Predicting The Outcome of Football Matches

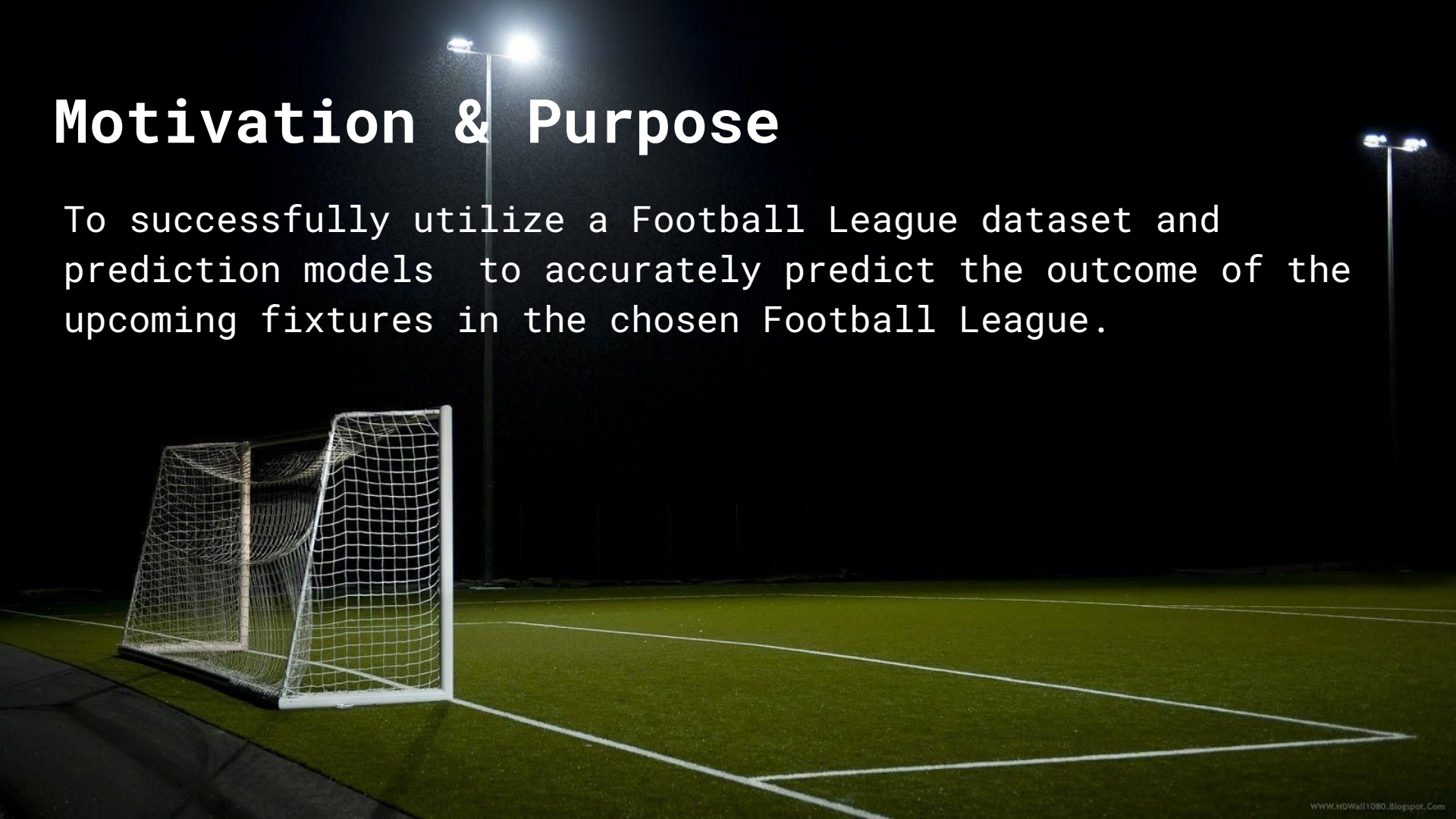
Ingo Davila Banghard

108990005

01.04.2022

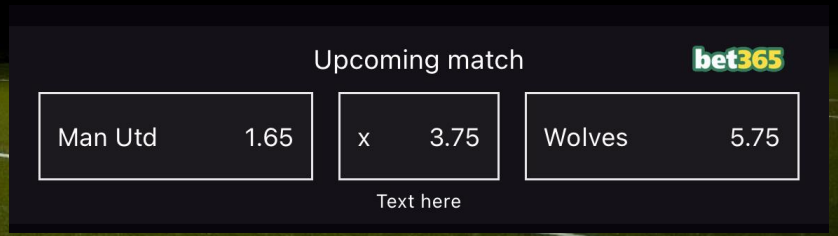
Motivation & Purpose

To successfully utilize a Football League dataset and prediction models to accurately predict the outcome of the upcoming fixtures in the chosen Football League.



