig betrachten würden, denn somit könnte mit beiden Wertungen eine bessere Sehenswürdigkeit erstellt werden. Dies ist jedoch nicht der Fall, wie die Abbildung 4.11 zeigt. Die Abbildung plottet den Durchschnitt der

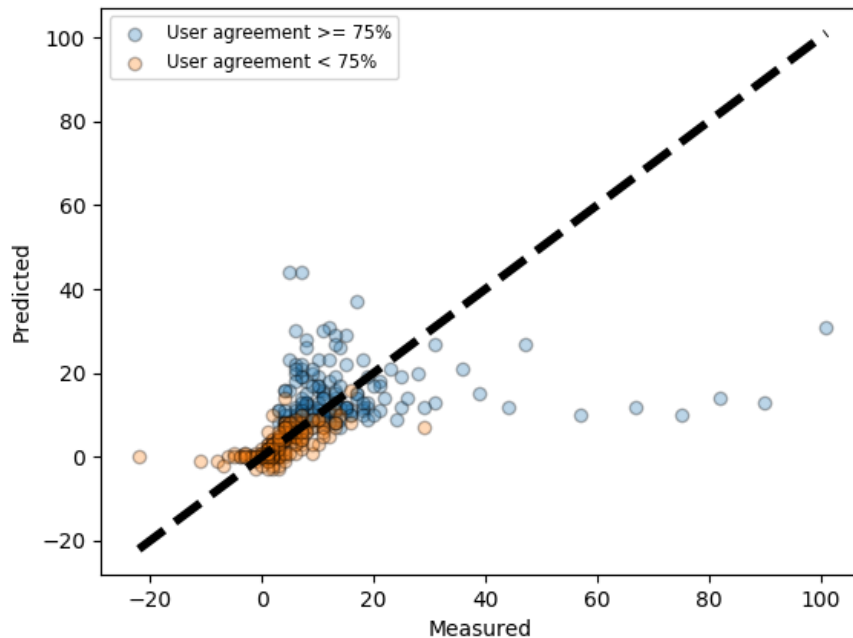


Abbildung 4.9: Random Forest Wikihoops Game-Rating Resultatwerte.

beiden Werte als Säule und der Strich zeigt auf, wie gross die durchschnittliche Differenz ist zu den gleichen Spielen der anderen Wertung. Da die durchschnittliche Differenz zu der anderen Wertung sehr gross ist, wird klar, dass die Reddit-Benutzer und die Wikihoop-Benutzer sich nicht einig sind, welche Spiele sehenswert sind.

Am Anfang der Arbeit dachte ich nicht, dass der Unterschied zwischen dem Wikihoops User-Rating und den Reddit-Ups so gross sein würde. Ein Unterschied war zu erwarten, da die Resultate unterschiedlich berechnet werden. Reddit-Ups basiert rein auf Benutzerbewertungen, während das Wikihoops User-Rating zum Teil mit einer Funktion berechnet wird. Eines der Probleme, welches noch nicht betrachtet wurde, ist, wie die Reddit-Ups entstehen und ob gewisse Teams mit einer grösseren Fangemeinschaft mehr Ups erhalten als andere Teams und somit die Reddit-Ups keine neutrale Sicht der Sehenswürdigkeit darstellen. Zusätzlich gibt es für das User-Rating im Vergleich zu den Reddit-Ups viel weniger Daten, was auch ein Grund dafür sein könnte, dass der Unterschied so gross ist.

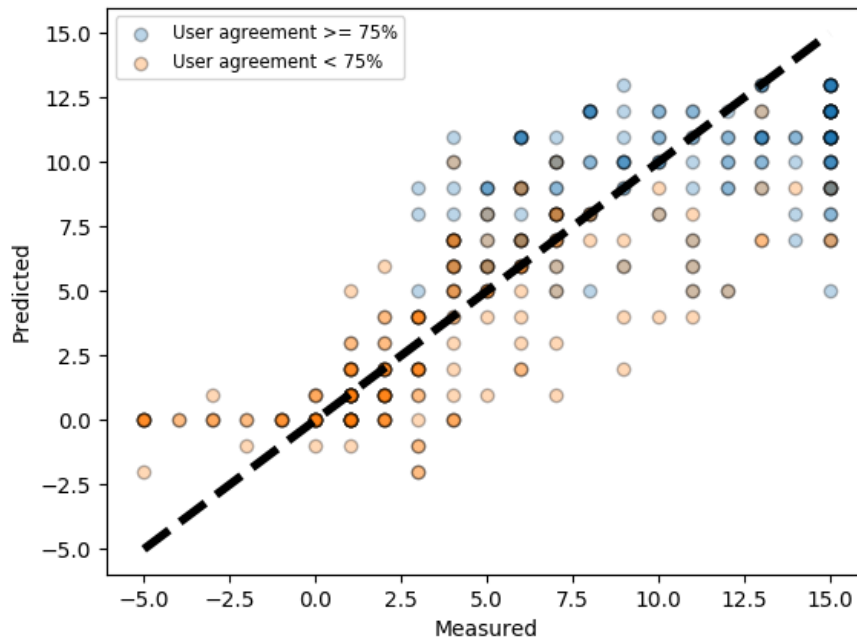


Abbildung 4.10: Random Forest Wikihoops Game-Rating normalisierte Resultatwerte.

