



FC BAYERN MÜNCHEN

Schießt Geld Tore?

vs

Oder ist Fußball Zufall?





Bundesliga Match Predictions

Data Exploration Project

Max Bernauer, Julius Koenning, Philipp Dingfelder, SAP, Schaeffler
July 05, 2022

Public

Agenda

1 Projekt-Idee

2 Datenbasis

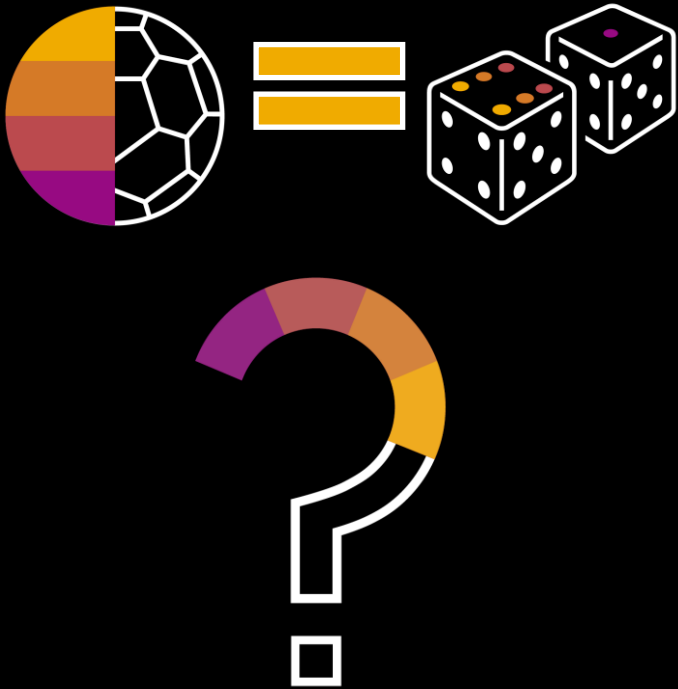
3 Umsetzung & Ergebnisse

4 Business Use Case

5 Kritische Reflexion

Projekt-Idee

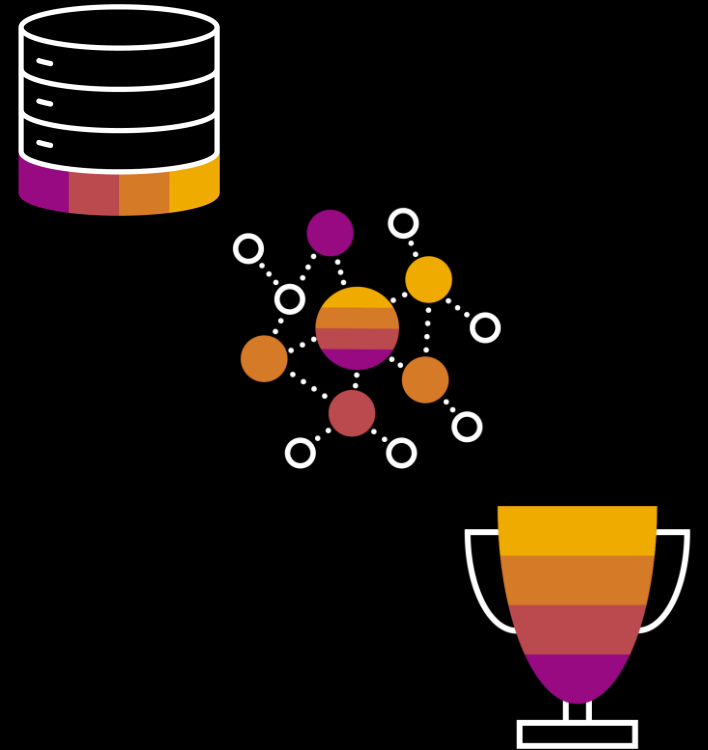
Motivation



Business Use Case



Idee



Datenbasis



Datenbasis



- Einzelner harmonisierter Datensatz
- Saisons 06/07 – 21/22
- 6 verwendete Features

- Fokus auf Differenzen / Quotienten
- Aussagekräftigere errechnete Features
- Feature Aggregation

- Einzelner harmonisierter Datensatz
- Saisons 06/07 – 21/22
- 9 verwendete Features

Datenbasis

Datensatz Vergleich

<i>Kriterium</i>	<i>Zwischenpräsentation</i>	<i>Abschlusspräsentation</i>
Anzahl Saisons	16	16
Anzahl genutzte Features	6	9
Ø Korrelation zu FTR	~11,58 %	~21,42 %
Skalierung	Nein	Ja



**Aussagekräftig ▲
Features ▲**

Datenbasis

Correlation Matrix

Nach Fokussierung auf Differenzen:

	FTR	FTHG	FTAG
DiffEloOld	0.265885	0.259722	-0.218671
DiffAttackOld	0.223388	0.200963	-0.182567
DiffDefendOld	0.115601	0.112869	-0.085549
PDiff3Matches	0.178228	0.164853	-0.145809
PDiff10Matches	0.276164	0.245761	-0.231113
PQuotAllMatches	0.187325	0.185435	-0.151971
MarketValueDiff	0.306257	0.308663	-0.254907
DirectComparisonHG	0.189993	0.239505	-0.112668
DirectComparisonAG	-0.185371	-0.155700	0.183249

