Prediction of Soccer Penalty Kicks using R

Group 1

Aravind Aashish Kalyanaraman

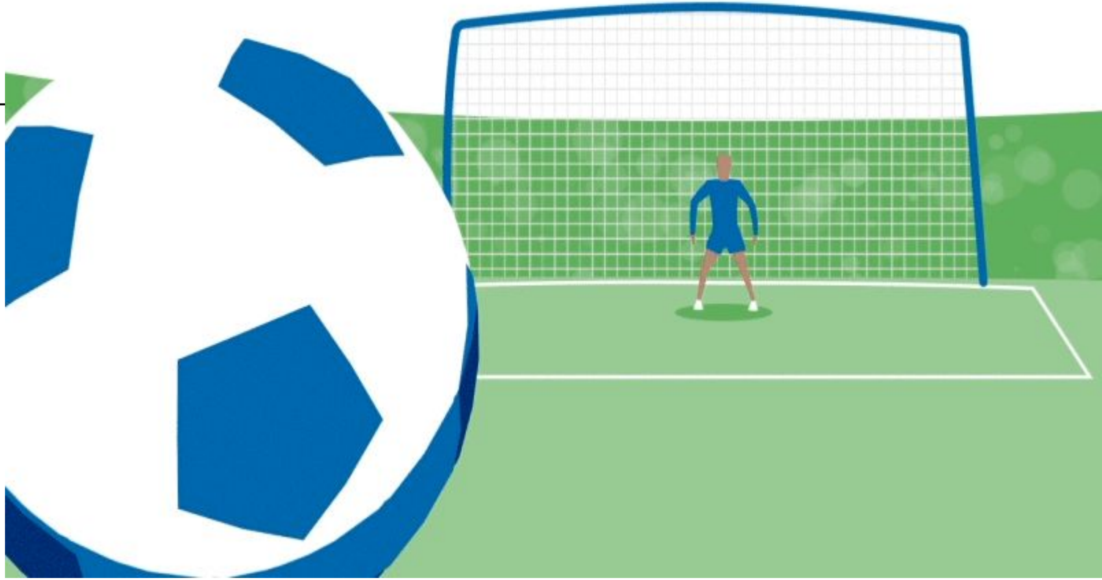
Nitish Aggarwal

Sukhman Singh

Analytics Project SoSe 2022
Prof. Dr. Sven Müller
August 5 2022

RWTH AACHEN UNIVERSITY

Introduction



IMPORTANCE OF PENALTIES IN FOOTBALL

50%

Probability of winning a match

52%

Probability of winning a match when the team has a ingame penalty

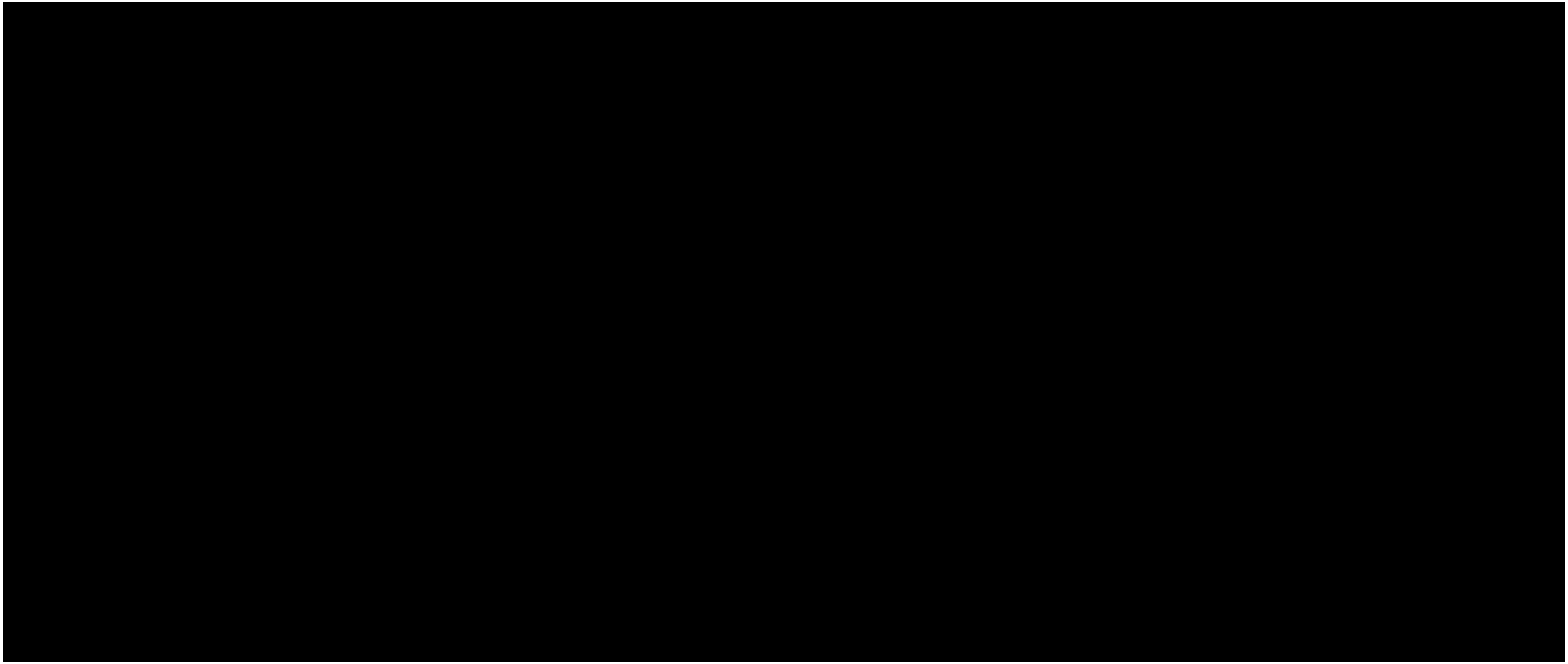
61%

Probability of winning a match when the team has a ingame penalty and scores it

[1]

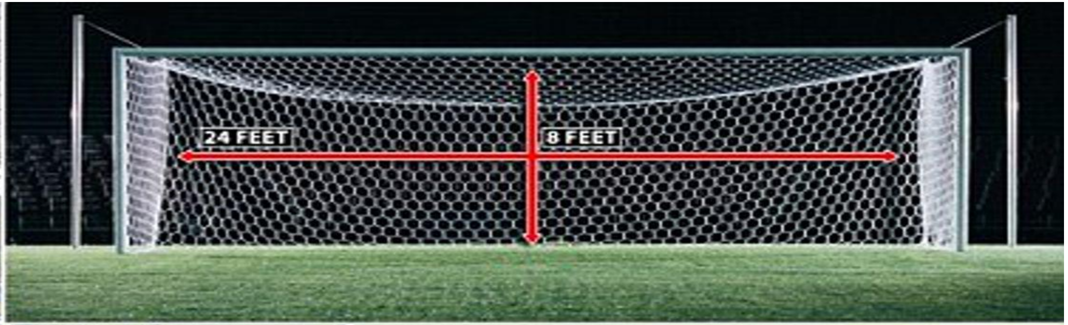
Source: [1] Dalton et al.(2015, p.815-827)