Problem Statement

Just as a stock can go up, down, or remain the same. A match outcome can also be categorized by win, lose, or draw.

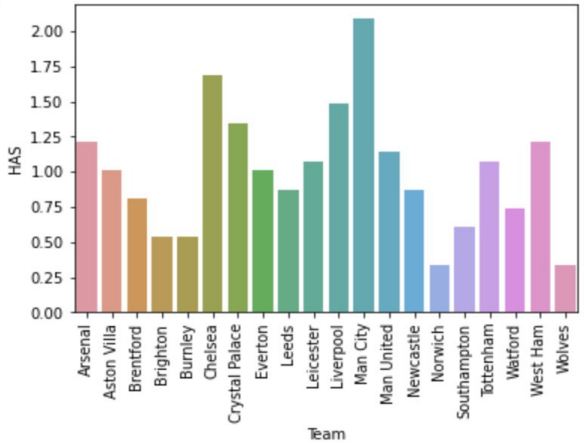



Football-Data.co.uk
Results / Odds / Tipsters
PREMIER LEAGUE
LaLiga
Santander
SERIE A
TIM
BUNDES LIGA
LIGUE 1
WWW.HDWall1080.Blogspot.Com

```
(1)Premier League (2)La Liga (3)Bundesliga (4)League One (5)Serie A
Please Enter the Number of the League Desired: 1
```

	Div	Date	Time	HomeTeam	AwayTeam	FTHG	FTAG	FTR	HTHG	HTAG	HTR	HS	AS	HST	AST	HF	AF	HC	AC	HY	AY	HR
0	E0	13/08/2021	20:00	Brentford	Arsenal	2	0	H	1	0	H	8	22	3	4	12	8	2	5	0	0	0
1	E0	14/08/2021	12:30	Man United	Leeds	5	1	H	1	0	H	16	10	8	3	11	9	5	4	1	2	0
2	E0	14/08/2021	15:00	Burnley	Brighton	1	2	A	1	0	H	14	14	3	8	10	7	7	6	2	1	0
3	E0	14/08/2021	15:00	Chelsea	Crystal Palace	3	0	H	2	0	H	13	4	6	1	15	11	5	2	0	0	0
4	E0	14/08/2021	15:00	Everton	Southampton	3	1	H	0	1	A	14	6	6	3	13	15	6	8	2	0	0
5	E0	14/08/2021	15:00	Leicester	Wolves	1	0	H	1	0	H	9	17	5	3	6	10	5	4	1	2	0

	Team	HGS	AGS	HAS	AAS	HGC	AGC	HDS	ADS
0	Arsenal	18	15	1.212121	1.214575	8	17	0.647773	1.144781
1	Aston Villa	15	10	1.010101	0.809717	14	16	1.133603	1.077441
2	Brentford	12	11	0.808081	0.890688	12	14	0.971660	0.942761
3	Brighton	8	12	0.538721	0.971660	9	11	0.728745	0.740741
4	Burnley	8	8	0.538721	0.647773	8	19	0.647773	1.279461
5	Chelsea	25	20	1.683502	1.619433	10	6	0.809717	0.404040



KNN

K-NN algorithm assumes the similarity between the new case/data and available cases and put the new case into the category that is most similar to the available categories.

XGB

XGBoost is an optimized distributed gradient boosting library designed to be highly efficient, flexible and portable. It implements Machine Learning algorithms under the Gradient Boosting framework. It provides a parallel tree boosting to solve many data science problems in a fast and accurate way.