4.2.4 Schlussfolgerung

Um nun die Frage zu beantworten, wie sehenswert ein Spiel ist, muss ein Modell gewählt werden. Am Anfang der Arbeit dachte ich, dass, wenn für das Wikihoops User-Rating und für die Reddit-Ups ein Modell erstellt wird, diese die gleichen Spiele als sehenswert bewerten würden. Dies ist leider nicht der Fall (s. 4.2.3), wobei dies auch das Problem der Frage nach der Sehenswürdigkeit aufzeigt. Meinungen sind suggestiv. Der Weg, welchen die Webseite Wikihoops eingeschlagen hat, namentlich einen Wert dafür zu berechnen, wie sehenswert ein Spiel ist (Game-Rating) und dann die Benutzer darüber abstimmen lassen, ob die Berechnung stimmt oder nicht, ist sicherlich ein gangbarer Weg. Die Bewertungen der Benutzer werden jedoch nicht in den Algorithmus einbezogen. Dies liesse sich jedoch gut umsetzen, wie in dieser Arbeit gezeigt wurde (s. 4.2.2). Spannend wäre nun, eine personalisierte Sehenswürdigkeit pro Benutzer zu berechnen. Dies war am Anfang der Arbeit auch das Ziel, jedoch zeigte sich schnell, dass dazu die Daten fehlen. Allgemein gibt es zu diesem Thema noch wenig Daten. Zwar gibt es detaillierte Statistiken für Spiele und Spieler seit dem Jahr 1963¹², jedoch fehlt eine Bewertung, wie sehenswert diese Spiele waren.

¹²<https://www.basketball-reference.com/>

4.2. Sehenswürdigkeit verbessern

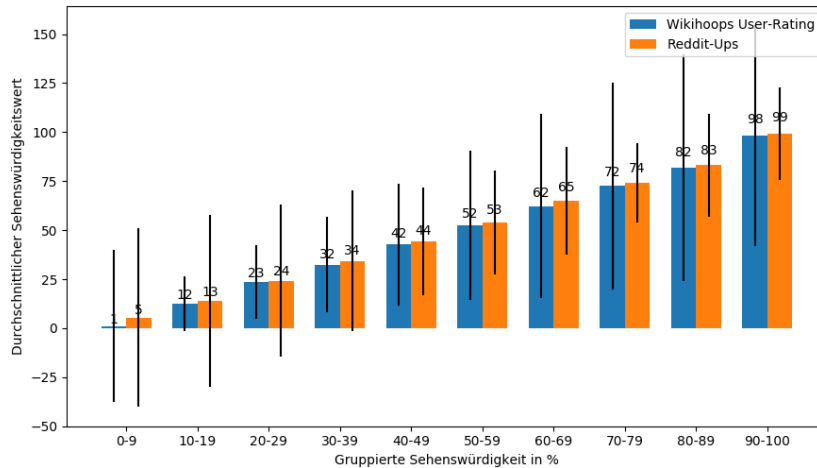


Abbildung 4.11: Sehenswürdigkeit von Wikihoops User-Rating und Reddit-Ups, wobei der Durchschnitt der jeweiligen Wertung als Säulen gezeigt wird (y Achse), diese nach deren Sehenswürdigkeit aufgeteilt sind (x Achse) und die durchschnittliche Differenz als schwarzer Strich dargestellt wird.

Die Frage nach der Sehenswürdigkeit eines Spieles kann nicht endgültig beantwortet werden. In dieser Arbeit werden daher die vorausgesagten Resultatwerte des Modells von Reddit-Up und Wikihoops-User-Rating dargestellt. Der Benutzer muss dann entscheiden, welcher Wert mehr zu seiner persönlichen Meinung passt.

Kapitel 5

Ergebnis

Um das Resultat zu visualisieren, wurde eine Webseite mit Django entwickelt, welche pro Tag die Spiele auflistet und zu jedem Spiel das Resultat des Wikihoops User-Rating sowie den Reddit-Ups Modell anzeigt. Die Daten pro Spiel sowie die Resultate werden via einem Job stündlich für die neuen Spielen berechnet.

5.1 Webseite

Die Webseite ist sehr schlicht gehalten. Der Benutzer hat die Möglichkeit, via einem Kalender die einzelnen Spiele pro Tag anzuzeigen. Dabei werden zu jedem Spiel die Resultate der Modelle angezeigt. Auf eine Benutzerbewertung wurde verzichtet, da dies schon von der Webseite Wikihoops umgesetzt wurde.

5.2 Offene Punkte

Spannend wäre gewesen, eine personalisierte Sehenswürdigkeit zu erstellen. Dafür braucht es jedoch viele Benutzerdaten und diese zu beschaffen, ist während der Zeit der Arbeit unrealistisch.

5.3 Ausblick

Momentan sind Abklärungen mit den Ersteller der Webseite Wikihoops im Gange, um das Modell, welches auf Wikihoops zur Berechnung der Sehenswürdigkeit verwendet wird, anzupassen und die Erkenntnisse dieser Arbeit in ihre Funktion einfließen zu lassen.