IMPORTANCE OF PENALTIES IN FOOTBALL



Source: Youtube

SCIENCE OF A PENALTY SHOOTOUT

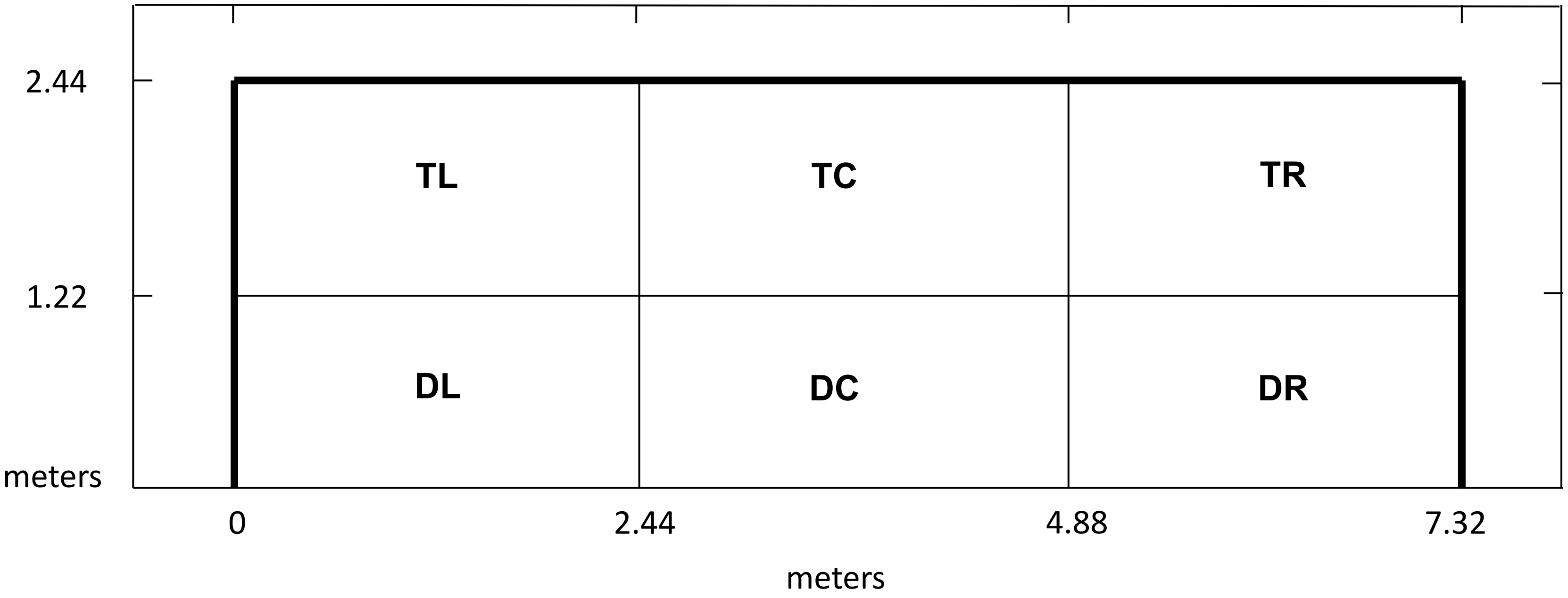


Source: daily mail

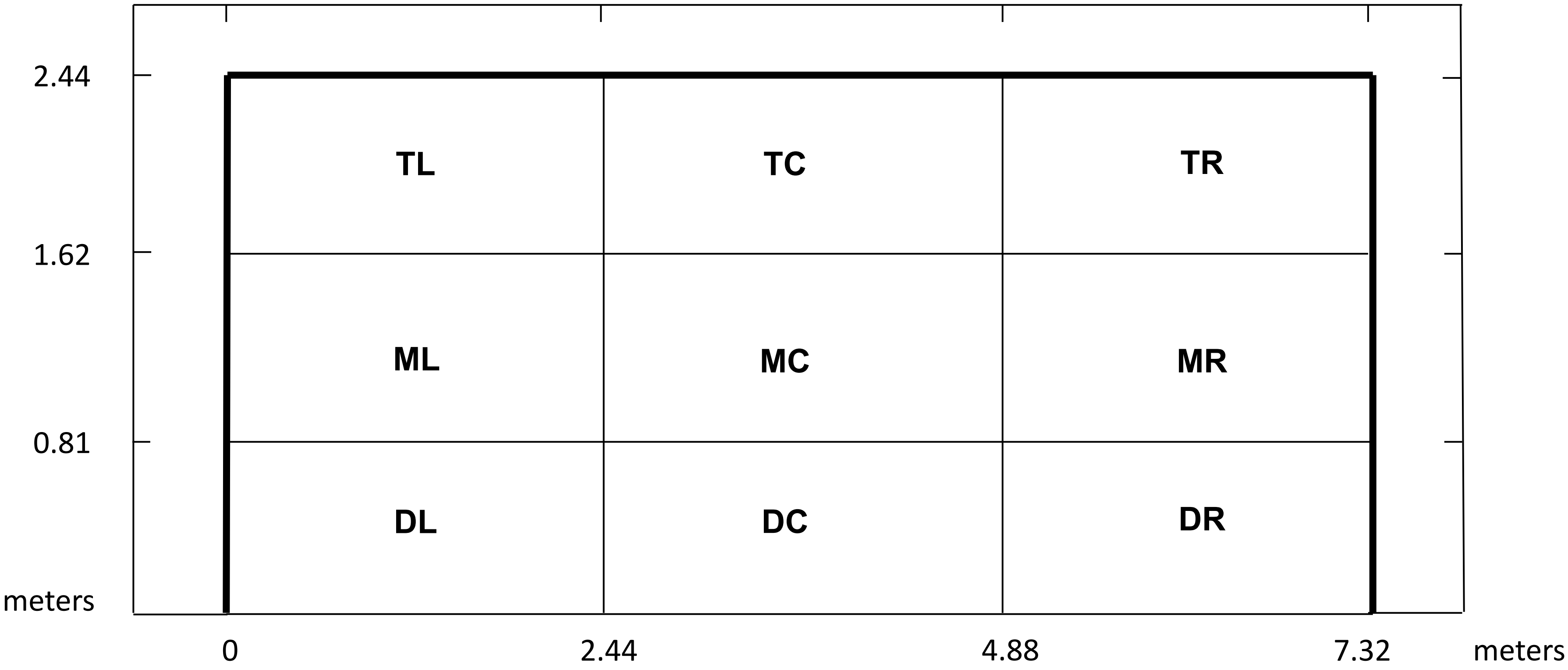
AGENDA

DEEP DIVING INTO ALTERNATIVES

6 ALTERNATIVE MODEL



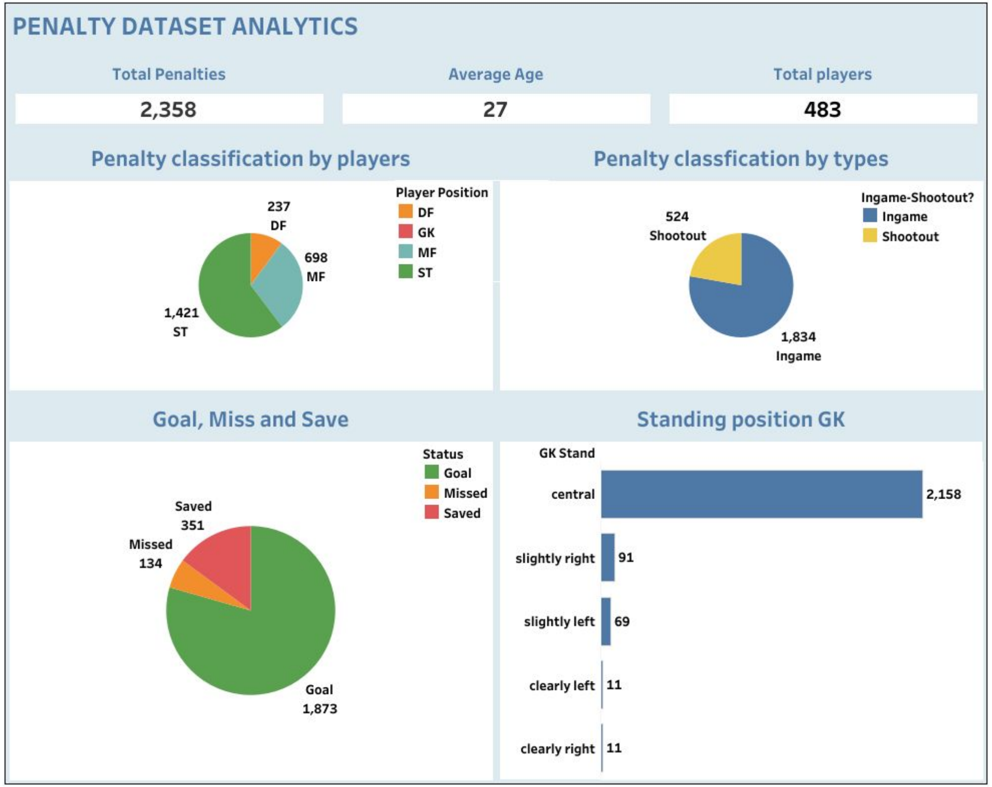
9 ALTERNATIVE MODEL



RWTH AACHEN UNIVERSITY

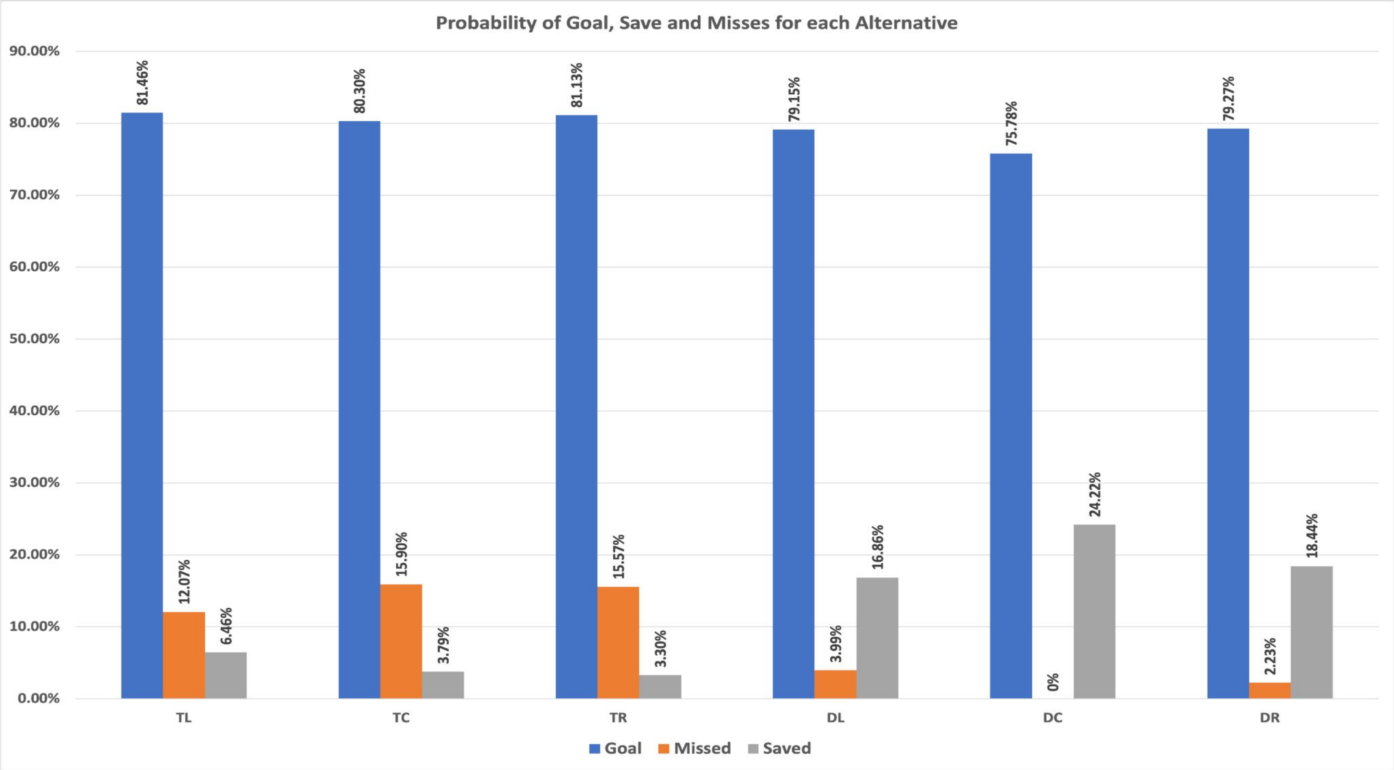
DATA ANALYSIS

DATA ANALYSIS DASHBOARD



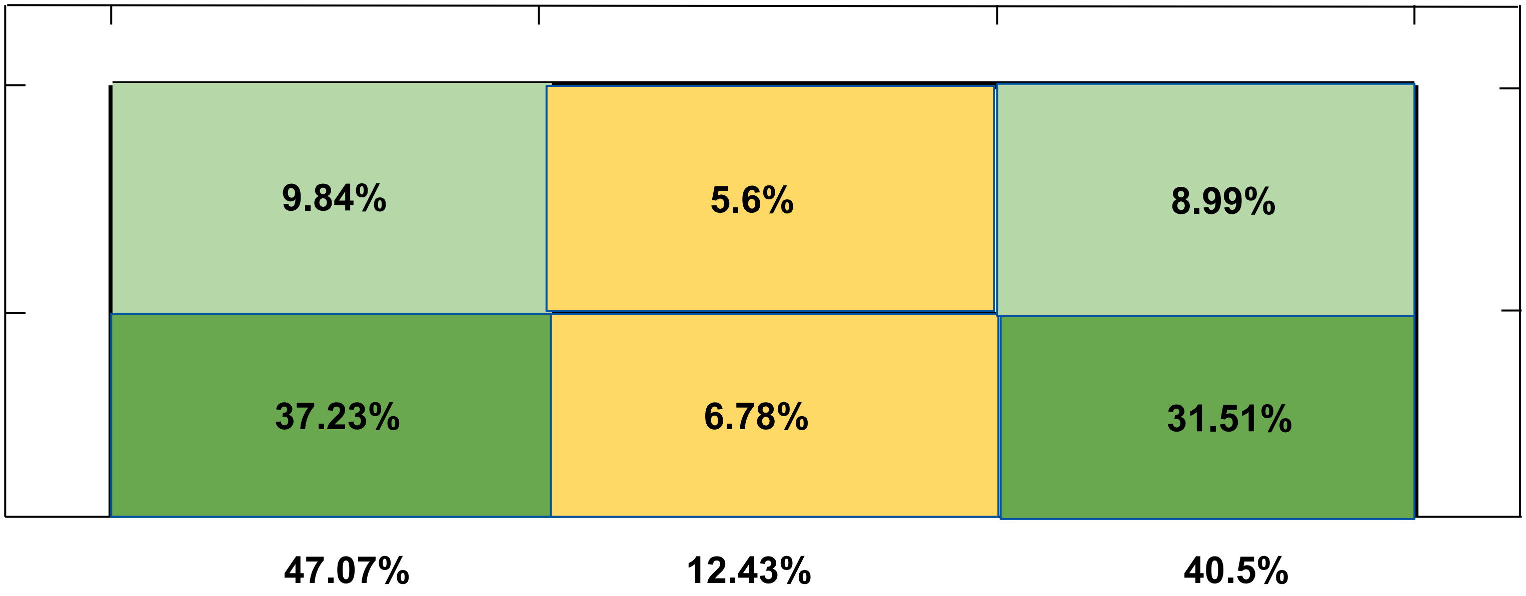
Reference: Penalty Dataset

PROBABILITY OF 'GOAL', 'SAVE' AND 'MISS' FOR EACH ALTERNATIVE



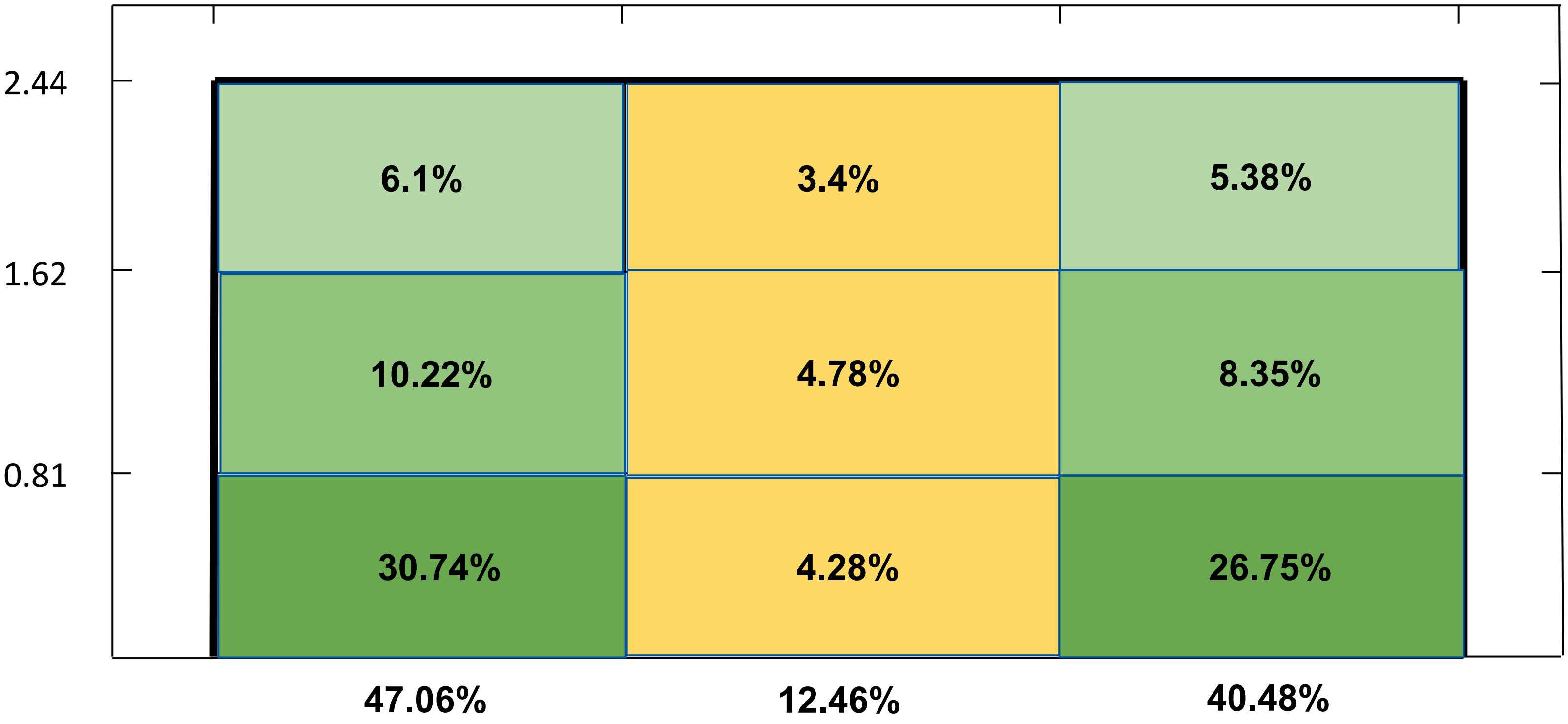
Reference: Penalty Dataset

Probability of a goal for a 6 alternative model



Reference: Penalty Dataset

Probability of a goal for a 9 alternative model



Reference: Penalty Dataset

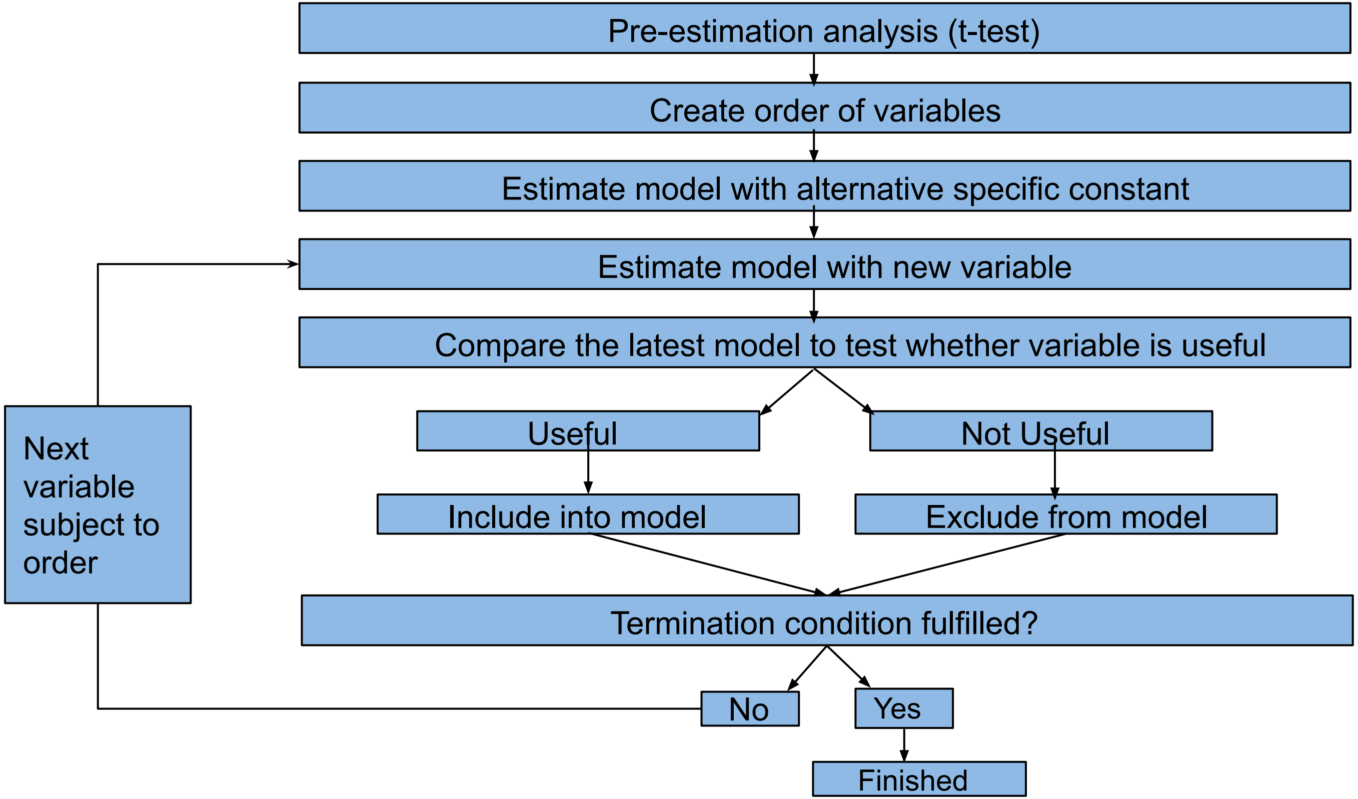
Data Dictionary

Variable	Description	Values
Date	Range of dates for the matches	Int Ex: 2004-2021
Player	Names of the Penalty Takers	String Ex: Aguero
Nation	Nationality of the penalty takers	String Ex: France
Player Position	Refers to the position which the player takes on the field	String : DF(Defender)/GK(Goalkeeper)/MF(Midfielder)/ST(Striker)
Player Foot	Whether the player is left footed or right footed	String : L(Left)/R(Right)
Age	Age of the players throughout different championships over the years (Age of a single player v	Int: 16 to 40
Age in groups	Age of the players in designated age groups	String:Under21, 21-33, Over33
GK(Goalkeeper)	Names of the Goalkeepers	String Ex: Buffon
Great GK?	Whether Goalkeeper is great	Boolean:Yes/No
GK already saved one	Whether Goalkeeper was successful in saving the goal	Boolean:Yes/No
Height of GK	Recorded heights of all goalkeepers (in cms)	Int: 176 to 203 (cms)
Height of GK (Grouped)	Recorded heights of all goalkeepers in designated groups	String:Under183, 183-196, Over196
Team	Clubs/National Teams that players belong to	String: Ex: Manchester City
Club/National Team	Whether players belong to a club or a national team for the games	String: Club/National Team
Table Position	Team standings	String: Upper/Middle/Bottom
Opponent	Opponent Team	String Ex: Real Madrid
Opponent Table position	Opponent Team standings	String: Upper/Middle/Bottom
Location(H-A-N)	Whether the match venue for the team is at home,away or neither	String: H(Home)/A(Away)/N(Neutral)
