Game Champion Prediction Using Machine Learning

Ch. Ranjith Kumar[1]M.Tech, (Ph.D) N.Raji[2], T. Shalini Lakshmi[3], A.Haritha[4], K.N.V.S.Tulasi[5]

[1]Associate Professor, Dept of CSE, BVC Institute of Technology and Science, Batlapalem, Amalapuram A.P, India

[2]B.Tech Student, Dept of CSE, BVC Institute of Technology and Science, Batlapalem, Amalapuram A.P, India

[3]B.Tech Student, Dept of CSE, BVC Institute of Technology and Science, Batlapalem, Amalapuram A.P, India

[4]B.Tech Student, Dept of CSE, BVC Institute of Technology and Science, Batlapalem, Amalapuram A.P, India

[5]B.Tech Student, Dept of CSE, BVC Institute of Technology and Science, Batlapalem, Amalapuram A.P, India

Abstract:

Many techniques to expect the outcome of professional football matches have traditionally used the number of goals scored by each team as a base measure for evaluating a team's concert and valuing future results. However, the number of goals scored during a match possesses an important random element which leads to large conflicts in many games between a team's performance and number of goals scored. We will discover different model design theories and assess our models performance against target techniques.The main objective of this project is to explore different Machine Learning techniques to predict the score and result of football matches, using in-game match events rather than the number of goals scored by each team. In this project, we developed an 'Win conditions' metric which helps us to estimate a team's performance used to build a classification model predicting the result of future matches, as well as a regression model predicting the score of future games.

Keywords: Prediction, Classification model, Regression model.

1. Introduction:

As a result, Machine Learning is presently a greatly trending approach. The model used by them for making calculations however is extremely personal and it is amazing as to how predictions were made with 100% accuracy.

Outcomes from sports matches can be difficult to predict. Football in particular is an exciting example as matches have fixed length (as opposed to racket sports such as tennis, where the game is played until a player wins). It also keeps a single type of scoring event: goals (as different to a sport like rugby where different events score a different number of points) that can hap-pen an infinite amount of times during a match, and which are all value one point.

The possible results for a team taking part in a football match are win, loss or draw. It can therefore look rather direct to predict the result of a game.Traditional predictive methods have simply used match results to estimate team performance and build statistical models to predict the results of future games.

2. An Overview of Proposed System:

In this paper, we propose a model to predict the results of Football matches in the FIFA worldcup. We train the final data-set on variousmachine learning classifiers and choose the one that returns the best result. Then, we optimize the classifier that produces the best result to further enhance the model accuracy in making predictions