Datenbasis

Berechnete Features

	DiffEloOld	DiffAttackOld	DiffDefendOld	PDiff3Matches	PDiff10Matches	PQuotAllMatches	MarketValueDiff	DirectComparisonHG	DirectComparisonAG
4886	172	0	0	0.0	7.0	1.527778	408.45	2.000000	0.800000
4887	683	0	0	2.0	8.0	1.894737	57.87	1.333333	0.666667
4888	-603	-2	-2	-2.0	-18.0	0.482759	-452.10	1.000000	1.000000
4889	230	2	-1	-1.0	7.0	1.970588	428.85	2.400000	1.200000
4890	48	0	3	5.0	2.0	1.107143	227.60	1.600000	1.400000
4891	5	1	0	4.0	1.0	1.095238	-130.75	1.000000	1.000000
4892	22	0	2	4.0	7.0	0.914894	92.35	1.600000	1.400000
4893	-28	-1	-1	-4.0	-6.0	0.584906	92.97	1.400000	1.000000
4894	357	1	1	1.0	7.0	1.279070	40.90	1.800000	1.100000
4895	-293	-1	1	3.0	-7.0	0.545455	-565.15	0.600000	3.000000

Elo, Angriffs und Verteidigungswert

Punkte aus den letzten Spielen

Marktwert

Direkter Vergleich (Tore)



Fokus auf direktem Vergleich durch Differenzen

Umsetzung & Ergebnisse



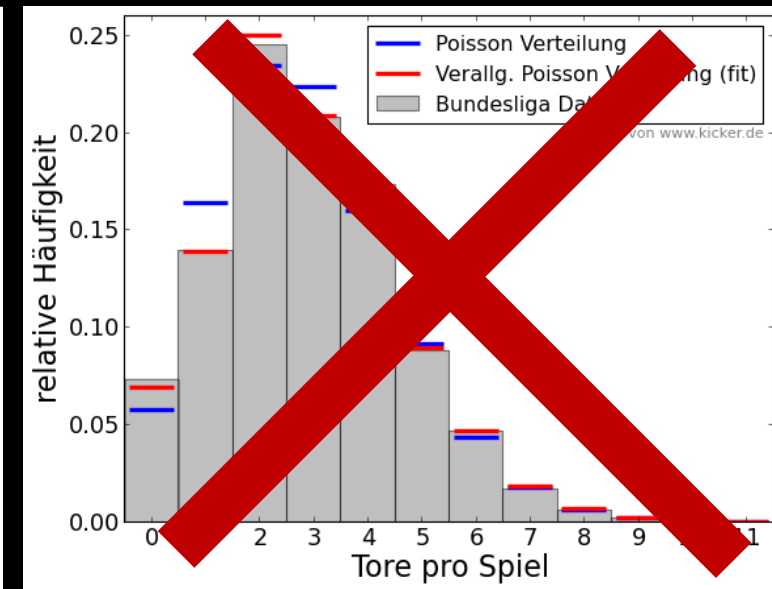
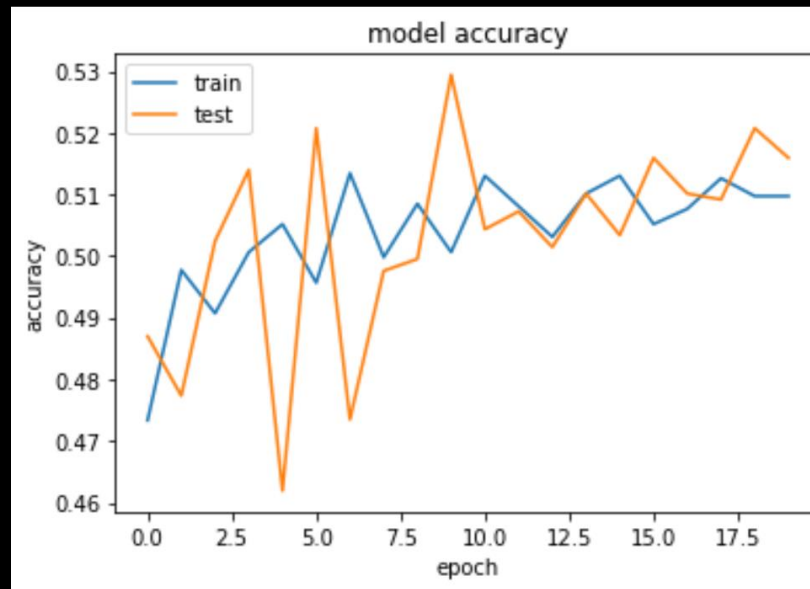
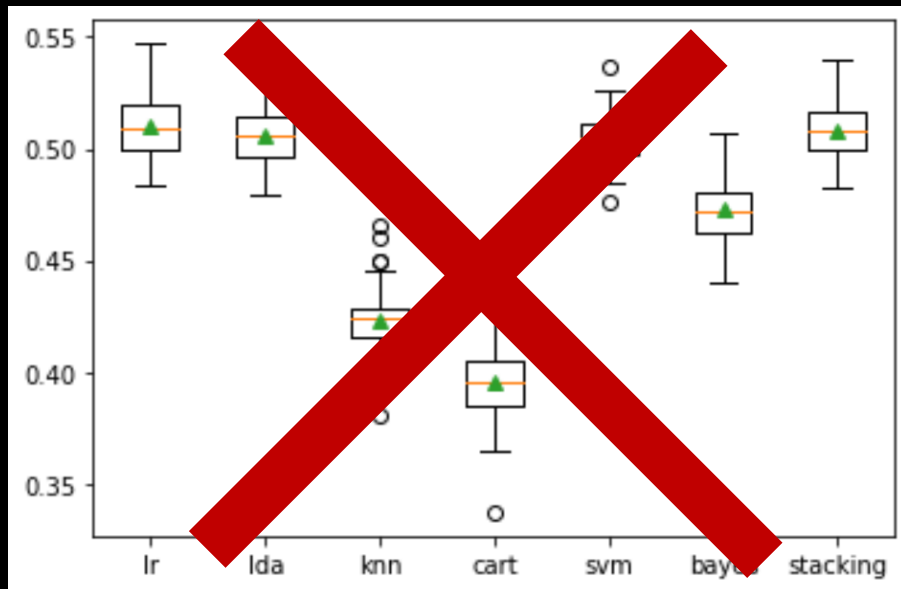
Umsetzung und Ergebnisse