Round Number	Number for each round	String: Ex: Quarter-Final, Final
Competition	The name of the different football leagues /tournaments played	String: Ex: League, UCL
Special match?	Whether the match is a special match (For ex. world cup final)	Boolean:Yes/No

https://docs.google.com/spreadsheets/d/1RMvy-auTaM4igd7T6o49wOnpvPrP5qmbVs5BTxbBy_Y/edit#gid=0

METHODOLOGY

Steps involved



Pre-estimation analysis (t-test)

Explanators	TL	TC	TR	DL	DC	DR
Explanator 1 (foot_cat), mean when alt is chosen:	0.82	0.85	0.61	0.80	0.76	0.70
Explanator 1 (foot_cat), mean when alt is not chosen:	0.75	0.75	0.77	0.73	0.75	0.78
Explanator 1 (foot_cat), t-test (mean if chosen - mean if not chosen)	**2.65**	**3.05**	**-4.61**	**4.35**	0.10	**-3.94**

Create order of variables

2 values > 1.96

- foot
- mvmtGK
- standposGK
- ingameshoot
- compleague
- lastpenaltygoal
- opplstpengoal
- shothardness
- tappedball
- lastGKdive
- clubornational
- compNational
- secondlastGKdivedir

1 value > 1.96

- heightGK
- decider
- GKmvmttyp
- compCup
- jerseyblueGK
- jerseygoldPL
- jerseypinkPL
- jerseypurplePL
- jerseyredPL
- jerseygreenPL
- secondlastpen

No Value > 1.96

- gameimportance
- specialmatch
- compECC
- compQuali
- jerseygreyGK
- jerseyredGK
- jerseyblackGK
- jerseygreenGK
- jerseyyellowGK
- jerseyorangeGK
- jerseyblackPL
- jerseywhitePL
- posdef
- roundnogroup

MODEL ESTIMATION

Multinomial Logit Model

Logit Model is a mathematical model used to estimate the probability of an event occurring (choice) having been given some previous data.

Multinomial Logit Model is a classification method that generalizes logistic regression and is best suited for :

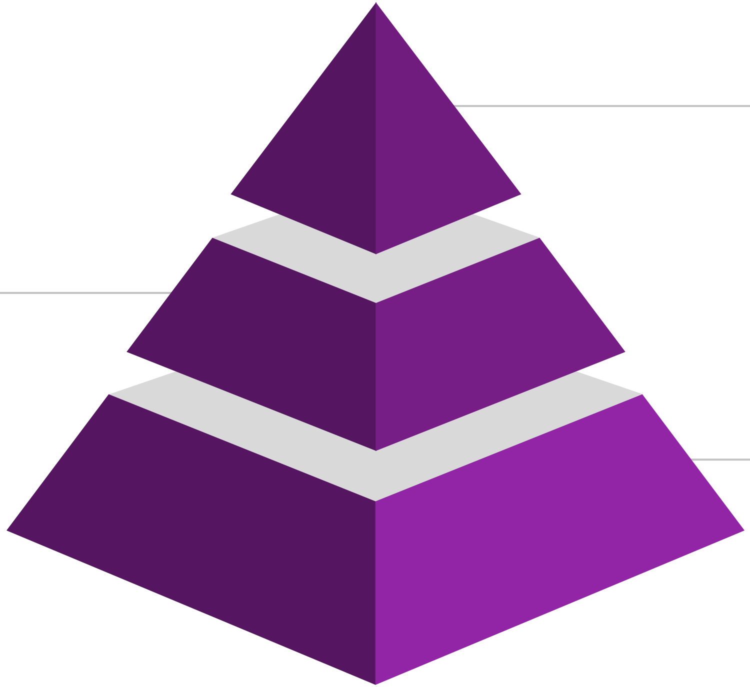
- Modelling multiclass problems, i.e. with more than two possible discrete outcomes.
- Useful when the dependent variable in question is nominal.

$$
\begin{aligned}
U_{nC} &= V_{nC} + \epsilon_{nC} & &= \beta_0 + \beta_1 \mathsf{TC}_{nC} + \epsilon_{nC} \\
U_{nT} &= V_{nT} + \epsilon_{nT} & &= \beta_2 + \beta_1 \mathsf{TC}_{nT} + \epsilon_{nT} \\
U_{nB} &= V_{nB} + \epsilon_{nB} & &= \qquad \beta_1 \mathsf{TC}_{nB} + \epsilon_{nB}
\end{aligned}
$$

$$
P_n\left(C \,|\{C, T, B\}\right) = \frac{e^{\mu V_{nC}}}{e^{\mu V_{nC}} + e^{\mu V_{nT}} + e^{\mu V_{nB}}}
$$

Reference: Wikipedia

For every model, ….