combination of features and classifiers required to make predictions.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
1	Year	Datetime	Stage	Stadium	City	Home Tea	Home Tea	Away Tea	Away Tea	Win condi	Attendan	Half-time	Half-time	Referee	Assistant	Assistant	RoundID	MatchID	Home Team	Initials
2	1930	13 Jul 193	Group 1	Pocitos	Montevid	France	4	1	Mexico	France	4444	3	0	LOMBARD	CRISTOPH	REGO Gilb	201	1096	FRA	
3	1930	13 Jul 193	Group 4	Parque Ce	Montevid	USA	3	0	Belgium	USA	18346	2	0	MACIAS Jo	MATEUCC	WARNKEN	201	1090	USA	
4	1930	14 Jul 193	Group 2	Parque Ce	Montevid	Yugoslavi	2	1	Brazil	Yugoslavi	24059	2	0	TEJADA Ar	VALLARIN	BALWAY T	201	1093	YUG	
5	1930	14 Jul 193	Group 3	Pocitos	Montevid	Romania	3	1	Peru	Romania	2549	1	0	WARNKEN	LANGENU	MATEUCC	201	1098	ROU	
6	1930	15 Jul 193	Group 1	Parque Ce	Montevid	Argentina	1	0	France	Argentina	23409	0	0	REGO Gilb	SAUCEDO	RADULESC	201	1085	ARG	
7	1930	16 Jul 193	Group 1	Parque Ce	Montevid	Chile	3	0	Mexico	Chile	9249	1	0	CRISTOPH	APHESTEG	LANGENU	201	1095	CHI	
8	1930	17 Jul 193	Group 2	Parque Ce	Montevid	Yugoslavi	4	0	Bolivia	Yugoslavi	18306	0	0	MATEUCC	LOMBARD	WARNKEN	201	1092	YUG	
9	1930	17 Jul 193	Group 4	Parque Ce	Montevid	USA	3	0	Paraguay	USA	18306	2	0	MACIAS Jo	APHESTEG	TEJADA Ar	201	1097	USA	
10	1930	18 Jul 193	Group 3	Estadio Ce	Montevid	Uruguay	1	0	Peru	Uruguay	57735	0	0	LANGENU	BALWAY T	CRISTOPH	201	1099	URU	
11	1930	19 Jul 193	Group 1	Estadio Ce	Montevid	Chile	1	0	France	Chile	2000	0	0	TEJADA Ar	LOMBARD	REGO Gilb	201	1094	CHI	
12	1930	19 Jul 193	Group 1	Estadio Ce	Montevid	Argentina	6	3	Mexico	Argentina	42100	3	1	SAUCEDO	ALONSO G	RADULESC	201	1086	ARG	
13	1930	20 Jul 193	Group 2	Estadio Ce	Montevid	Brazil	4	0	Bolivia	Brazil	25466	1	0	BALWAY T	MATEUCC	VALLEJO G	201	1091	BRA	
14	1930	20 Jul 193	Group 4	Estadio Ce	Montevid	Paraguay	1	0	Belgium	Paraguay	12000	1	0	VALLARIN	MACIAS Jo	LOMBARD	201	1089	PAR	
15	1930	21 Jul 193	Group 3	Estadio Ce	Montevid	Uruguay	4	0	Romania	Uruguay	70022	4	0	REGO Gilb	WARNKEN	SAUCEDO	201	1100	URU	
16	1930	22 Jul 193	Group 1	Estadio Ce	Montevid	Argentina	3	1	Chile	Argentina	41459	2	1	LANGENU	CRISTOPH	SAUCEDO	201	1084	ARG	
17	1930	26 Jul 193	Semi-fina	Estadio Ce	Montevid	Argentina	6	1	USA	Argentina	72886	1	0	LANGENU	VALLEJO G	WARNKEN	202	1088	ARG	
18	1930	27 Jul 193	Semi-fina	Estadio Ce	Montevid	Uruguay	6	1	Yugoslavi	Uruguay	79867	3	1	REGO Gilb	SAUCEDO	BALWAY T	202	1101	URU	
19	1930	30 Jul 193	Final	Estadio Ce	Montevid	Uruguay	4	2	Argentina	Uruguay	68346	1	2	LANGENU	SAUCEDO	CRISTOPH	405	1087	URU	
20	1934	27 May 19	Prelimina	Stadio Ber	Turin	Austria	3	2	France	Austria	16000	0	0	VAN MOC	CAIRONI G	BAERT Lo	204	1104	AUT	
21	1934	27 May 19	Prelimina	Giorgio As	Naples	Hungary	4	2	Egypt	Hungary	9000	2	2	BARLASSI	DATTILO G	SASSI Ote	204	1119	HUN	
22	1934	27 May 19	Prelimina	San Siro	Milan	Switzerlar	3	2	Netherlar	Switzerlar	33000	2	1	EKLIND Iv	BERANEK	BONIVEN	204	1133	SUI	
23	1934	27 May 19	Prelimina	Littorale	Bologna	Sweden	3	2	Argentina	Sweden	14000	1	1	BRAUN Eu	CARRARO	TURBIANI	204	1102	SWE	

Fig: FIFA world cup matches

3. Correlation Matrix

The results of the paper were based on a data-set that only included information about those who are participate in FIFA world cup. To further improve this research, machine learning algorithm is developed. In machine learning techniques supervised learning is used in real world application .In this paper it can be done, using classification model.In the dataset drop NAN values and also drop unwanted columns information. Draw correlation matrix for the dataset and estimate the future winner in the football world cup.Before prediction splitting is performed for train and test the dataset then apply classification model.

