

MACHINE LEARNING BASED IPL WINNER ANALYSIS

Mrs.Ranjani.J
dept.of InformationTechnology
SriSairamEngineeringCollege
Chennai,India
ranjani.it@sairam.edu.in

Mrs.B.Kamala
dept.of InformationTechnology
SriSairamEngineeringCollege
Chennai,India
kamala.it@sairam.edu.in

Mrs.V.K.G.Kalaiselvi
dept.of InformationTechnology
SriSairamEngineeringCollege
Chennai,India
Kalaiselvi.it@sairam.edu.in

KarthikRaja Y
Dept.InformationTechnology
SriSairamEngineeringCollege
Chennai,India
sec20it004@sairamtap.edu.in

SrikantA
dept.InformationTechnology
SriSairamEngineeringCollege
Chennai,India
sec20it049@sairamtap.edu.in

1. Abstract:

Ipl is a most famous league in India as well as all over the world. Nowadays T20 format of cricket game is mostly enjoyed by the viewers. A tournament named IPL which a T20 format has been one of the reasons for growth of T20 format of this game. The experts used to call cricket as a funny game as anything can happen in this form of cricket. In our project we are going to analyze which team has majority of chances in winning. Some of the major techniques used in our project are SVM, logistic regression, KNN, decision tree, etc. Finally, the conclusion, based upon which the algorithm gives the accuracy. Decision tree and logistic regression had given a accuracy over 89% and 97% respectively.

2. INTRODUCTION

Machine learning is a branch of broader field of artificial intelligence that makes use of statistical models to develop prediction. In basic technical terms machine learning uses algorithms that take empirical or historical data in, analyze it, and generate output based on that analysis. The main aim is to identify necessary parameters that influence the outcome of the match and choose best machine learning model that fits the data and provides good outcome. Some work is published in this domain of forecasting the match outcome. Since only a few important factors are used to forecast in certain papers but the accuracy

is lower. Billions of viewers from across the globe watch the T20 cricket competition known as IPL in India. There is an exciting and thrilling expectation around every match in a league, which consists some of the top cricket players in the globe. With the development of artificial intelligence and machine learning it is nowadays more interesting among people to predict the result of the match. Here we will talk about how to analyze the winning team of the IPL. **LITERATURE SURVEY**

Considering many strategies and variables that must be considered during the procedure, we have discovered that techniques such as random forest, support vector machines were not exploited completely. They created and verified a more virtual technique for predicting scores with greater possibility in this research. We collected dataset from the last year and checked the probability from these models. Ravindra and Azeem [1] used data from IPL official website to execute this document. Since the data has a lot of features, we looked into it and picked out a few most important ones. Jaiswal [2] provides the idea for developing a prediction system which uses previous data to check whether the upcoming IPL matches will be won or lost. Ananda [3] found the Duckworth-Lewis method and tells the accuracy. Shilpi, Sujal, Jayash [4] analyze the winner of the match using 3 different ML algorithm techniques. Vardar [5] got the technique the splitting is the

best approach to analyze the winner of the match

4. EXISTING ANALYSIS

In the existing system analysis, there is a formula available to avail the projected score and a win analysis based on the winning percentage of a poll and teams. These methods will not give exact outputs because they are based on assumption and predictions based on a particular instant. The formula to avail a projected score is given as follows:

$$\text{Projected score} = \text{current run rate} * \text{overs in an innings}$$

DISADVANTAGE:

The mean precision gained from following the given method is very less.

5. PROPOSED ANALYSIS

The main goal of our project is to predict the runs using labeled data. Huge amount of data has to be analyzed in order to get accuracy and so the first step in achieving the purpose is to collect data of all IPL matches. Data which gives complete information about each IPL match were collected. The attributes is gathering information of all IPL matches which uses few amount of core knowledge.

ADVANTAGES:

1.) Decision Tree is one of the supervised learning Algorithms that is used for Regression as well as classification

2.) Random forest is supervised learning method. In the Random forest classifier, if the number of the trees is more then there will be best accuracy for the model.

6. METHODOLOGY

1.1. LOADING THE DATASET

The dataset is matches.csv (IPL Matches data from 2008

to 2023) whose size is 2,08,060 bytes and it is taken from the Kaggle Repository. The number of attributes is 20 and total number of records is 686.