Estimate model with alternative specific constant

1 **Utility Equation**

$$V(TL) = ASC_{TL}$$
$$V(TC) = ASC_{TC}$$
$$V(TR) = ASC_{TR}$$
$$V(DL) = ASC_{DL}$$
$$V(DC) = ASC_{DC}$$
$$V(DR) = ASC_{DR}$$

Estimate model with alternative specific constant

2 **Log likelihood values**

LL(start)	-4224.97
LL(0)	-4224.97
LL(C)	-3586.82
LL(final)	-3586.82
Rho-square (0)	0.15
Adj.Rho-square (0)	0.15
Rho-square (c)	0.00
Adj.Rho-square (c)	0.00
AIC	7183.63
BIC	7212.46

Estimate model with alternative specific constant

3 Estimates, Standard error and t-ratio

Alternatives	Estimate	s.e.	t.rat.(0)	Rob.s.e.	Rob.t.rat.(0)
asc_TL	0.37	0.10	3.56	0.10	3.56
asc_TC	-0.20	0.12	-1.69	0.12	-1.69
asc_TR	0.28	0.10	2.63	0.10	2.63
asc_DL	1.70	0.09	19.79	0.09	19.78
asc_DC	0.00	0.00	0.00	0.00	0.00
asc_DR	1.53	0.09	17.59	0.09	17.59

Estimate model with new variable

1 Utility Equation

$$V(TL) = ASC_{TL} + \beta_{foot_{TL}} * foot\_cat$$
$$V(TC) = ASC_{TC} + \beta_{foot_{TC}} * foot\_cat$$
$$V(TR) = ASC_{TR} + \beta_{foot_{TR}} * foot\_cat$$
$$V(DL) = ASC_{DL} + \beta_{foot_{DL}} * foot\_cat$$
$$V(DC) = ASC_{DC} + \beta_{foot_{DC}} * foot\_cat$$
$$V(DR) = ASC_{DR} + \beta_{foot_{DR}} * foot\_cat$$

Estimate model with new variable

② **Log likelihood values**

LL(start)	-4224.97
LL(0)	-4224.97
LL(C)	-3586.82
LL(final)	-3558.25
Rho-square (0)	0.16
Adj.Rho-square (0)	0.16
Rho-square (c)	0.01
Adj.Rho-square (c)	0.01
AIC	7136.49
BIC	7194.15

Estimate model with new variable

3 **Estimates, Standard error and t-ratio**

Alternatives	Estimate	s.e.	t.rat.(0)	Rob.s.e.	Rob.t.rat.(0)
asc_TL	0.07	0.22	0.33	0.22	0.33
asc_TC	-0.67	0.28	-2.43	0.28	-2.43
asc_TR	0.76	0.19	3.89	0.19	3.89
asc_DL	1.49	0.18	8.40	0.18	8.40
asc_DC	0.00	0.00	0.00	0.00	0.00
asc_DR	1.74	0.17	10.02	0.17	10.01
b_foot1	0.37	0.25	1.47	0.25	1.47
b_foot2	0.58	0.30	1.91	0.30	1.91
b_foot3	-0.70	0.23	-3.02	0.23	-3.02
b_foot4	0.26	0.20	1.31	0.20	1.31
b_foot5	0.00	0.00	0.00	0.00	0.00
b_foot6	-0.29	0.20	-1.43	0.20	-1.43

Compare the latest model to test whether variable is useful

Model	LL	par
Null model	-3586.82	5
Model with foot	-3558.24	10
Difference	28.58	5

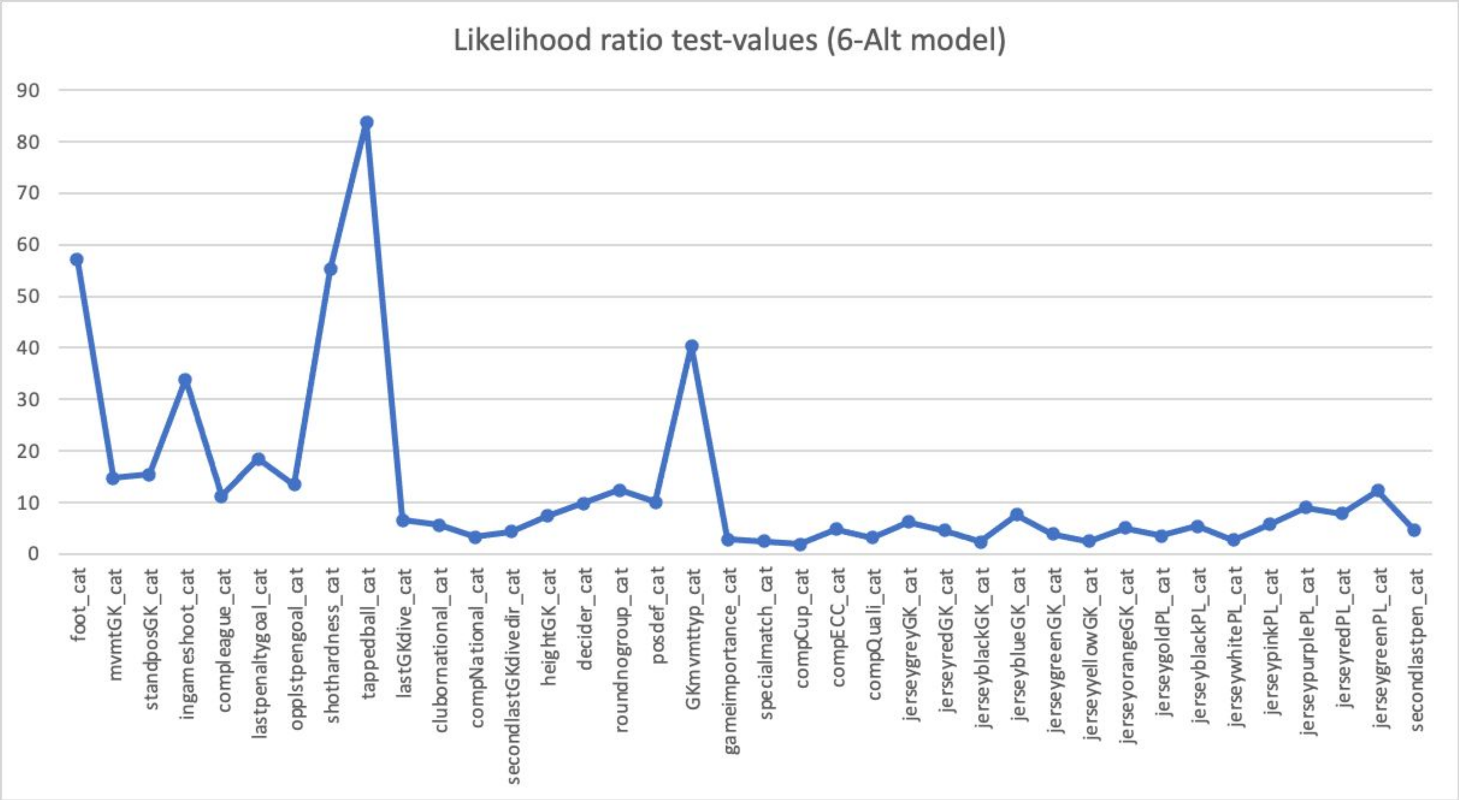
Level of confidence number of restrictions	90%	95%	99%
1	2.71	3.84	6.63
2	4.61	5.99	11.34
5	9.24	11.07	15.09

[1]

Likelihood ratio test-value:	57.16
Degrees of freedom:	5
Likelihood ratio test p-value:	4.69E-11

Reference: mbounthavong.com [1]

Feature extraction graph analysis (6 Alt model)



Data Dictionary for 6 Alt model

Variable	Equivalent feature variable	Description
foot_cat	Foot	strong foot of penalty-taker. Coding; 0: Left, 1: Right
mvmtGK_cat	moveGK	movement of goalkeeper during run-up of penalty-taker. Coding; 0: no/null, 1: movement to left , 2: jumping left and right, 2: jumping on spot, 2: step forward, 3: movement to right
standposGK_cat	GKS	standing position of goalkeeper during penalty kick. Coding; 1: clearly or slightly left, 2: central, 3:clearly or slightly right, 0/NA: not available
ingameshoot_cat	Ingso	whether penalty is shot in-match or in a penalty shootout. Coding; 0: Ingame, 1: shootout
compleague_cat	Competition	whether team of penalty-taker is playing in a "league" match or not. Coding; 0: no, 1: yes
lastpenaltygoal_cat	lpbg	whether last penalty of penalty-taker was a goal, missed or saved . Coding; 1: 'missed' or 'saved', 2: 'goal'
opplstpengoal_cat	Solsbgopponent	whether last shot penalty of the opponent of the penalty-taker in a penalty shootout was a goal, missed or saved. Coding; 1: 'goal', 2: 'missed', 0:'saved'
shothardness_cat	Shothardness	whether the penalty shot by the penalty-taker is classified as strong/normal/weak. Coding; 1:'strong', 2:'normal', 3:'weak'
tappedball_cat	tappedtheball	whether the goalkeeper tapped the ball during the penalty shootout. Coding; 1:'yes',0:'no'
roundnogroup_cat	rnoGroup	whether team of penalty-taker is playing in a group stage match or not. Coding; 0: no, 1: yes
GKmvmttyp_cat	nmovGK	whether goalkeeper is moving when the penalty is hit by penalty-taker. Coding; 0/NA: not available, 1: moving, 2: still

Final 6 Alt Model after Feature Selection

2 **Log likelihood values**

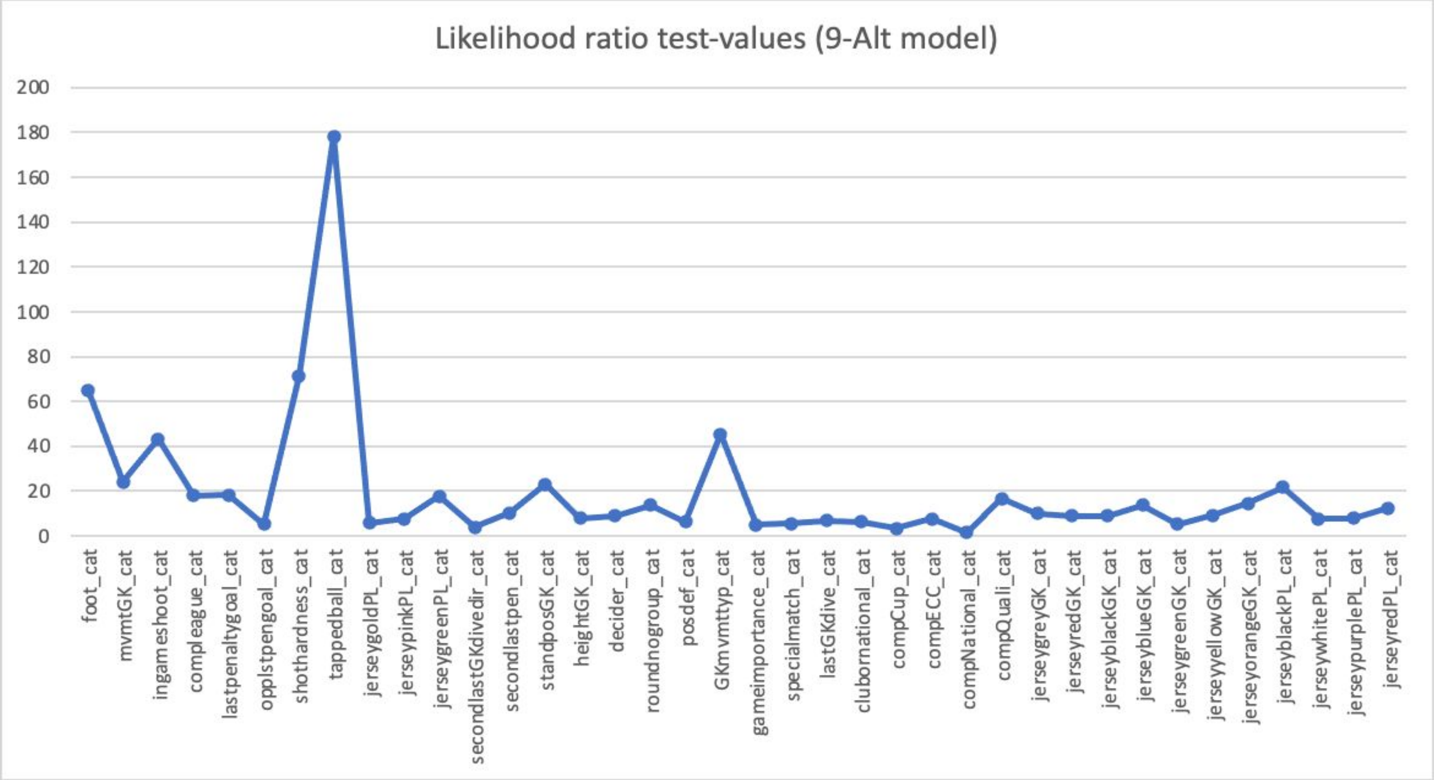
LL(start)	-4224.97
LL(0)	-4224.97
LL(C)	-3586.82
LL(final)	-3395.11
Rho-square (0)	0.1964
Adj.Rho-square (0)	0.181
Rho-square (c)	0.0534
Adj.Rho-square (c)	0.0353
AIC	6920.21
BIC	7294.97

FINAL MODEL AFTER FEATURE SELECTION (6 ALT)