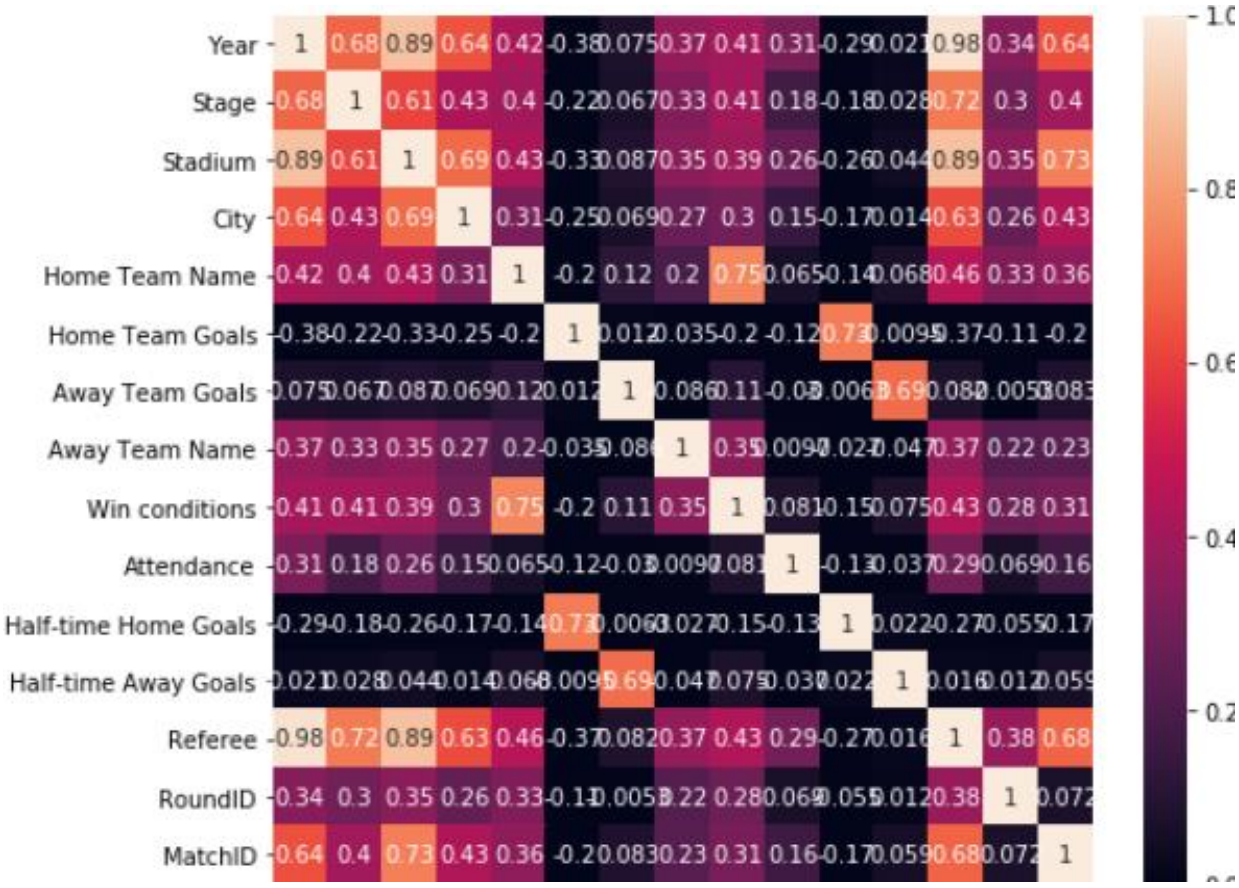


Fig: Correlation Matrix.

Before applying the algorithm we draw the correlation matrix, in this matrix we have negative values so we have to drop the columns with negative values and consider positive value columns.

4. Algorithm

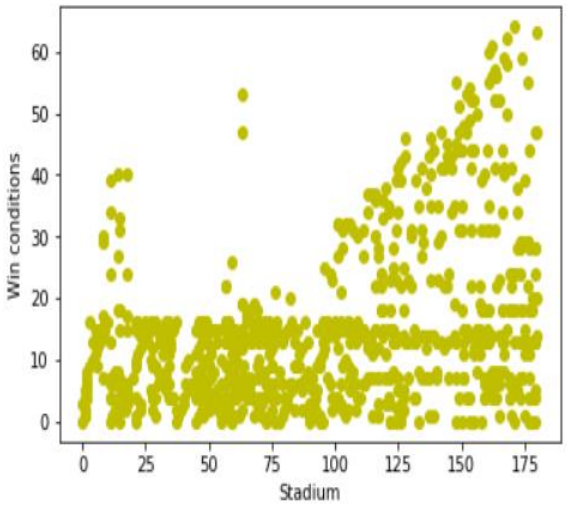
Decision Tree Classifier: Decision trees are a popular Machine learning methods to link input variables signified in the trees, branches and nodes with an output value represented in the trees leaves. Tress can both be used in classification problems, by outputting a real number. Decision trees can be fixed using different algorithms.