Vorhersage des Endergebnisses - Supervised Learning-Problem

Ensemble Learning

Neural Network

Mathematischer Ansatz



Neural Network

Neural Network bei
Zwischenpräsentation

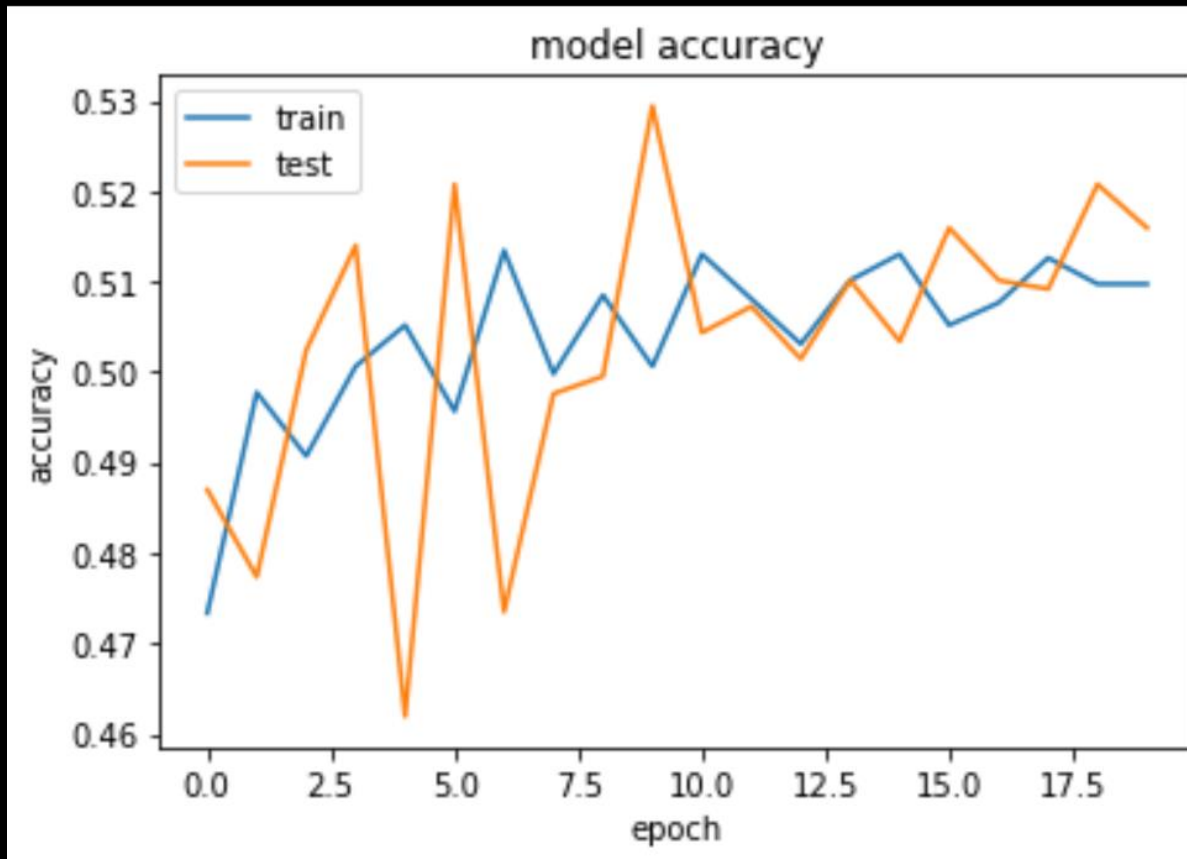