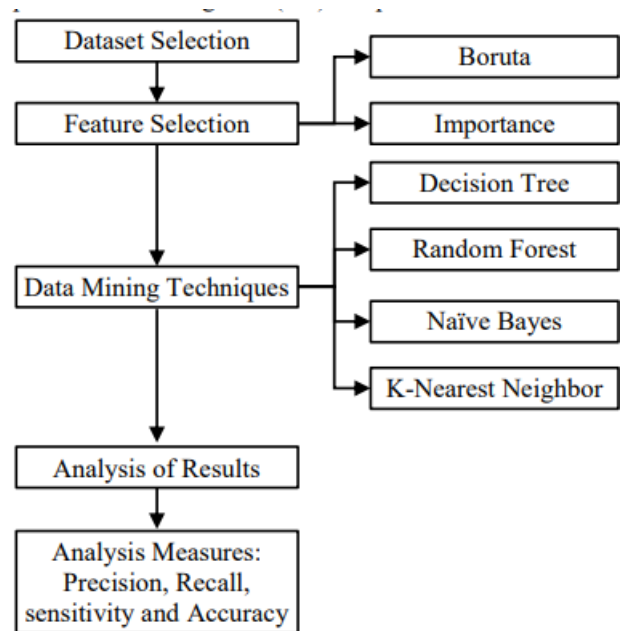
1.2. DATA PRE-PROCESSING

Data Pre-Processing is important in machine learning. It transforms raw data as a useful data format. Mostly it is used as a primary step for cleaning the data. Data Pre-Processing transforms the data into a format for error free processing for the classification.

1.3. FEATURE SELECTION

Feature selection is the next step which is the use of particular parameters in the dataset to maximize efficiency. Another name of feature selection is also known as variable selection.

7. ARCHITECTURE OF PROPOSED SYSTEM:



DATASET SELECTION:

A machine learning dataset is a collection of data that is used to train the model. A dataset acts as an example to teach the machine learning algorithm how to make predictions.

FEATURE SELECTION:

Feature selection is also called variable selection or attribute selection. It is the automatic selection of attributes in our data (such as columns in tabular data) that are most relevant to the predictive modeling problem you are working on.

DATAMINING TECHNIQUES:

Data mining includes the utilization of refined data analysis tools to find previously unknown, valid patterns and relationships in huge datasets.

It includes:

Data Source, Software Framework, Graphical Interface, Graph, Linear Regression Algorithm, Chart Plot.

ANALYSIS OF RESULTS:

After the selection of dataset, feature and data mining techniques, the results are analyzed.

ANALYSIS MEASURES:

Finally, analysis measures such as Precision, Recall, Sensitivity and Accuracy are taken.

8. WORKING

A). READING THE DATASET

	match_id	inning	battling_team	bowling_team	over	ball	batsman	non_striker
0	1	1	Sunrisers Hyderabad	Royal Challengers Bangalore	1	1	DA Warner	S Dhawan
1	1	1	Sunrisers Hyderabad	Royal Challengers Bangalore	1	2	DA Warner	S Dhawan
2	1	1	Sunrisers Hyderabad	Royal Challengers Bangalore	1	3	DA Warner	S Dhawan
3	1	1	Sunrisers Hyderabad	Royal Challengers Bangalore	1	4	DA Warner	S Dhawan
4	1	1	Sunrisers Hyderabad	Royal Challengers Bangalore	1	5	DA Warner	S Dhawan

In this we read the dataset using head() function and the command used is:

df.head(10)