Logistic Regression

Logistic regression is a statistical analysis method used to predict a data value based on prior observations of a data set. Based on historical data about earlier outcomes involving the same input criteria, it then scores new cases on their probability of falling into a particular outcome category.

```
Enter the Home Team please: Chelsea
Enter who plays against Chelsea: Liverpool
```

	HomeTeam	AwayTeam	FTR	FTHG	FTAG	HS	AS	HC	AC	pastHS	pastHC	pastAS	pastAC	pastHG	pastAG	HAS	HDS	AAS	ADS
191	Chelsea	Liverpool	D	0	0	0	0	0	0	12.000000	6.000000	10.333333	4.333333	2.000000	1.333333	1.683502	0.809717	1.619433	0.404040
190	Man United	Wolves	A	0	1	9	19	3	8	11.333333	5.666667	14.666667	8.333333	1.666667	0.333333	1.144781	1.214575	1.052632	0.808081
189	Chelsea	Liverpool	D	2	2	15	10	6	7	10.000000	5.333333	6.666667	6.000000	1.333333	1.666667	1.683502	0.809717	1.619433	0.404040
188	Leeds	Burnley	H	3	1	22	8	9	3	22.666667	4.000000	17.666667	7.666667	1.000000	0.333333	0.875421	1.295547	0.647773	1.414141
187	Everton	Brighton	A	2	3	17	12	8	5	16.666667	6.000000	11.666667	7.666667	1.333333	1.000000	1.010101	1.295547	0.647773	1.077441

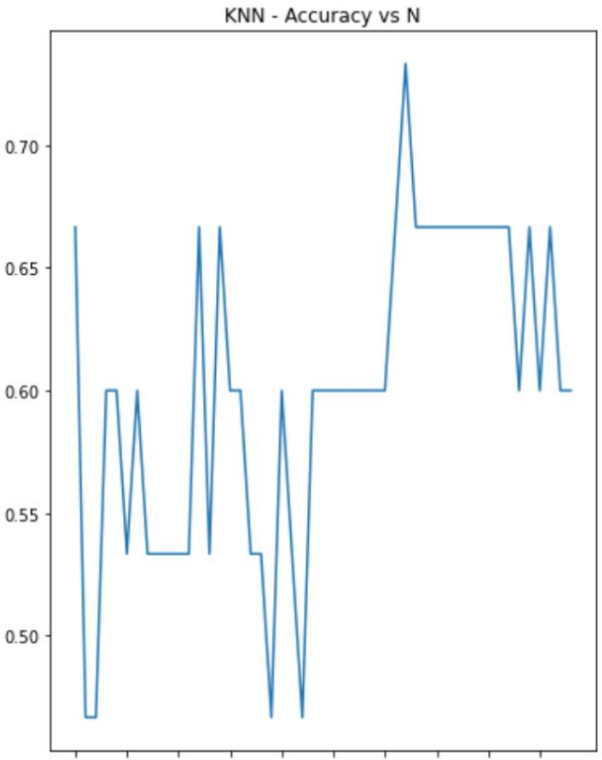
	pastCornerDiff	pastGoalDiff	pastShotsDiff	HAS	HDS	AAS	ADS
191	0.555556	0.222222	3.555556	1.683502	0.809717	1.619433	0.40404

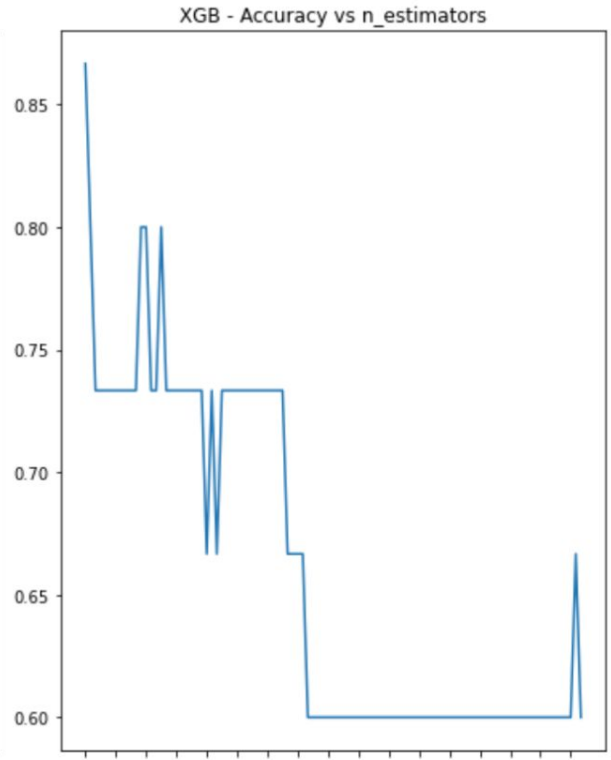
KNN = 0.867
XGB = 0.671
Logistic Regression = 0.534