	Estimate	s.e.	t.rat.(0)	Rob.s.e.	Rob.t.rat.(0)
asc_TL	0.44	1.18	0.37	1.08	0.41
asc_TC	0.14	1.30	0.11	1.49	0.10
asc_TR	1.67	1.15	1.46	1.09	1.54
asc_DL	1.71	0.97	1.76	0.96	1.78
asc_DC	0.00	NA	NA	NA	NA
asc_DR	3.28	0.96	3.42	0.92	3.58
b_foot1	0.39	0.25	1.55	0.25	1.56
b_foot2	0.61	0.31	1.97	0.31	1.98
b_foot3	-0.73	0.24	-3.09	0.23	-3.10
b_foot4	0.31	0.21	1.53	0.20	1.54
b_foot5	0.00	NA	NA	NA	NA
b_foot6	-0.27	0.20	-1.34	0.20	-1.35
b_mvmtGK1	0.08	0.22	0.38	0.22	0.39
b_mvmtGK2	-0.07	0.25	-0.27	0.27	-0.25
b_mvmtGK3	0.14	0.22	0.62	0.22	0.62
b_mvmtGK4	0.34	0.18	1.87	0.19	1.81
b_mvmtGK5	0.00	NA	NA	NA	NA
b_mvmtGK6	-0.31	0.19	-1.62	0.19	-1.58

https://docs.google.com/spreadsheets/d/1b5zdbrU7qKk0EHTXGprB9SEaf3VgbvDyuCBwCRmo8aQ/edit?usp=sharing

Feature extraction graph analysis (9 Alt model)



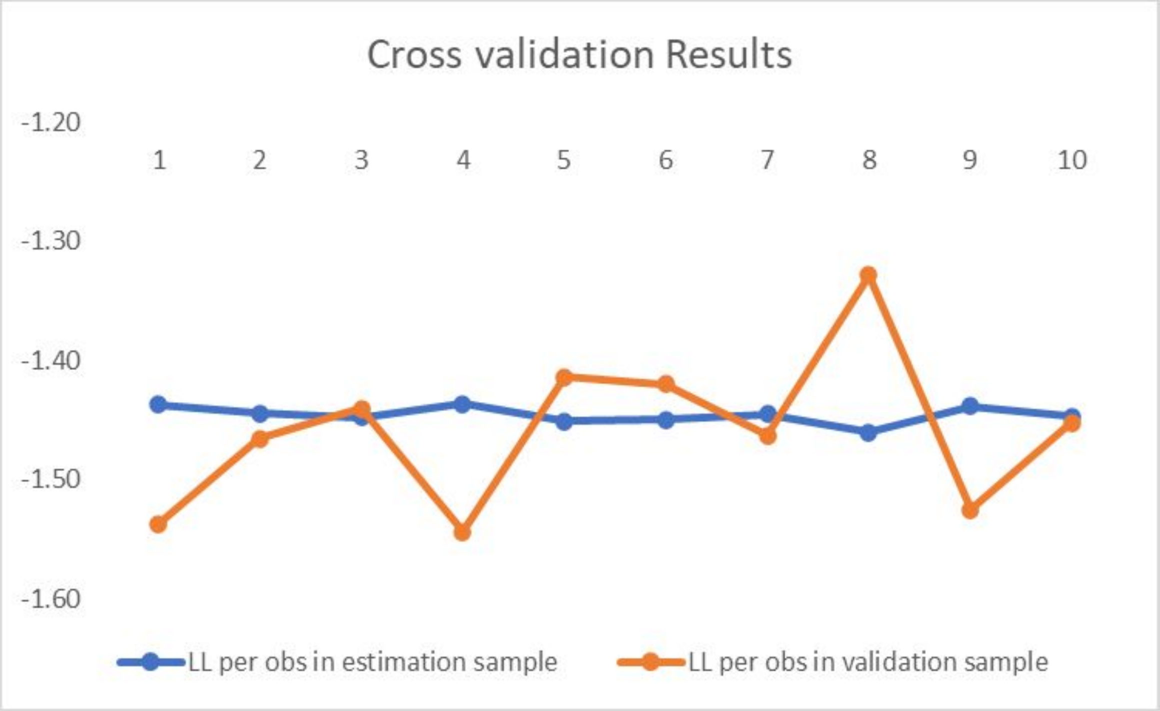
RWTH AACHEN UNIVERSITY

Data Dictionary for 9 Alt model

Variable	Equivalent feature variable	Description
foot_cat	Foot	strong foot of penalty-taker. Coding; 0: Left, 1: Right
mvmtGK_cat	moveGK	movement of goalkeeper during run-up of penalty-taker. Coding; 0: no/null, 1: movement to left , 2: jumping left and right, 2: jumping on spot, 2: step forward, 3: movement to right
ingameshoot_cat	Ingso	whether penalty is shot in-match or in a penalty shootout. Coding; 0: Ingame, 1: shootout
compleague_cat	Competition	whether team of penalty-taker is playing in a ”league” match or not. Coding; 0: no, 1: yes
lastpenaltygoal_cat	lpbg	whether last penalty of penalty-taker was a goal, missed or saved . Coding; 1: 'missed' or 'saved', 2: 'goal'
shothardness_cat	Shothardness	whether the penalty shot by the penalty-taker is classified as strong/normal/weak. Coding; 1:'strong', 2:'normal', 3:'weak'
tappedball_cat	tappedtheball	whether the goalkeeper tapped the ball during the penalty shootout. Coding; 1:'yes',0:'no'
standposGK_cat	GKS	standing position of goalkeeper during penalty kick. Coding; 1: clearly or slightly left, 2: central, 3:clearly or slightly right, 0/NA: not available
GKmvmttyp_cat	nmovGK	whether goalkeeper is moving when the penalty is hit by penalty-taker. Coding; 0/NA: not available, 1: moving, 2: still
compQuali_cat	Competition	whether team of penalty-taker is playing in a qualification match or not

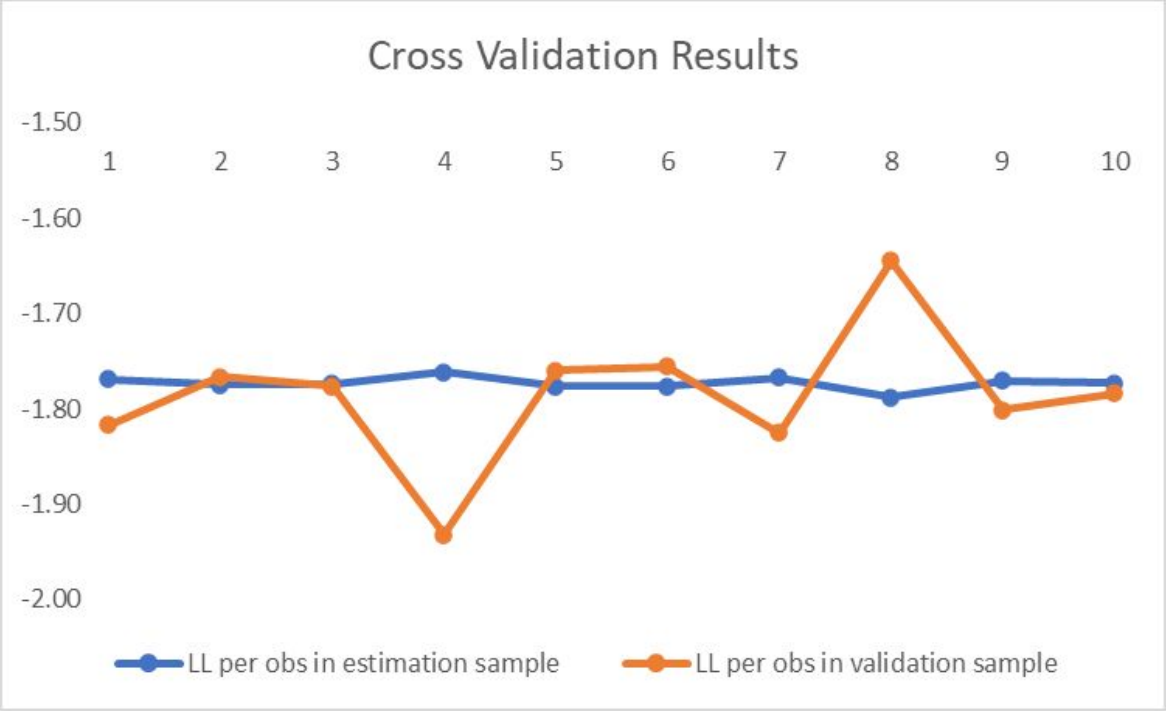
Cross Validation results of the final model (6 alt)

Estimation Sample (LL per observation)	Validation Sample (LL per observation)
-1.44	-1.54
-1.44	-1.47
-1.45	-1.44
-1.44	-1.54
-1.45	-1.41
-1.45	-1.42
-1.45	-1.46
-1.46	-1.33
-1.44	-1.53
-1.45	-1.45



Cross Validation results of the final model (9 alt)

Estimation Sample (LL per observation)	Validation Sample (LL per observation)
-1.77	-1.82
-1.77	-1.77
-1.77	-1.78
-1.76	-1.93
-1.78	-1.76
-1.78	-1.76
-1.77	-1.83
-1.79	-1.64
-1.77	-1.80
-1.77	-1.78

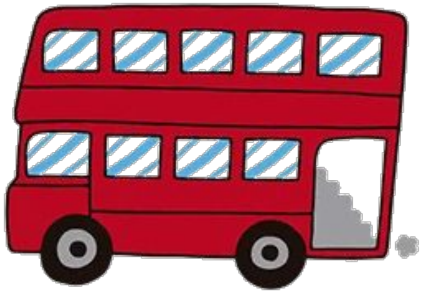


IIA VERIFICATION

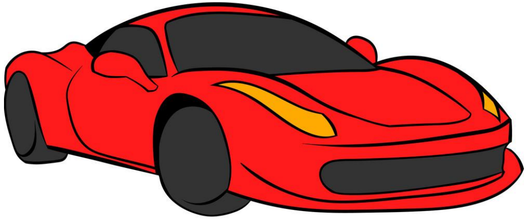
Independence of Irrelevant Alternatives (IIA)

Independence of irrelevant alternatives is the property stating that any item added to the set of choices will decrease all other items’ likelihood by an equal fraction.

Red Bus - Blue Bus Paradox

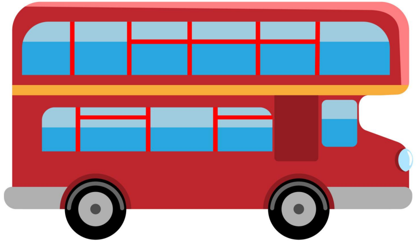


Independence of Irrelevant Alternatives (IIA)



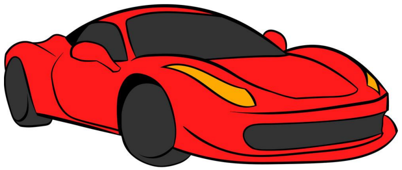
$$U_{car} = \beta_{TT} + \epsilon_{car}$$

$$U_{bus} = \beta_{TT} + \epsilon_{bus}$$

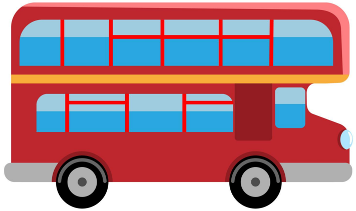


$$P(car|car, bus\} = \left(e^{\beta_{TT}} / \left(e^{\beta_{TT}} + e^{\beta_{TT}}\right)\right)$$

Independence of Irrelevant Alternatives (IIA)

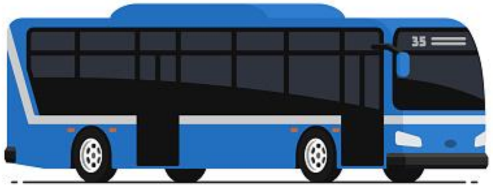


$$U_{car} = \beta_{TT} + \epsilon_{car}$$



$$U_{bluebus} = \beta_{TT} + \epsilon_{bluebus}$$

$$U_{redbus} = \beta_{TT} + \epsilon_{redbus}$$



$$P(car|car, bluebus, redbus\} = \left(e^{\beta_{TT}}/\left(e^{\beta_{TT}} + e^{\beta_{TT}} + e^{\beta_{TT}}\right)\right)$$

IIA for the 6 Alt model

b_foot1	0.36886	0.2508	1.4706	0.2509	1.4703		
b_foot2	0.58231	0.3046	1.9119	0.3046	1.9114		
b_foot3	-0.69952	0.2316	-3.0204	0.2316	-3.0197		
b_foot4	0.26442	0.2026	1.3053	0.2026	1.305		
b_foot5	0	0	0	0	0		
b_foot6	-0.28742	0.2007	-1.4324	0.2007	-1.4321		