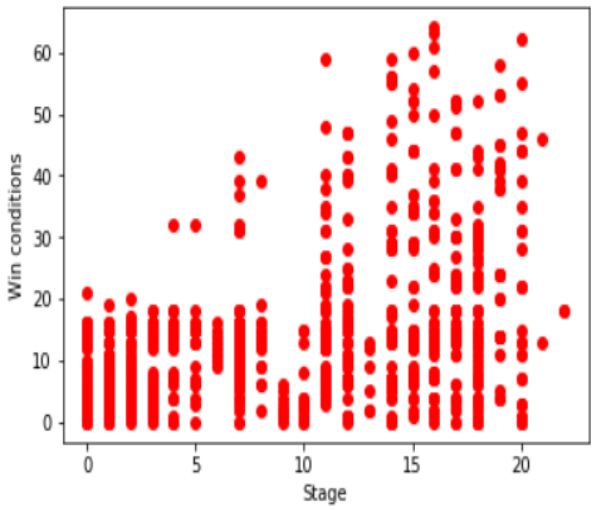
```
decisiontree=DecisionTreeClassifier(criterion='entropy',max_depth=4)

obtained_tree=decisiontree.fit(X_train,Y_train)

print("Extracted classes",decisiontree.classes_)

Predicted_Winconditions=obtained_tree.predict(X_test)

print(Predicted_Winconditions)
```

```
print(df2['Win conditions'].tolist())

print(confusion_matrix(Y_test,Predicted_Winconditions))

importseaborn

plt.figure(figsize=(22,22))

print(seaborn.heatmap(confusion_matrix(Y_test,Predicted_Winconditions),annot=True,cmap="Blues",xti
cklabels=decisiontree.classes_,yticklabels=decisiontree.classes_))

print(decisiontree.score(x,y)) print(decisiontree.predict([[2020,0,0,0]]))
```

Scatter Plots for Prediction

A graph in which the values of two values are plotted along two axes, the pattern of the resulting points revealing any correlation present. A scatter plot can be used either when one continuous variable that is under the control of the experimenter and the other combinations of categorical and quantitative variables. A mosaic plot, fluctuation diagram, or faceted bar chart may be used to display two categorical variables. Other plots are used for one categorical and one quantitative variable. It depends on it or when both continuous variables are independent. If a parameter exists that is systematically incremented and/or decremented by the other, it is called the control parameter or independent variable and is customarily plotted along the horizontal axis. The measured or dependent variable is customarily plotted along the vertical axis. If no dependent variable exists, either type of variable can be plotted on either axis or a scatter plot will illustrate only the degree of correlation (not causation) between two variables. A generalized scatter plot matrix offers a range of displays of paired.





5. Architecture/Design

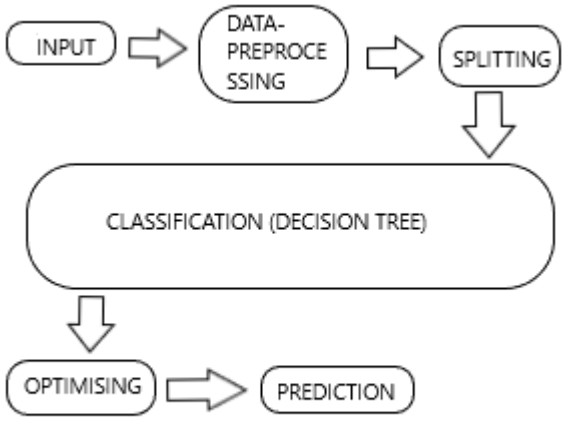


Fig: Architecture Diagram

A. Dataset Description

We are going to be predicting match results using data from past games for a few seasons. We obtained this data from football matches. We filtered these elements into a final list which proved to be the most significant for predicting the result.

B. Pre-Processing

The data-set we obtained from previous football matches several elements from each season.A lot of these features are pretty much unnecessary for making outcome predictions. Hence, our primary task is to clean the data to only retain the features or elements most. We calculate the correlation matrix to observe how much one element affects another set and their correlations. This will help us pick the most significant features that we want to use to build our new dataset.

C. Data Splitting:

Once we finish building our new set of vital elements, we split the data into training and testing data.

D. Optimizing

Once we select the best performing classifier, we optimize its hyper-parameters, to further develop the performance and accuracy of the model in making our prediction. Finally, we obtain our target variable for predicting the result of the match.