Accuracy: 48,70%

Neural Network mit
neuen Features

Accuracy: 51,40%

Neural Network nach
Hyperparameter-
Tuning

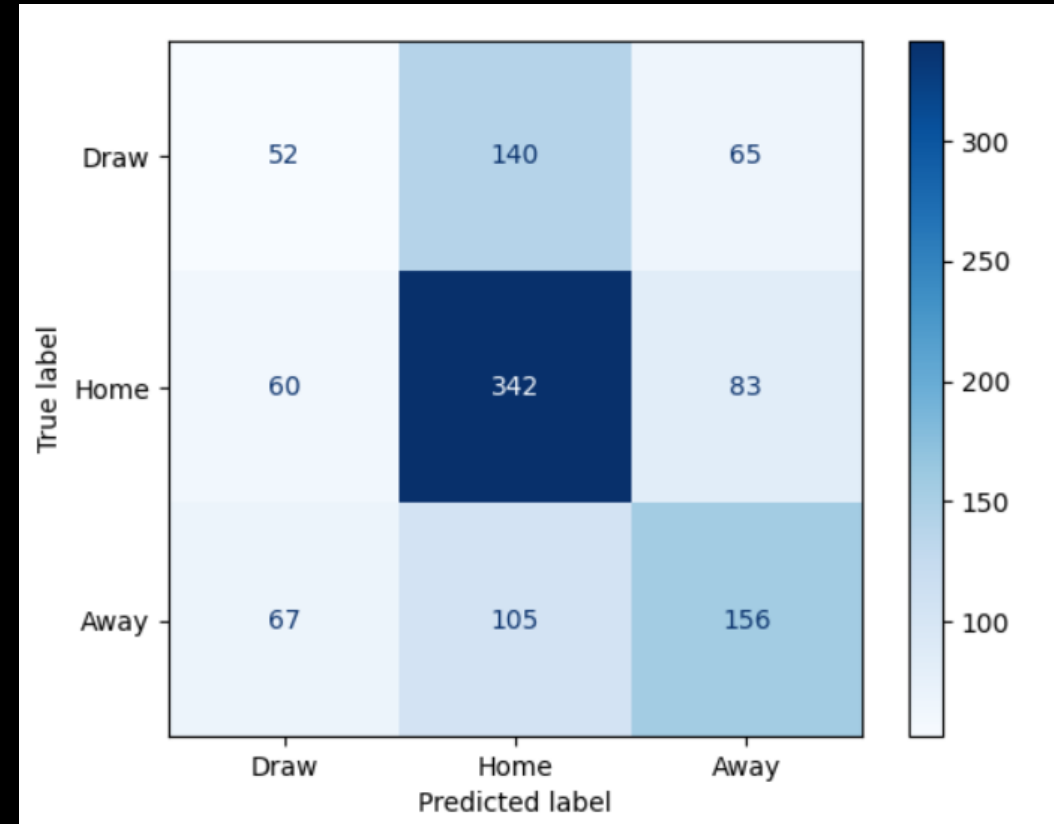
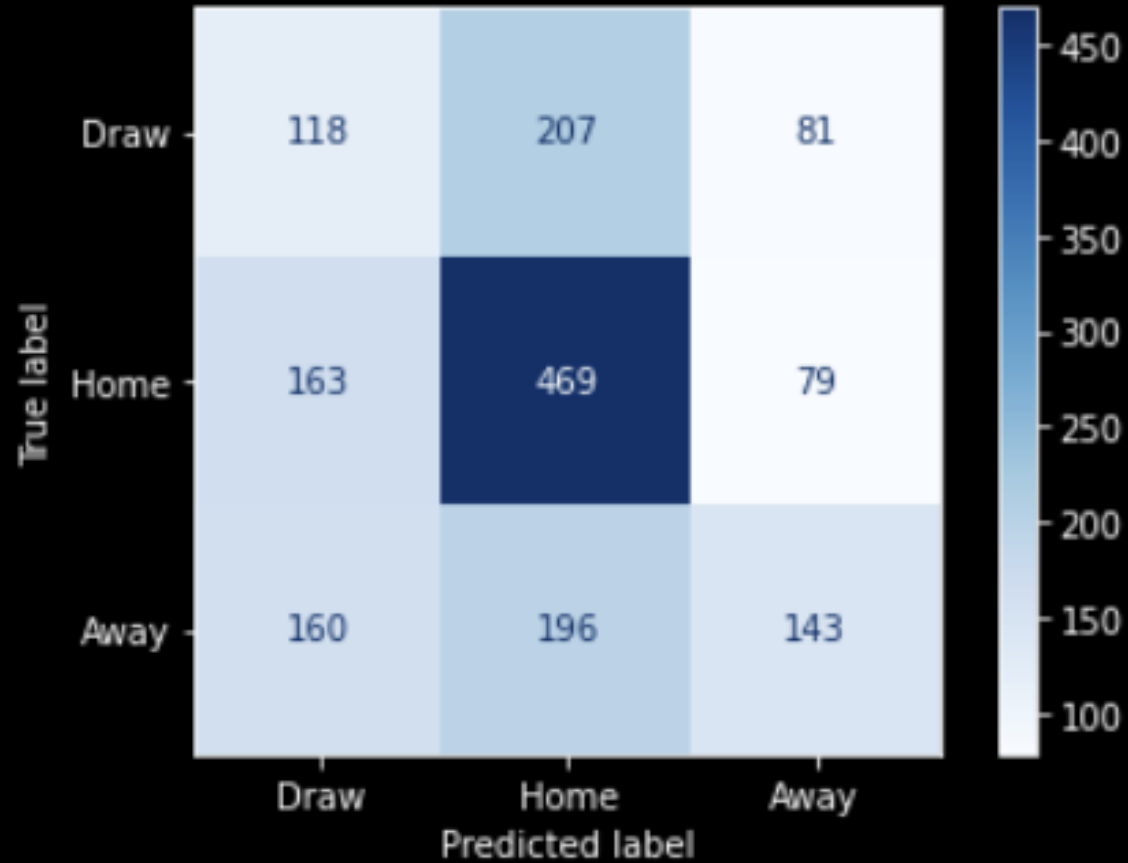
Accuracy: 53,11%