Model without TC alternative

Alternativ	Estima	standard e	t.rat.(	Rob.s	Rob.t.rat	T Ratio calcul	IIA verificat
asc_TL	0.07416	0.2224	0.3335	0.2224	0.3334		
asc_TR	0.75529	0.1941	3.8905	0.1942	3.8896		
asc_DL	1.48974	0.1773	8.4042	0.1773	8.4022		
asc_DC	0	0	0	0	0		
asc_DR	1.73913	0.1736	10.0165	0.1737	10.0142		
b_foot1	0.36885	0.2508	1.4706	0.2509	1.4703	-3.98724E-05	IIA holds
b_foot3	-0.69949	0.2316	-3.0203	0.2317	-3.0196	0.000129534	IIA holds
b_foot4	0.26444	0.2026	1.3054	0.2026	1.3051	9.87167E-05	IIA holds
b_foot5	0	0	0	0	0	0	
b_foot6	-0.28739	0.2007	-1.4323	0.2007	-1.432	0.000149477	IIA holds

Model without DL alternative

Alternativ	Estima	standard e	t.rat.(	Rob.s	Rob.t.rat	T Ratio calcul	IIA verificat
asc_TL	0.07411	0.2224	0.3333	0.2225	0.3332		
asc_TC	-0.66783	0.275	-2.4282	0.2751	-2.4274		
asc_TR	0.75529	0.1941	3.8905	0.1942	3.8892		
asc_DC	0	0	0	0	0		
asc_DR	1.73912	0.1736	10.0165	0.1737	10.0131		
b_foot1	0.36889	0.2508	1.4707	0.2509	1.4703	0.000119617	IIA holds
b_foot2	0.58231	0.3046	1.9119	0.3047	1.9112	0	IIA holds
b_foot3	-0.69949	0.2316	-3.0203	0.2317	-3.0193	0.000129534	IIA holds
b_foot5	0	0	0	0	0	0	
b_foot6	-0.28739	0.2007	-1.4323	0.2007	-1.4318	0.000149477	IIA holds

Model without DR alternative

Alternativ	Estima	standard e	t.rat.(	Rob.s	Rob.t.rat	T Ratio calcul	IIA verificat
asc_TL	0.07415	0.2224	0.3334	0.2224	0.3333		
asc_TC	-0.66776	0.275	-2.428	0.2751	-2.4272		

b_foot3	-0.6995	0.2316	-3.02	0.2317	-3.019	8.63558E-05	IIA holds
b_foot4	0.2645	0.2026	1.306	0.2026	1.305	0.000394867	IIA holds
b_foot5	0	0	0	0	0	0	
b_foot6	-0.2874	0.2007	-1.432	0.2007	-1.432	9.96512E-05	IIA holds

Model without TR alternative

Alternative	Estima	standard e	t.rat.(0)	Rob.s.e.	Rob.t.rat.(0	T Ratio calcul	IIA verificat
asc_TL	0.0741	0.2224	0.3332	0.2224	0.3331		
asc_TC	-0.66781	0.275	-2.4281	0.2751	-2.4276		
asc_DL	1.48973	0.1773	8.4042	0.1773	8.4022		
asc_DC	0	0	0	0	0		
asc_DR	1.73912	0.1736	10.0165	0.1737	10.0142		
b_foot1	0.36891	0.2508	1.4708	0.2509	1.4705	0.000199362	IIA holds
b_foot2	0.58229	0.3046	1.9118	0.3046	1.9114	-6.56599E-05	IIA holds
b_foot4	0.26445	0.2026	1.3054	0.2026	1.3051	0.000148075	IIA holds
b_foot5	0	0	0	0	0	0	
b_foot6	-0.28739	0.2007	-1.4322	0.2007	-1.4319	0.000149477	IIA holds

Model without DC alternative

Alternative	Estima	standard e	t.rat.(0)	Rob.s.e.	Rob.t.rat.(0	T Ratio calcul	IIA verificat
asc_TL	0.7421	0.2717	2.7315	0.2717	2.731		
asc_TC	0	0	0	0	0		
asc_TR	1.423	0.2491	5.7128	0.2491	5.7116		
asc_DL	2.1575	0.2362	9.1352	0.2362	9.1334		
asc_DR	2.4069	0.2335	10.3098	0.2335	10.3077		
b_foot1	-0.2135	0.2966	-0.7197	0.2967	-0.7196	-1.963452461	IIA does not hold
b_foot2	0	0	0	0	0	0	
b_foot3	-1.2817	0.2806	-4.568	0.2806	-4.567	-2.074768354	IIA does not hold
b_foot4	-0.3178	0.2571	-1.2357	0.2572	-1.2355	-2.264566317	IIA does not hold
b_foot6	-0.8696	0.2556	-3.4017	0.2557	-3.401	-2.277699531	IIA does not hold

https://docs.google.com/spreadsheets/d/1pLtspuIP6GrOXdavZPqDZorn-hLZQ0UBZGNJmUKFvHs/edit#gid=1519839050

RWTH AACHEN UNIVERSITY

IIA for the 9 alt model

9 Alt model with all alts

Alternati	Estimate	standard e	t.rat.(	Rob.s	Rob.t.rat
b_foot1	0.30	0.34	0.89	0.34	0.89
b_foot2	0.27	0.40	0.68	0.40	0.68
b_foot3	-1.16	0.30	-3.86	0.30	-3.86
b_foot4	-0.91	0.28	-3.22	0.28	-3.22
b_foot5	0.00	0.00	0.00	0.00	0.00
b_foot6	-0.15	0.29	-0.52	0.29	-0.52
b_foot7	-0.06	0.26	-0.24	0.26	-0.24
b_foot8	-0.41	0.33	-1.22	0.33	-1.22
b_foot9	-0.54	0.26	-2.10	0.26	-2.10

Without TL

Alternati	Estimate	standard e	t.rat.(	Rob.s	Rob.t.rat	T Ratio calcul	IIA verificati
b_foot2	0.27	0.40	0.68	0.40	0.68	-0.000555836	IIA holds
b_foot3	-1.16	0.30	-3.87	0.30	-3.86	-6.6357E-05	IIA holds
b_foot4	-0.15	0.29	-0.52	0.29	-0.52	2.643475251	IIA does not hold
b_foot5	0.00	0.00	0.00	0.00	0.00		IIA holds
b_foot6	-0.91	0.28	-3.22	0.28	-3.22	-2.691046881	IIA does not hold
b_foot7	-0.06	0.26	-0.24	0.26	-0.24	0.000231214	IIA holds
b_foot8	-0.41	0.33	-1.23	0.33	-1.22	-0.00027084	IIA holds
b_foot9	-0.54	0.26	-2.10	0.26	-2.10	0.000310438	IIA holds

Without TC

Alternatives	Estimate	standard erro	t.rat.(0)	Rob.s.e.	Rob.t.rat.(0	T Ratio calculati	IIA verification
b_foot1	0.30	0.34	0.89	0.34	0.89	-2.95683E-05	IIA holds
b_foot3	-1.16	0.30	-3.87	0.30	-3.86	-0.000165893	IIA holds
b_foot4	-0.15	0.29	-0.52	0.29	-0.52	2.643406023	IIA does not hold
b_foot5	0.00	0.00	0.00	0.00	0.00		IIA holds
b_foot6	-0.91	0.28	-3.22	0.28	-3.22	-2.691223123	IIA does not hold
b_foot7	-0.06	0.26	-0.24	0.26	-0.24	-3.85356E-05	IIA holds
b_foot8	-0.41	0.33	-1.22	0.33	-1.22	-6.01866E-05	IIA holds
b_foot9	-0.54	0.26	-2.10	0.26	-2.10	7.76096E-05	IIA holds

Without TR

Alternatives	Estimate	standard erro	t.rat.(0)	Rob.s.e.	Rob.t.rat.(0	T Ratio calculati	IIA verification
b_foot1	0.30	0.34	0.89	0.34	0.89	-0.000680071	IIA holds
b_foot2	0.27	0.40	0.68	0.40	0.68	-0.000505306	IIA holds
b_foot4	-0.15	0.29	-0.52	0.29	-0.52	2.643094496	IIA does not hold
b_foot5	0.00	0.00	0.00	0.00	0.00		IIA holds
b_foot6	-0.91	0.28	-3.22	0.28	-3.22	-2.691857596	IIA does not hold
b_foot7	-0.06	0.26	-0.24	0.26	-0.24	-0.000423892	IIA holds
b_foot8	-0.41	0.33	-1.23	0.33	-1.22	-0.000541679	IIA holds
b_foot9	-0.54	0.26	-2.10	0.26	-2.10	-0.000349243	IIA holds

https://docs.google.com/spreadsheets/d/19mOWgOyM3ci1NoxCAGz6wBLBYMArRKHrT9BgqxT1pmw/edit#gid=1826902422

ANALYSIS BASED ON LITERATURE FINDINGS

PLAYER QUALITY ANALYSIS

MODEL 1

➔ **Player Quality**
- Player fit low (OR<75)
- Player fit medium (75<OR<85)
- Player fit high (OR>85)