E.Prediction

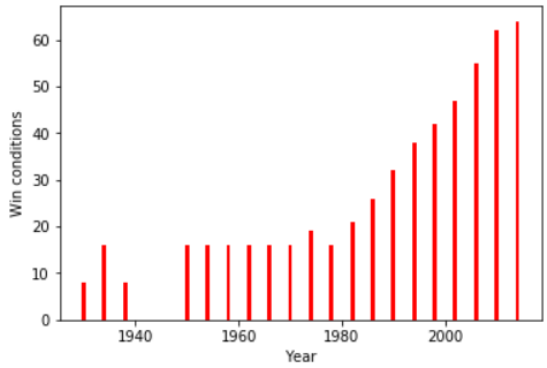
In this module, we apply the machine learning classifiers required for making our prediction.

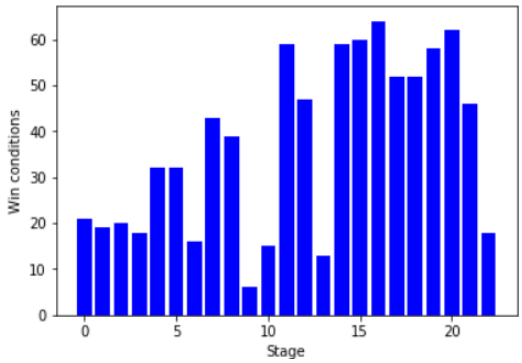
Bar Graph for accuracy Prediction

A bar graph is a pictorial rendition of statistical data in which the independent values can attain only certain discrete values. The most common form of bar graph is the vertical bar graph, also called as column graph. A bar graph is a chart that uses bars to show comparisons between categories of data. The bars can be either horizontal or vertical. Bar graphs with vertical bars are sometimes called vertical bar graphs. A bar graph will have two axes.

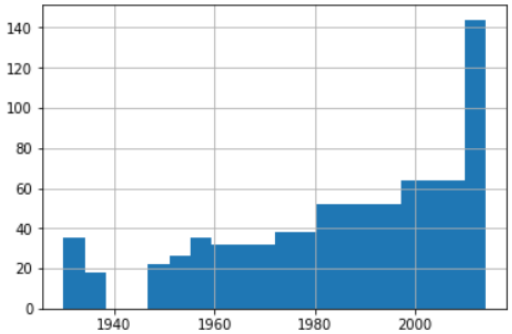
One axis will describe the types of categories being compared, and the other will have numerical values that represent the values of the data. It does not matter which axis is which, but it will determine what bar graph is shown.

If the descriptions are on the horizontal axis, the bars will be oriented vertically, and if the values are along the horizontal axis, the bars will be oriented horizontally.





Mat Plot for accuracy



6. Conclusion

The model we designed is based on dataset of past football games. We will be able to make properly accurate predictions. Although the accuracy of this model is attractive good, it's not sure to be always right and there is a lot of scope for future work in this regard. We could bringfeatures such as team performance metrics.

7. References

[1] *Size of the betting industry* [http://www.bbc.co.uk/sport/football/24354124[1]]. pages

[2] *Average goals per match in the Premier League* [http: //www.skysports.com/football/news/11095/10752408/ the-evolution-of-full-backs-how-they-became-integral-to-success[2]].pages 2

[3] *Linear Model diagram from scikit-learn website* [http://scikit-learn.org/stable/modules/linear_model.html[3]]. pages 8

[4] *Logistic Function plot* [http://www.thefactmachine.com/logistic-regression/[4]]. pages 10

[5] *Architecture of the Random Forest model* [https://www.researchgate.net/figure/Architecture-of-the-random-forest-model_fig1_301638643[5]].pages 11

[6] *Diagram of the Gaussian Naive Bayes model* [https://www.researchgate.net/ figure/Illustration-of-how-a-Gaussian-Naive-Bayes-GNB-classifier-works-For-each fig1_255695722[6]]. pages 12

[7] *Diagram of the k-Nearest-Neighbors model* [http://scikit-learn.org/stable/auto_examples/neighbors/plot_classification.html[7]]. pages 13

[8] *Plot of the SVM Hyperplane* [http://scikit-learn.org/stable/modules/svm.html[8]]. pages 14

[9] *Diagram of the neural network model* [http://scikit-learn.org/stable/modules/neural_networks_supervised.html[9]]. pages 15

[10] M. J. Moroney. *Facts from figures, 3rd edn.*. Penguin: London, 1956. pages 16

[11] C. Reep. *Skill and chance in ball games*. Journal of the Royal Statistical Society Series A 131: 581-585, 1971. pages 16

[12] I.D. Hill. *Association football and statistical inference*. Applied Statistics 23: 203-208, 1974. pages 16

M. J. Maher. *Modelling association football scores*. StatisticaNeerlandica, 1982. pages 16