Trial 18 Complete [00h 00m 14s]
val_accuracy: 0.5247589945793152

Best val_accuracy So Far: 0.5311130732297897

Vergleich zur Zwischenpräsentation



Business Use Case



Business Use Case

Validierung durch simuliertes Wetten



Validierung durch simuliertes Wetten

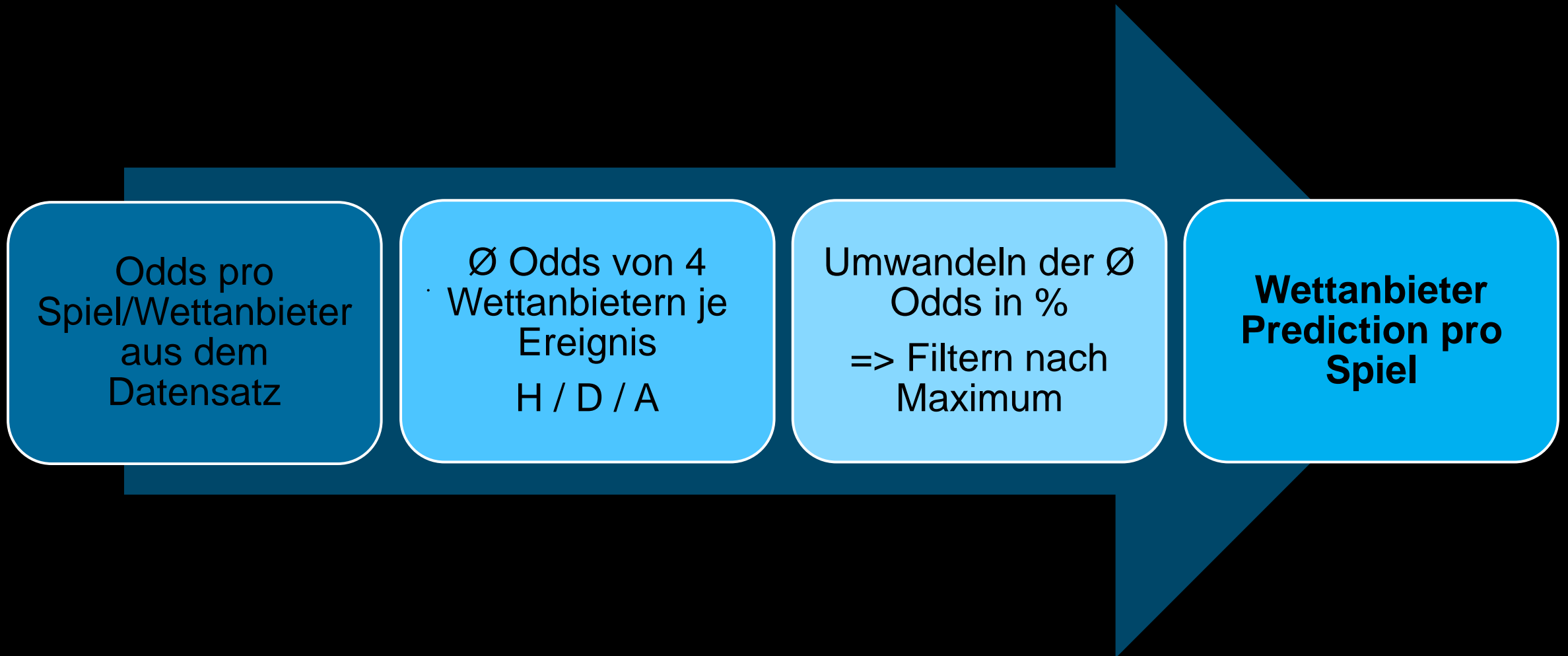
Wettanbieter

VS

Model

Business Use Case

Wettanbieter Prediction



Business Use Case

Modell Auswertung

	Overall	Saison 21/22
Wettanbieter	51,6%	50,4%
Unser Modell	53,1%	50,0%

Gründe Unterschied Overall:

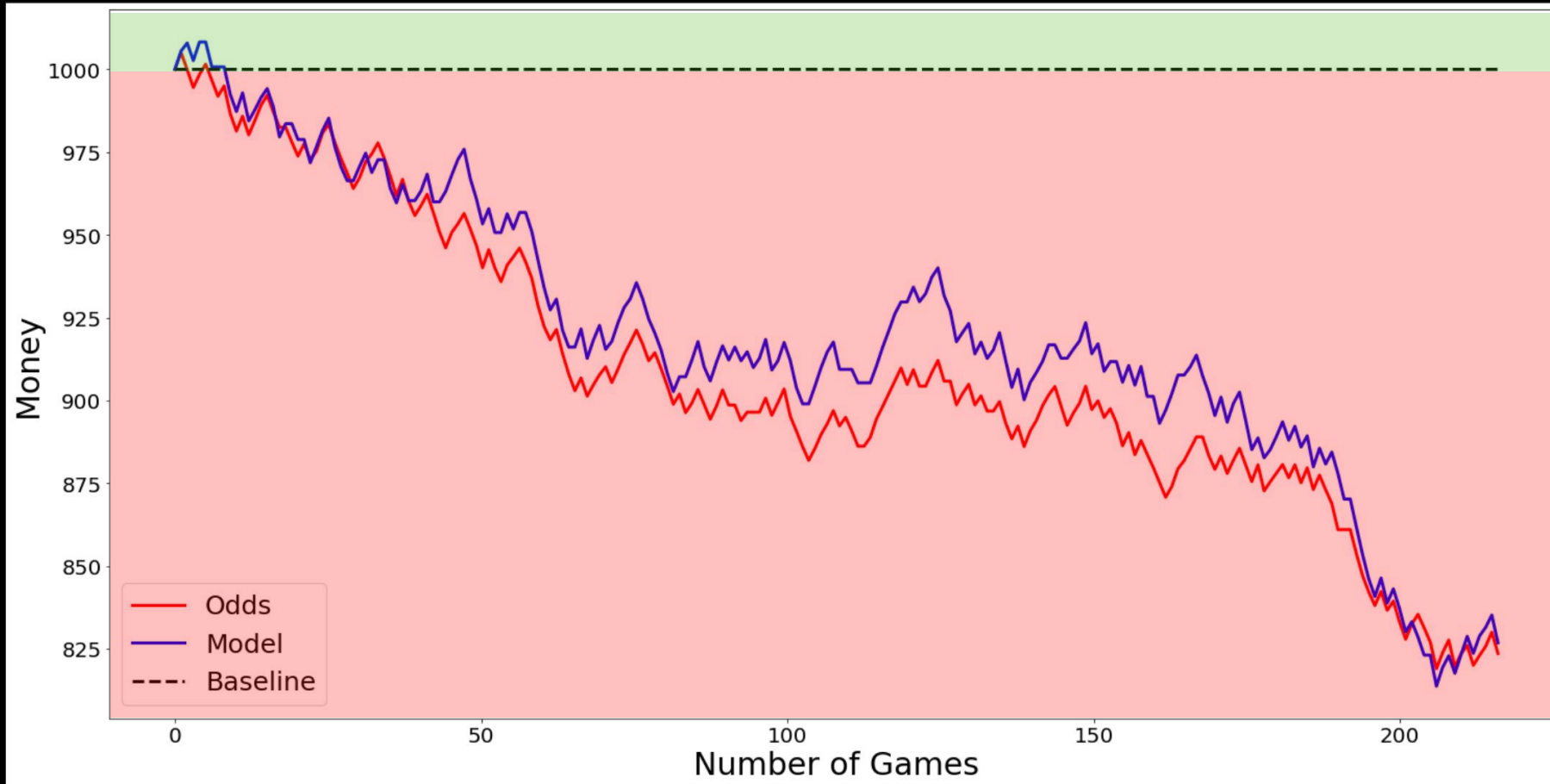
1. Mehr Daten zum Trainieren als Wettanbieter
2. Daten vorhanden, die nach dem Ereignis liegen
3. Optimierte Modelle/ Techniken

Gründe Unterschied Saison:

1. Verwenden Saisondaten nicht zum Trainieren
2. Modelloptimierung auf den vergangenen Saisons
3. Bestes Ergebnis des Modells