➔ **Status of the goal**
- Goal/ No Goal

MODEL 2

➔ **Player Quality**

➔ **Status of the goal**

➔ **GK Quality**
- Great/ Not Great

MODEL 3

➔ **Player Quality**

➔ **Status of the goal**

➔ **Movement of GK**
- Moving/ Not Moving

MODEL 4

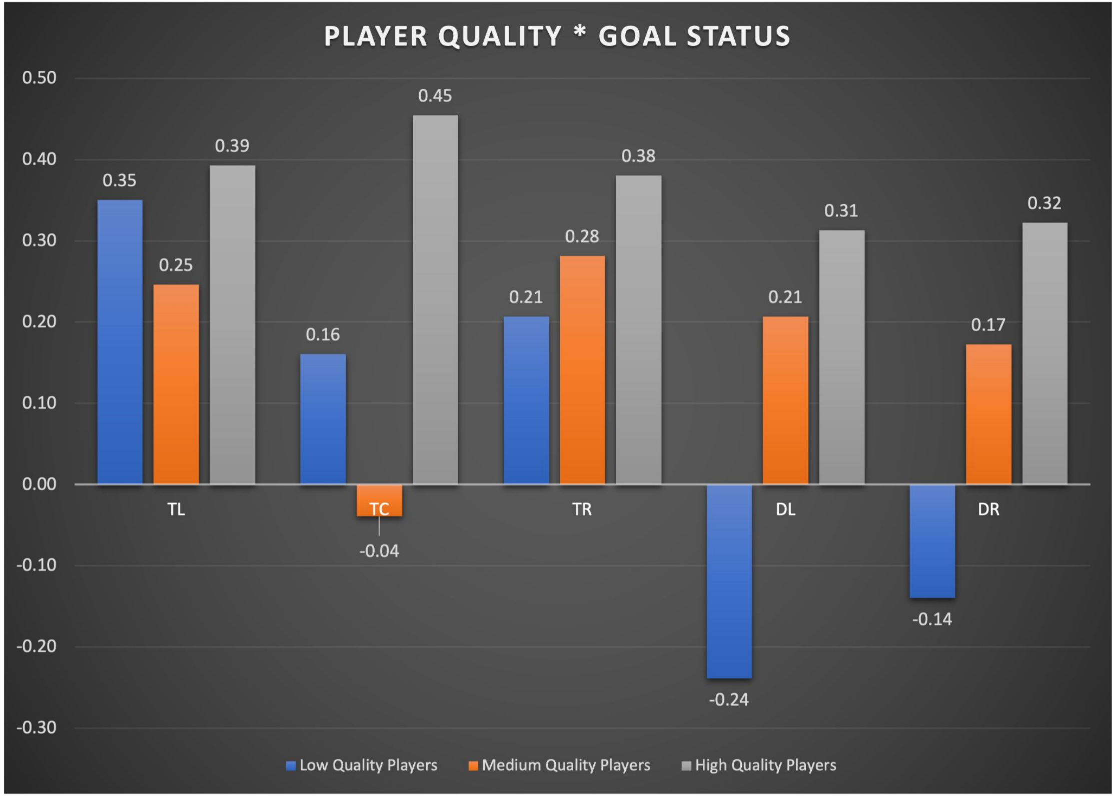
➔ **Player Quality**

➔ **Status of the goal**

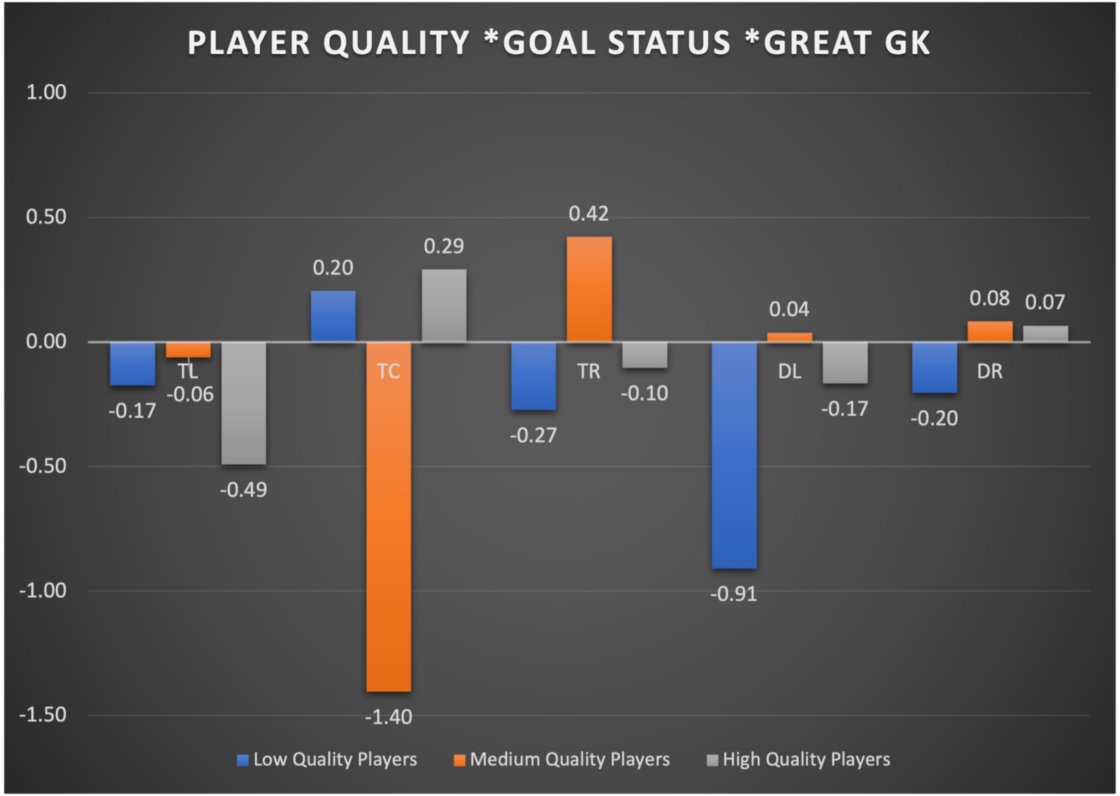
➔ **Height Of GK**
- Height >= 190 is 1
- Height < 190 is 0

****OR= Overall Rating**

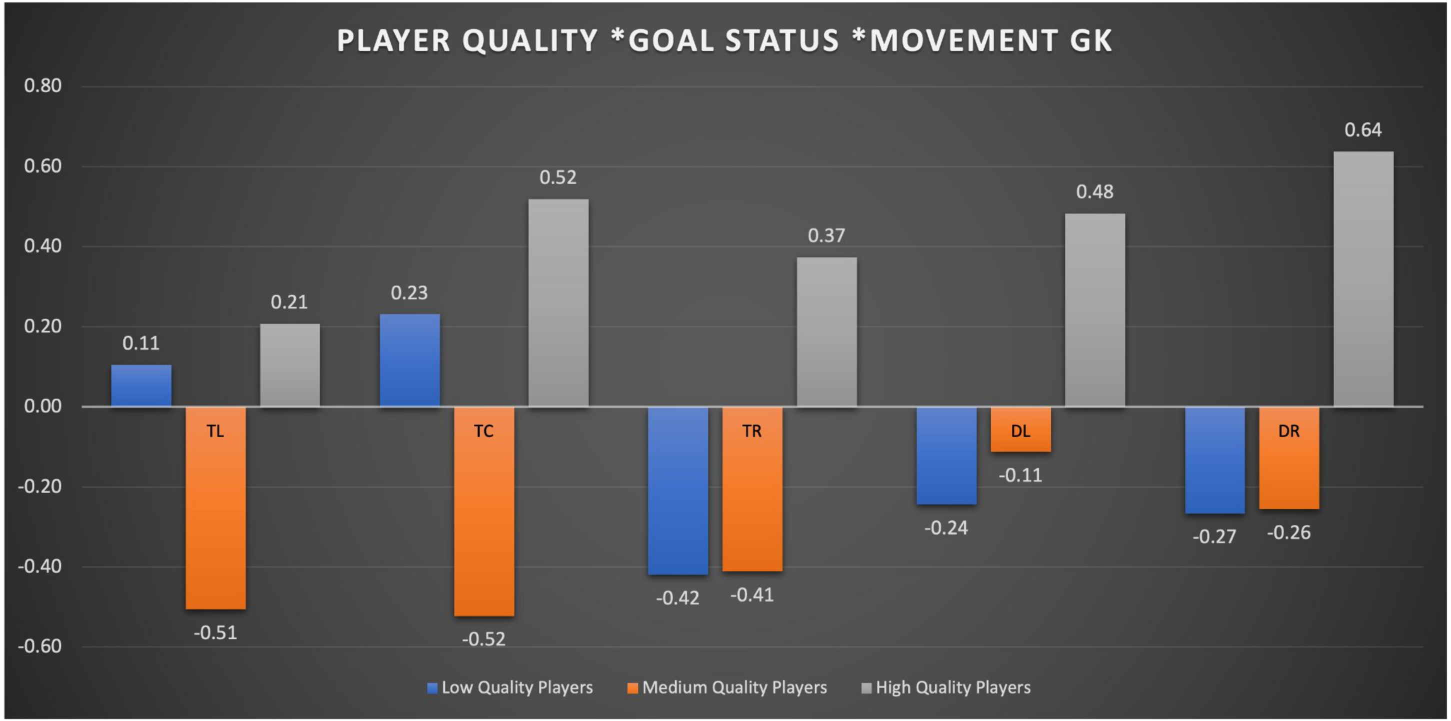
MODEL 1


PLAYER QUALITY * GOAL STATUS
0.50
0.40
0.30
0.20
0.10
0.00
-0.10
-0.20
-0.30
0.35
0.25
0.39
0.16
-0.04
0.45
0.21
0.28
0.38
-0.24
0.21
0.31
-0.14
0.17
0.32
TL
TC
TR
DL
DR
Low Quality Players
Medium Quality Players
High Quality Players

MODEL 2


PLAYER QUALITY *GOAL STATUS *GREAT GK
1.00
0.50
0.00
-0.50
-1.00
-1.50
TL
TC
TR
DL
DR
-0.17
-0.06
-0.49
0.20
-1.40
0.29
-0.27
0.42
-0.10
-0.91
0.04
-0.17
-0.20
0.08
0.07
Low Quality Players
Medium Quality Players
High Quality Players

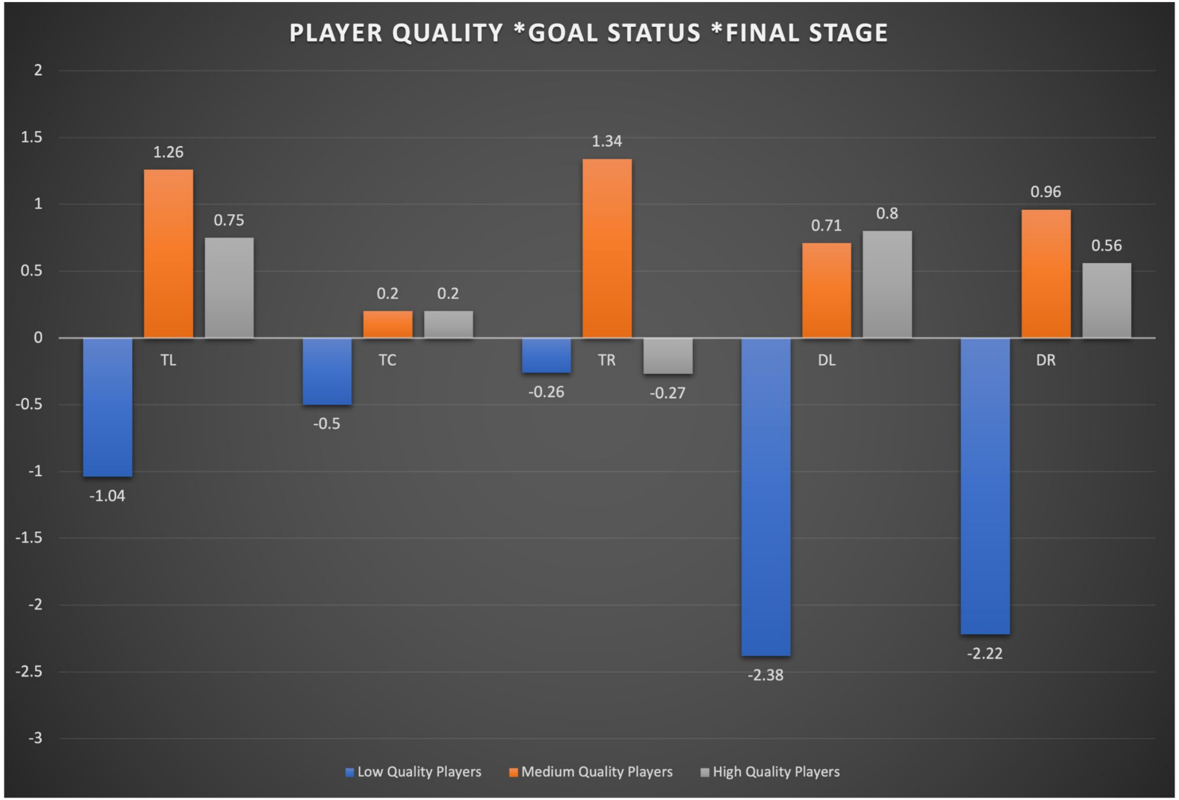
MODEL 3


PLAYER QUALITY *GOAL STATUS *MOVEMENT GK
0.80
0.60
0.40
0.20
0.00
-0.20
-0.40
-0.60
0.11
-0.51
0.21
0.23
-0.52
0.52
-0.42
-0.41
0.37
-0.24
-0.11
0.48
-0.27
-0.26
0.64
TL
TC
TR
DL
DR
Low Quality Players
Medium Quality Players
High Quality Players

MODEL 4

RWTH AACHEN UNIVERSITY

MODEL 5



PLAYER QUALITY *GOAL STATUS *FINAL STAGE
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
-2.5
-3
TL
TC
TR
DL
DR
-1.04
1.26
0.75
-0.5
0.2
0.2
-0.26
1.34
-0.27
-2.38
0.71
0.8
-2.22
0.96
0.56
Low Quality Players
Medium Quality Players
High Quality Players

MODEL PERFORMANCE ON UNSEEN DATA

Test Data 6 alt (Penalties Happened in Last 2 Years)