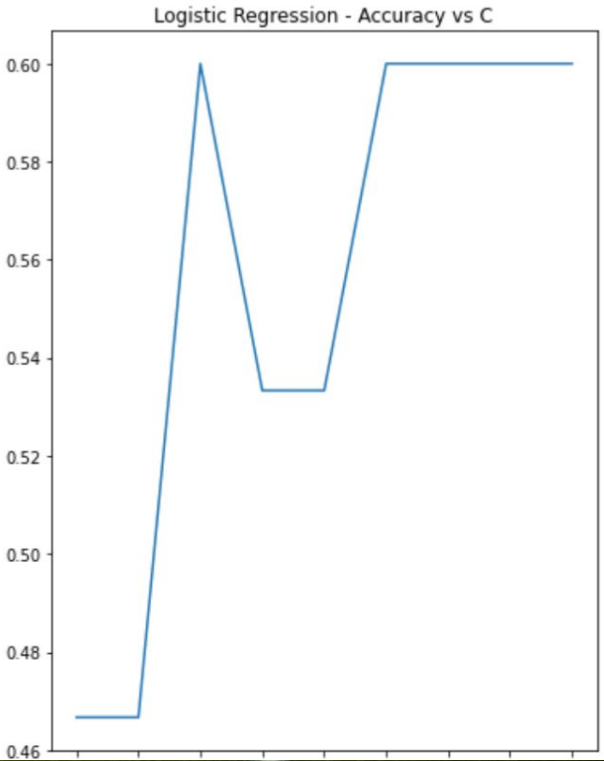
KNN

XGBoost

Logistic Regression







	HomeTeam	AwayTeam	Res_knn	Res_XGB	Res_logreg
183	Man United	Wolves	A	H	H

	HomeTeam	AwayTeam	Res_knn	Res_XGB	Res_logreg
183	Arsenal	Man City	D	H	H

	HomeTeam	AwayTeam	Res_knn	Res_XGB	Res_logreg
183	Watford	Tottenham	D	A	A

	HomeTeam	AwayTeam	Res_knn	Res_XGB	Res_logreg
191	Brentford	Aston Villa	A	A	A

	HomeTeam	AwayTeam	Res_knn	Res_XGB	Res_logreg
191	Leeds	Burnley	A	A	D

	HomeTeam	AwayTeam	Res_knn	Res_XGB	Res_logreg
191	Everton	Brighton	A	A	A

	HomeTeam	AwayTeam	Res_knn	Res_XGB	Res_logreg
191	Crystal Palace	West Ham	H	H	H

	HomeTeam	AwayTeam	Res_knn	Res_XGB	Res_logreg
191	Chelsea	Liverpool	H	H	H

	HomeTeam	AwayTeam	Res_knn	Res_XGB	Res_logreg
191	Villarreal	Levante	D	D	H

this_week

	HomeTeam	AwayTeam	Res_knn	Res_XGB	Res_logreg
191	Cadiz	Sevilla	H	A	A

Jan 1, 2022 (Saturday)

FT	1	Arsenal
	2	Manchester City
FT	0	Watford
	1	Tottenham Hotspur

Jan 2, 2022 (Sunday)

FT	2	Crystal Palace
	3	West Ham
FT	2	Brentford FC
	1	Aston Villa
FT	3	Leeds United
	1	Burnley
FT	2	Everton
	3	Brighton & Hove Albion