Business Use Case

Vergleich Wettanbieter / Model



Zusätzliche Informationen

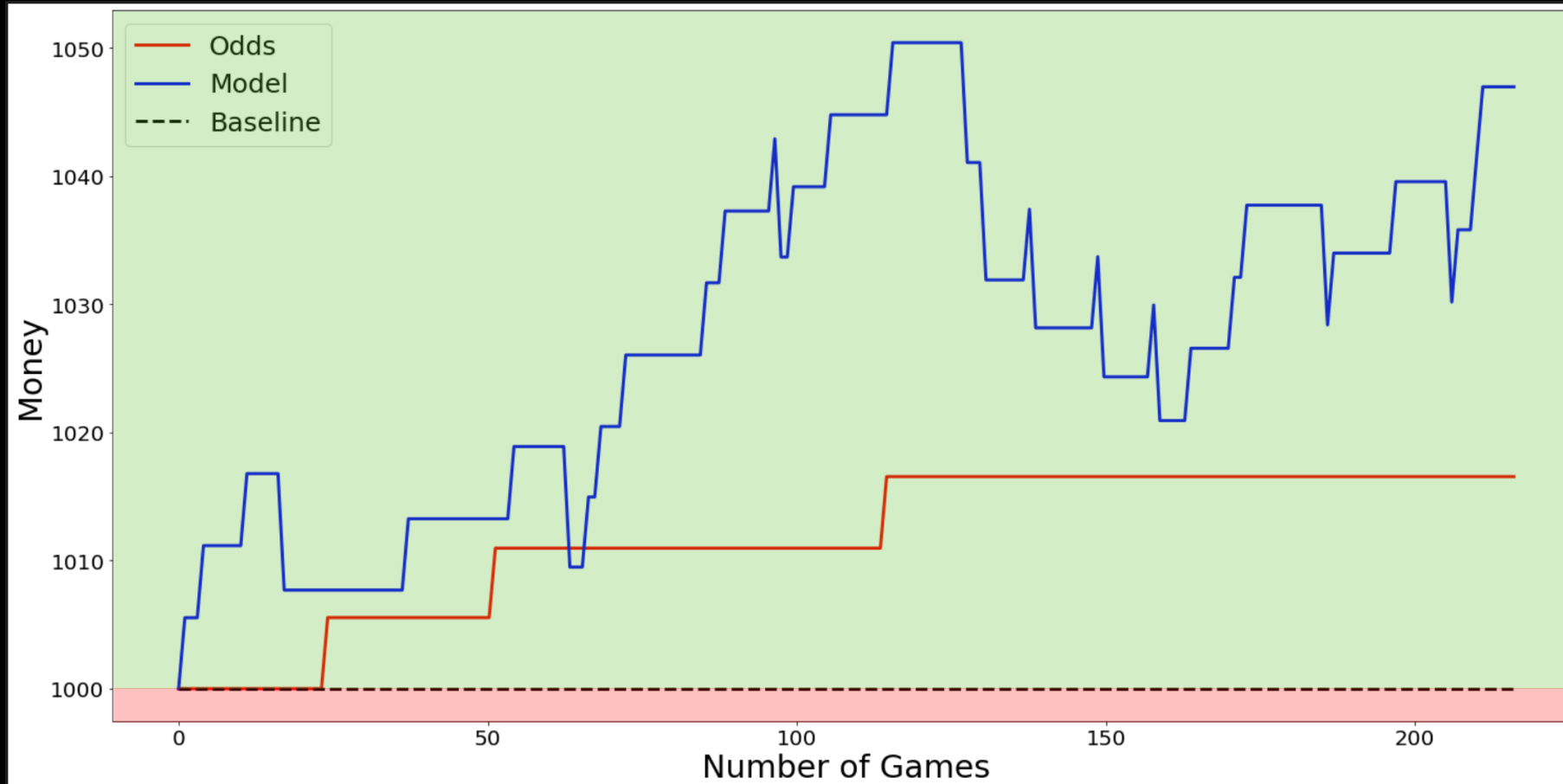
1. Wetten erst ab 40 %
2. 189 Wetten Modell
3. 204 Wetten Wettanbieter



Business Use Case (✓)