B). GROUP BATSMAN ACCORDING TO MATCHES

In this we will group all the batsman according to their matches played

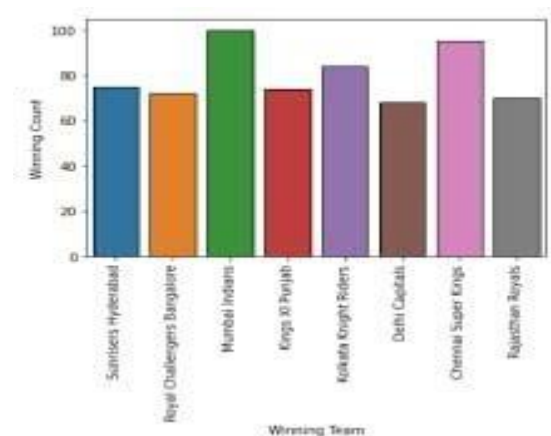
	season	match_id	inning	battling_team	batsman	batsman_runs	balls_faced	4s	6s	SR	dismissal_kind	fielder
0	2017	1	1	Sunrisers Hyderabad	BCJ Cutting	16	6.0	0.0	2.0	266.67	NaN	NaN
1	2017	1	1	Sunrisers Hyderabad	DA Warner	14	8.0	2.0	1.0	175.00	caught	Mandeep Singh
2	2017	1	1	Sunrisers Hyderabad	DJ Hooda	16	12.0	0.0	1.0	133.33	NaN	NaN
3	2017	1	1	Sunrisers Hyderabad	MC Henriques	52	37.0	3.0	2.0	140.54	caught	Sachin Baby
4	2017	1	1	Sunrisers Hyderabad	S Dhawan	40	31.0	5.0	0.0	129.03	caught	Sachin Baby
5	2017	1	1	Sunrisers Hyderabad	Yuvraj Singh	62	27.0	7.0	3.0	229.63	bowled	NaN
6	2017	1	2	Royal Challengers Bangalore	A Choudhary	6	2.0	0.0	1.0	300.00	NaN	NaN
7	2017	1	2	Royal Challengers Bangalore	CH Gayle	32	21.0	2.0	3.0	152.38	caught	DA Warner
8	2017	1	2	Royal Challengers Bangalore	KM Jadhav	31	16.0	4.0	1.0	193.75	run out	BCJ Cutting
9	2017	1	2	Royal Challengers Bangalore	Mandeep Singh	24	16.0	5.0	0.0	150.00	bowled	NaN

C). GROUP BOWLERS ACCORDING TO SET OF DATA

	season	match_id	inning	bowling_team	bowler	over	wide_runs	noball_runs	runs	extras	wickets	Econ
0	2017	1	1	Royal Challengers Bangalore	A Choudhary	4	3	1	55	4	1.0	13.75
1	2017	1	1	Royal Challengers Bangalore	S Arvind	3	0	0	36	0	0.0	12.00
2	2017	1	1	Royal Challengers Bangalore	SR Watson	3	0	0	41	0	0.0	13.67
3	2017	1	1	Royal Challengers Bangalore	STR Binny	1	0	0	10	0	1.0	10.00
4	2017	1	1	Royal Challengers Bangalore	TM Head	1	0	0	11	0	0.0	11.00
5	2017	1	1	Royal Challengers Bangalore	TS Mills	4	2	0	31	2	1.0	7.75
6	2017	1	1	Royal Challengers Bangalore	YS Chahal	4	0	0	22	0	1.0	5.50
7	2017	1	2	Sunrisers Hyderabad	A Nehra	4	1	0	42	1	2.0	10.50
8	2017	1	2	Sunrisers Hyderabad	B Kumar	4	1	0	27	1	2.0	6.75
9	2017	1	2	Sunrisers Hyderabad	BCJ Cutting	4	2	0	35	2	0.0	8.75

- In this we will group the bowlers according to the set of data

E). EXPLORATORY DATA ANALYSIS



- In this we will analyze and explore the dataset using statistical methods and the describe() method is used for EDA(Exploratory Data Analysis)

9. TECHNOLOGIES USED

A. Jupyter Notebook-

For designing frontend of webpage.

B. Netbeans-For designing the back end of the webpage.

C. MsExcel-For collecting datasets.

10. CONCLUSION

Our project helped us a lot in learning machine learning concepts. Analyzing a winner in IPL is challenging which involves very complicated process. But with the help of Machine Learning, it can be made easier. In our paper, several factors have been found that provide a way to the outcome of IPL matches.

11. REFERENCES

- 1.) "Predicting Outcome of Indian Premier League (IPL) Matches Using Machine Learning" (September 2018) Rabindra Lamsal and Ayesha Choudhary.
- 2.) "Winning Prediction Analysis in One Day- International (ODI) Cricket Using Machine Learning Techniques", International Journal of Emerging Technology and Computer Science, (April 2018) Abhishek Naik, Shivaneepawar, Minakshree Naik, Sahil Mulani.
- 3.) "Prediction of the outcome of a Twenty-20 Cricket Match." (2015) Arjun Singhvi, Ashish Shenoy, Shruthi Racha, and Srinivas Tunuguntla.
- 4.) "Analysis on Attributes Deciding Cricket Winning", International Research Journal of Engineering and Technology (IRJET), (March 2017) Swetha, Saravanan. KN.
- 5.) "A Review Paper on Cricket Predictions Using Various Machine Learning Algorithms and Comparisons among Them", International Journal for Research in Applied Science & Engineering. IJRASET 17099 (April 2018)