Yesterday

FT	2	Chelsea
	2	Liverpool

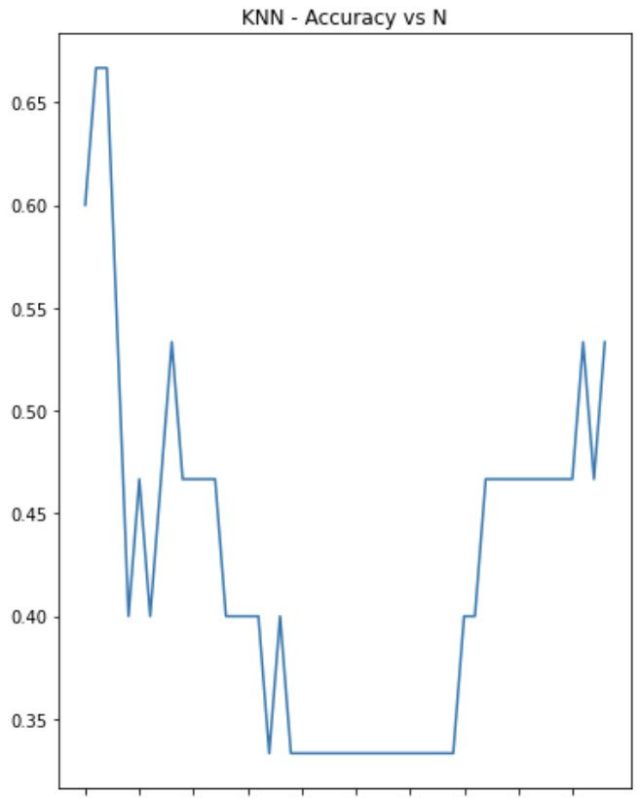
Today

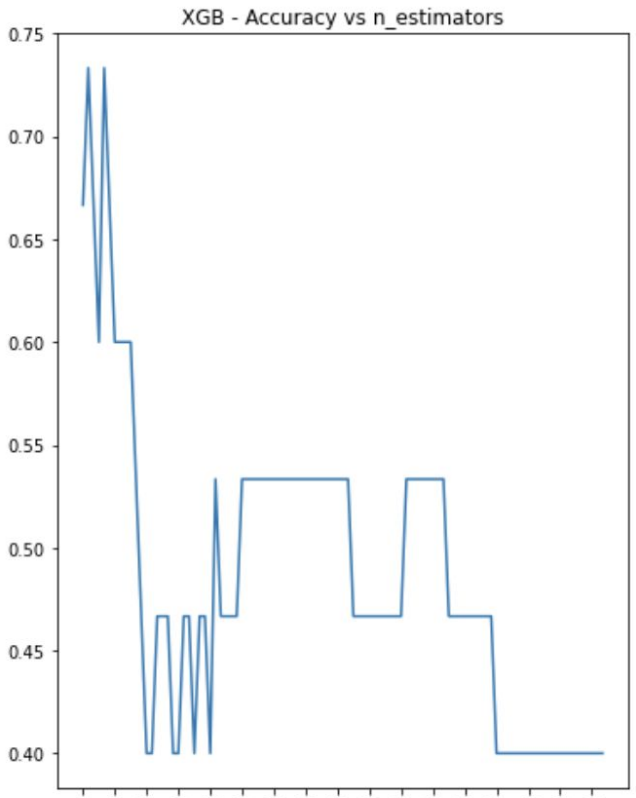
FT	0	Manchester United
	1	Wolves

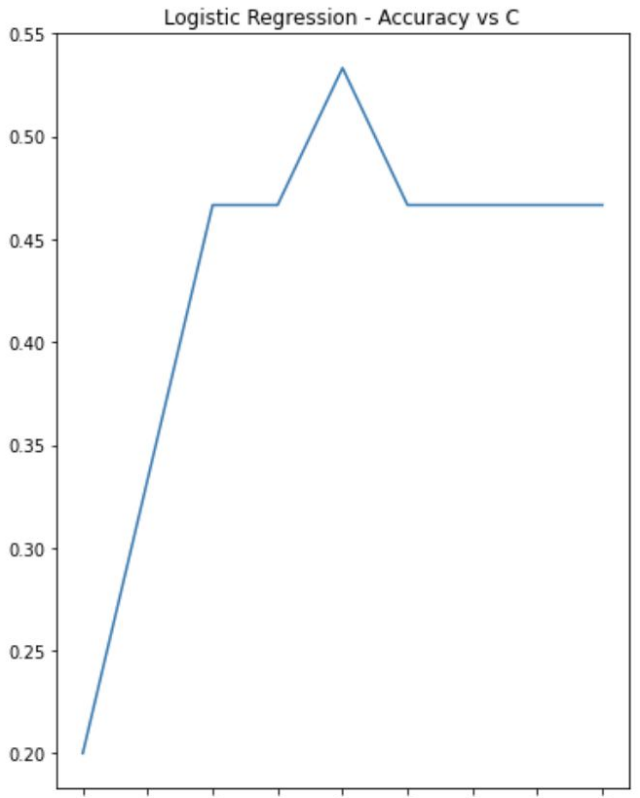
KNN
0.7333

XGBoost
0.6100

Logistic Regression
0.5333

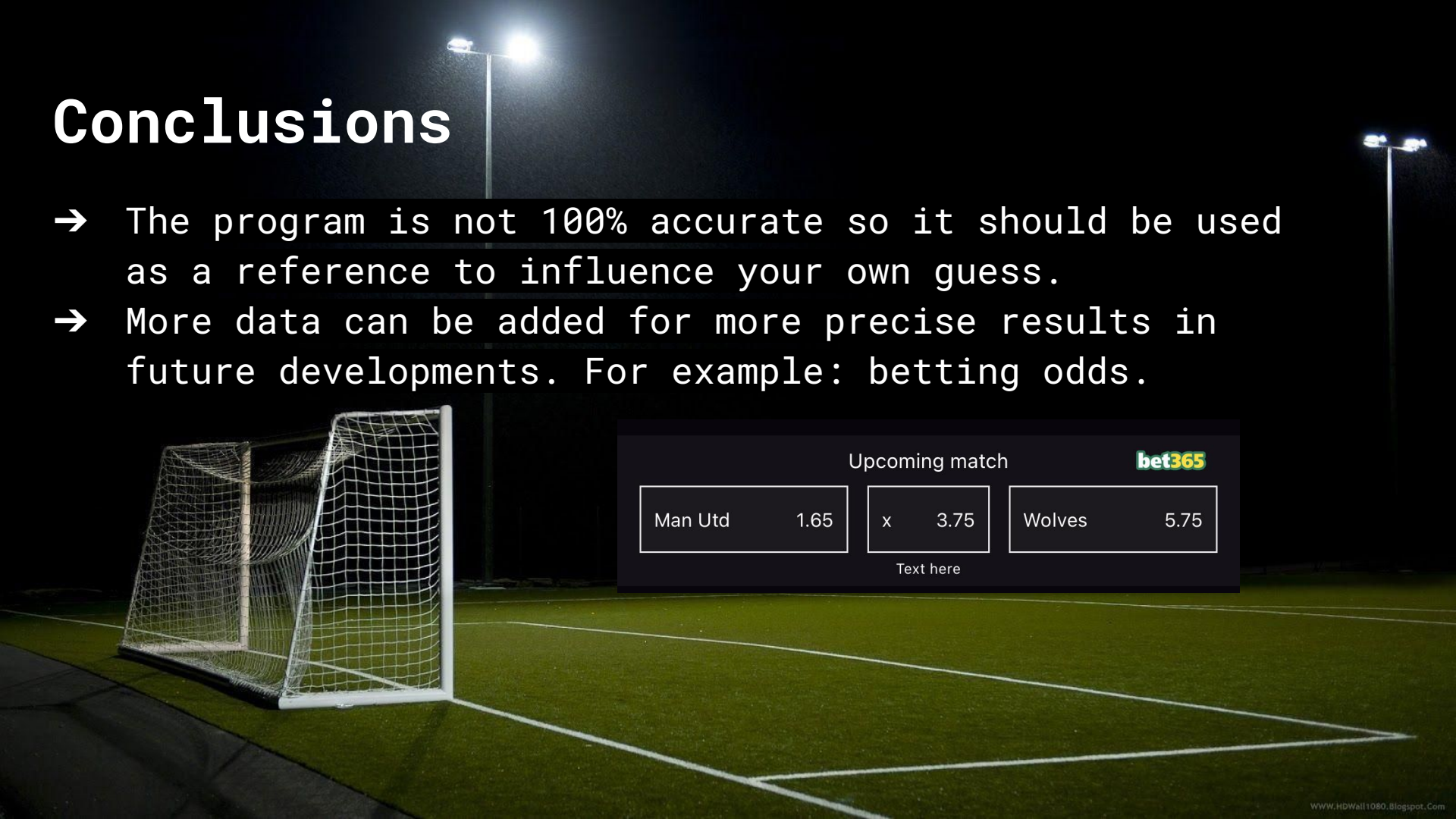






Conclusions

- ➔ The program is not 100% accurate so it should be used as a reference to influence your own guess.
- ➔ More data can be added for more precise results in future developments. For example: betting odds.



Future Development

- There is a lot of potential for this program. Especially regarding data.
- The more information is has to work with the more accurate it can be.
- Difference in points, Last 5 match results, weather conditions, referee, available players, individual player stats, etc.