Business Use Case

Vergleich Wettanbieter / Model



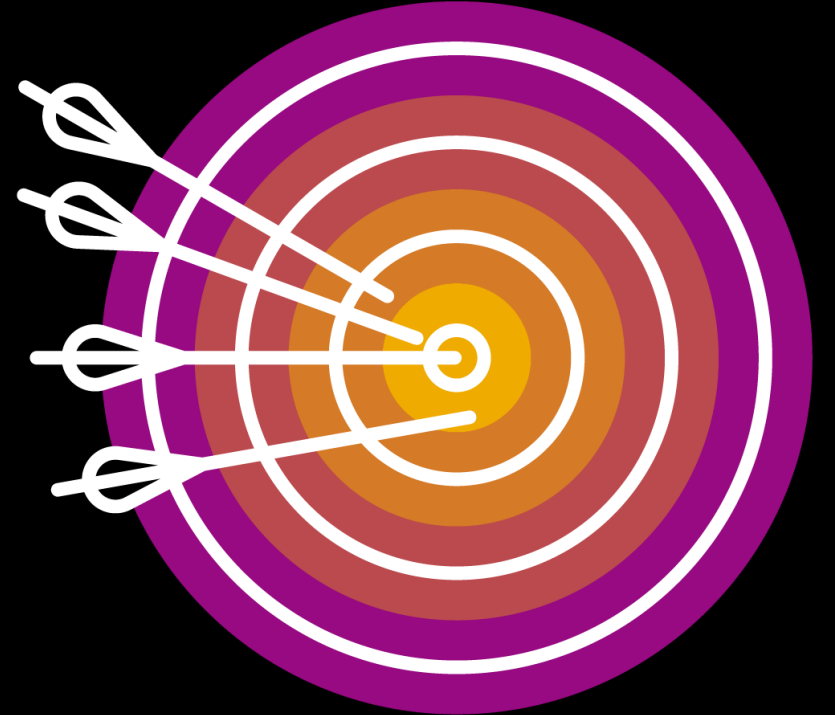
Zusätzliche Informationen

1. Wetten erst ab 90 %
2. 30 Wetten Modell
3. 3 Wetten Wettanbieter



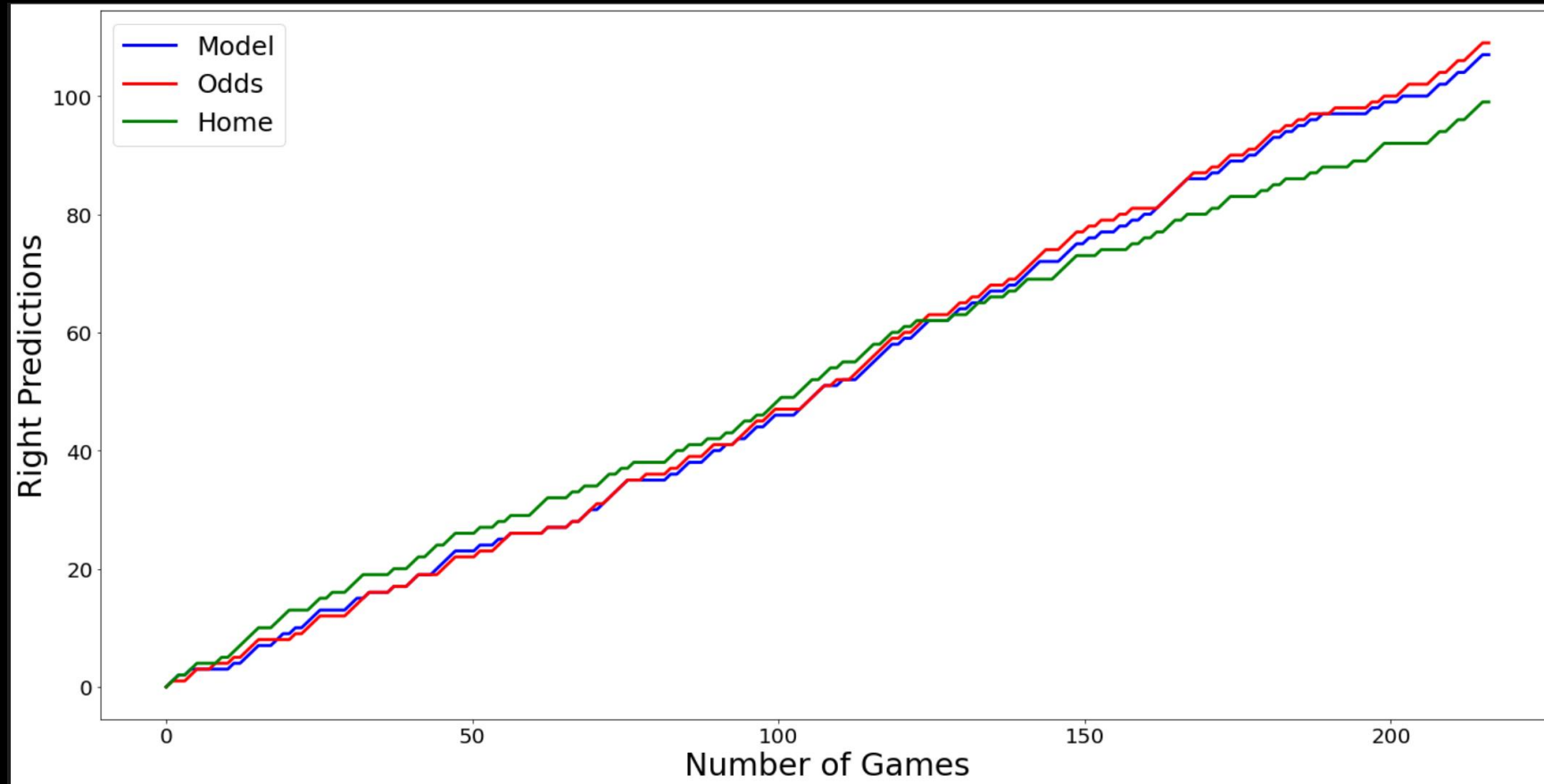
Business Use Case (✓)

Kritische Reflexion



Kritische Reflexion

Wurde das Ziel erreicht?



**Fußballergebnis schwer
genau vorherzusagen**

Kritische Reflektion

Resumé

Pro

- Herausfiltern von Komponenten die das Spielergebnis maßgeblich beeinflussen
- Zusammenstellen der Datenbasis
- Teilweise besseres Abschneiden als Vergleichsprojekte und Wettanbieter

Contra

- Accuracy vergleichsweise niedrig
 - Fußball nicht direkt vorhersehbar
 - Zu viele Zufallsvariablen für konstant richtige Vorhersagen
 - Weitere Features notwendig, aber schwer zugänglich
- Bei simulierten Wetten schnell im Minus

Thank you.

Contact information:

Max Bernauer, Julius Koenning, Philipp Dingfelder



Q & A