ID	playerid	name	TL	TC	TR	DL	DC	DR	chosen	predicted by the model	original value	Value Matches or not?
1	14	Alberto More	0.05	0.00	0.06	0.70	0.06	0.13	0.13	DL	DR	FALSE
2	355	Rashford	0.15	0.04	0.08	0.37	0.05	0.31	0.37	DL	DL	TRUE
3	327	Parejo	0.15	0.02	0.06	0.29	0.05	0.43	0.29	DR	DL	FALSE
4	89	Cavani	0.20	0.01	0.03	0.52	0.05	0.18	0.52	DL	DL	TRUE
5	172	Gomez	0.15	0.04	0.08	0.37	0.05	0.31	0.05	DL	DC	FALSE
6	151	Fred	0.08	0.01	0.25	0.35	0.04	0.27	0.27	DL	DR	FALSE
7	15	Albiol	0.15	0.04	0.08	0.37	0.05	0.31	0.31	DL	DR	FALSE
8	20	Alonso	0.10	0.07	0.12	0.43	0.05	0.23	0.43	DL	DL	TRUE
9	265	Lewandowski	0.16	0.05	0.05	0.22	0.05	0.47	0.47	DR	DR	TRUE
10	11	Alaba	0.08	0.01	0.27	0.35	0.04	0.25	0.25	DL	DL	TRUE
11	213	Hummels	0.19	0.01	0.03	0.61	0.04	0.11	0.61	DL	DL	TRUE
12	242	Kessie	0.06	0.13	0.19	0.35	0.10	0.17	0.19	DL	TR	FALSE
13	56	Berardi	0.06	0.01	0.14	0.65	0.08	0.06	0.01	DL	TC	FALSE
14	293	Messi	0.09	0.01	0.41	0.37	0.04	0.08	0.41	TR	TR	TRUE
15	367	Ronaldo	0.08	0.03	0.13	0.47	0.11	0.18	0.03	DL	TC	FALSE
16	419	Suarez	0.04	0.18	0.12	0.20	0.10	0.36	0.36	DR	DR	TRUE
17		Dalot	0.15	0.04	0.08	0.37	0.05	0.31	0.08	DL	TR	FALSE
18		Milner	0.16	0.02	0.57	0.08	0.06	0.10	0.57	TR	TR	TRUE
19		Thiago	0.18	0.06	0.08	0.19	0.08	0.41	0.19	DR	DL	FALSE
20		Reece James	0.17	0.01	0.09	0.63	0.04	0.06	0.01	DL	TC	FALSE
21		Firmino	0.05	0.00	0.05	0.77	0.04	0.09	0.05	DL	TL	FALSE
22		Dembélé	0.17	0.04	0.14	0.27	0.09	0.29	0.29	DR	DR	TRUE
23		Pjanić	0.07	0.00	0.04	0.78	0.06	0.05	0.78	DL	DL	TRUE
24		puig	0.12	0.04	0.27	0.32	0.07	0.19	0.19	DL	DR	FALSE
25		Veretout	0.02	0.02	0.03	0.31	0.15	0.46	0.31	DR	DL	FALSE

https://docs.google.com/spreadsheets/d/1_lj8VKUD6tf0TTfll7POaCD2v1qirb7wZLXQAO8yBTI/edit#gid=0

RWTHAACHEN UNIVERSITY

Test Data 9 alt (Penalties Happened in Last 2 Years)

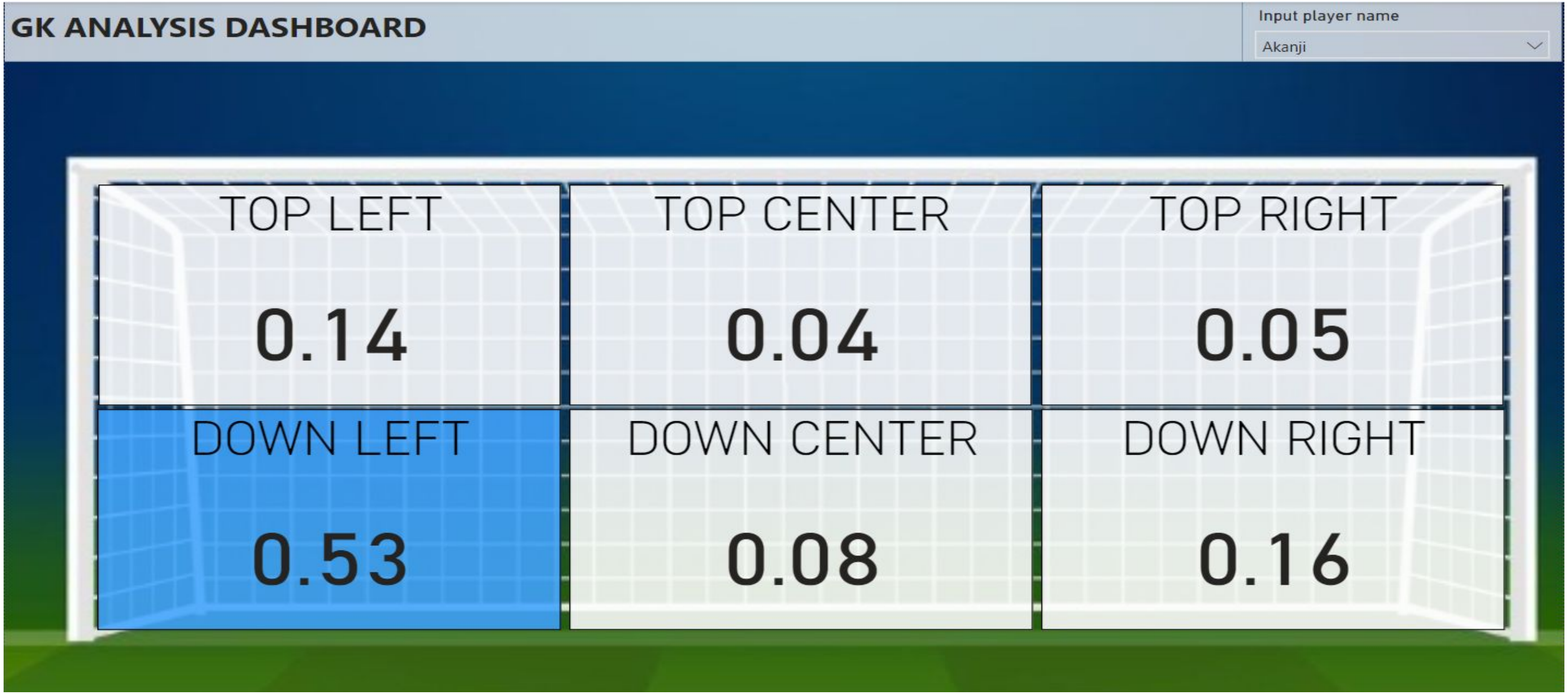
ID	playerid	name	TL	TC	TR	ML	MC	MR	DL	DC	DR	chosen	Predicted by the Model	Original Value	Value Match or not ?
1	14	Alberto Moren	0	0	0	0.05	0.02	0.31	0.03	0.11	0.48	0.48	DR	DR	TRUE
2	355	Rashford	0.12	0.03	0.08	0.01	0.02	0.11	0.35	0.05	0.22	0.35	DL	DL	TRUE
3	327	Parejo	0.13	0.01	0.05	0.02	0.01	0.13	0.25	0.06	0.34	0.25	DL	DL	TRUE
4	89	Cavani	0.13	0.01	0.04	0.03	0.01	0.11	0.11	0.07	0.48	0.48	DR	DR	TRUE
5	172	Gomez	0.35	0.03	0.08	0.01	0.02	0.11	0.12	0.05	0.22	0.35	TL	TL	TRUE
6	151	Fred	0.04	0.01	0.16	0.02	0	0.3	0.27	0.04	0.15	0.15	DL	DR	FALSE
7	15	Albiol	0.12	0.03	0.08	0.01	0.02	0.11	0.35	0.05	0.22	0.22	DL	DR	FALSE
8	20	Alonso	0.08	0.03	0.21	0.01	0.02	0.18	0.2	0.08	0.19	0.21	TR	DL	FALSE
9	265	Lewandowski	0.14	0.03	0.06	0.03	0.01	0.15	0.1	0.08	0.4	0.4	DR	DR	TRUE
10	11	Alaba	0.04	0.01	0.16	0.02	0	0.3	0.27	0.04	0.15	0.16	DL	TR	FALSE
11	213	Hummels	0.11	0.01	0.04	0.03	0	0.11	0.57	0.06	0.06	0.57	DL	DL	TRUE
12	242	Kessie	0.05	0.12	0.07	0.04	0.22	0.04	0.31	0.03	0.12	0.07	DL	TR	FALSE
13	56	Berardi	0.54	0.01	0.05	0.19	0.01	0.11	0.02	0.05	0.03	0.01	TL	TC	FALSE
14	293	Messi	0.05	0.01	0.32	0.03	0	0.29	0.22	0.03	0.04	0.32	TR	TR	TRUE
15	367	Ronaldo	0.07	0.05	0.03	0.15	0.03	0.39	0.09	0.05	0.15	0.05	MR	MR	TRUE
16	419	Suarez	0.04	0.15	0.3	0.03	0.19	0.03	0.18	0.04	0.04	0.04	TR	DR	FALSE
17	293	Messi	0.05	0.01	0.22	0.32	0	0.29	0.01	0.03	0.04	0.32	ML	ML	TRUE
19	89	Cavani	0.13	0.01	0.04	0.48	0.01	0.11	0.03	0.07	0.11	0.48	ML	ML	TRUE
18		Reece James	0.12	0.01	0.07	0.04	0.01	0.14	0.55	0.04	0.03	0.14	DL	MR	FALSE
20		Dalot	0.12	0.03	0.08	0.01	0.02	0.11	0.22	0.05	0.35	0.02	DR	MC	FALSE
21		Jesus	0.11	0.03	0.06	0.03	0.01	0.48	0.15	0.07	0.06	0.15	MR	MR	TRUE
22		Dalot	0.12	0.03	0.08	0.01	0.02	0.11	0.35	0.05	0.22	0.08	DL	TR	FALSE
23		Milner	0.48	0.03	0.06	0.03	0.01	0.15	0.11	0.07	0.06	0.06	TL	TR	FALSE
24		Thiago	0.14	0.03	0.07	0	0.07	0.04	0.17	0.06	0.42	0.17	DR	DL	FALSE
25		Reece James	0.12	0.55	0.07	0.04	0.01	0.14	0.03	0.04	0.03	0.01	TC	TC	TRUE
26		Firmino	0	0	0	0.06	0.02	0.22	0.59	0.07	0.03	0	DL	TR	FALSE
27		Dembélé	0.14	0.04	0.08	0.01	0.03	0.1	0.24	0.07	0.3	0.3	DR	DR	TRUE
28		Pjanić	0	0	0	0.06	0.02	0.17	0.62	0.09	0.03	0.62	DL	DL	TRUE
29		puig	0.11	0.05	0.2	0.01	0.03	0.13	0.3	0.05	0.12	0.12	TR	DR	FALSE
30		Veretout	0	0	0	0.12	0.07	0.09	0.25	0.07	0.4	0.25	DL	DL	TRUE

https://docs.google.com/spreadsheets/d/1Ac6OlKYKOy9Jp9qYH1r8QIN1JTbwQjiV_2nCxRwubhA/edit#gid=0

RWTH AACHEN UNIVERSITY

PREDICTION RESULTS

PREDICTION RESULTS



PREDICTION RESULTS

CHOOSING BETWEEN 6 ALT AND 9 ALT MODEL

Ockham’s Razor (Law of Parsimony)

[1]

Source: Medium

Conclusion

- Two models were proposed namely, the 6-alt & 9-alt model
- After performing the data & pre-estimation analysis, we were able to extract 12 variables out of 42 for 6-alt model using the LR test
- Similarly, for 9-alt model, we were able to extract 10 variables out of 42
- From the results of Cross-Validation for 6-alt & 9-alt model, we observed that the difference between the ‘LL per Observation’ for Estimation and Validation Samples is quite small
- We saw that the IIA was holding for both 6-alt & 9-alt models
- Literature findings proved to be quite successful as we were able to observe the influences clearly
- Model performance on unseen data was favourable to an extent
- Direction of the penalty-taker shot was identified by the interactive Power BI dashboard
- 6-alt model was chosen between the two models which were initially proposed

REFERENCES

- Dalton, K.; Guillon, M.; Naroo, S.A. (2015). An Analysis of Penalty Kicks in Elite Football Post 1997. International Journal of Sports Science & Coaching 10(5), 815-827
- https://dribbble.com/tags/penalty_kick
- https://www.youtube.com/watch?v=gJLq1ypgU_Y&t=1s&ab_channel=SirAlex
- https://www.google.com/url?sa=i&url=https%3A%2F%2Fwww.dailymail.co.uk%2Fsciencetech%2Farticle-2661383%2FThe-science-penalty-kicks-Size-goal-reaction-times-game-theory-odds-overwhelmingly-favour-shooter.html&psig=AOvVaw2E3bZkuIXeihYsHxcCuASB&ust=1653931605061000&source=images&cd=vfe&ved=0CAwQjRxqFwoTCMCQm-WdhfgCFQAAAAAdAAAAABBA
- Prediction of Soccer Penalty Kicks using a Discrete Choice Model (Master Thesis by Fritz Golz, Universität Magdeburg)
- https://mbounthavong.com/blog/tag/chi+square+test - chi quadrant table
- https://en.wikipedia.org/wiki/Multinomial_logistic_regression
- https://nikhilsamuel.medium.com/understanding-occams-razor-6734c8888a31 (Ockham's Razor)
- Braumann, Florian; Friehe, Tim; Wedow, Michael (2011). General Ability and Specialization: Evidence from penalty kicks in soccer, Journal of Sports Economics, 12(1) 81-105.
- Krumer, Alex (2020). Pressure versus ability: Evidence from penalty shoot-outs between teams from different divisions. Journal of Behavioural and Experimental Economics.

RWTH AACHEN UNIVERSITY