



CRICKET(IPL) PREDICTION BASED ON PREVIOUS DATA

L. Eswar Kumar , V. Yagnesh, U.Likith, M. Suresh
Department of Computer Science and Engineering,
Kalasalingam Academy of Research and Education,
Virudhnagar, Tamil nadu ,India

ABSTRACT:

Cricket is one of the most popular sport in nowadays, and success is determined by a variety of criteria such as home ground advantage, player performance in prior matches, and so on ,In recent days a lot of research has been done to measure the performances of player's and predicts the winning probability/percentage. In this work, we use a supervised learning strategy to predict the outcome of an Indian Premier League (IPL) cricket match from the standpoint of team composition. Our findings imply that predicting team strength is possible. We use statistic and recent performance of a team; we use machine learning technique to predict the outcome of a match based on the previous years data.

Keywords- Classification Algorithm, Decision tree, Random Forest.

I. INTRODUCTION

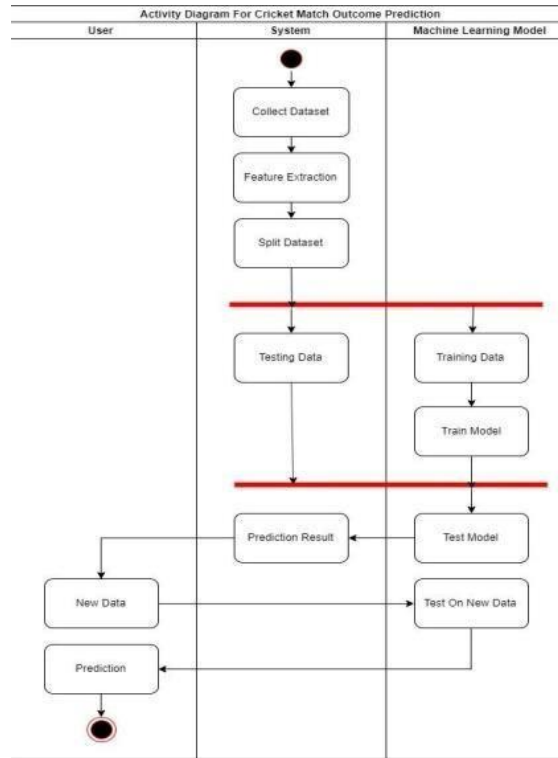
Sports analytics play a critical role in a variety of sports-related issues. The rating of individual players and their specialized talents, the construction of teams with an ideal balance of specialized skills, and the ranking of teams are some of these issues, the negotiation of contracts, their potential revenue streams, the planning of both physical and mental training, the development of strategies for winning games and injuries (health and insurance), the analysis and improvisation of rules, Quality of equipment and technology, prize determination, historical records, and the production of odds for gambling activities are all factors to consider. Sports analytics play a significant part in a variety of sports- related issues. Award determination, historical records, and the calculation of odds for gambling activities are all part of this process. Sports analytics has progressed to the point where both the technology that delivers data and the statistical approaches that give the tools for interpreting it have rapidly advanced. Despite the fact that sports analytics is fast evolving, this has not been the case with cricket. Due to historical reasons, cricket was not subject to large financial transactions because it was seen as a leisurely gentleman's game played without remuneration to players (until recently). This has changed in recent years as a result of shorter formats. The shortest and newest format, known as T20, generates intense interest and vast sums

of money, especially in the Indian subcontinent. The demand for cricket analytics has risen in tandem, with cricinfo.com serving as the primary source of cricket data and information. Our research is focused on IPL. We present an algorithm to predict the outcome of IPL cricket



matches. we propose an approach where we first estimate the previous year's performances of the team using statistics.

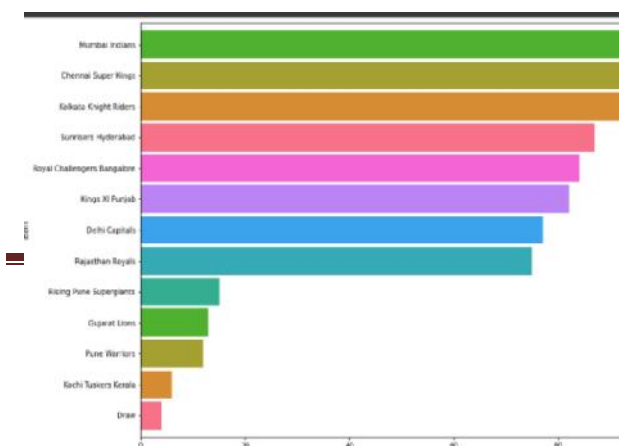
II. RELATED WORK



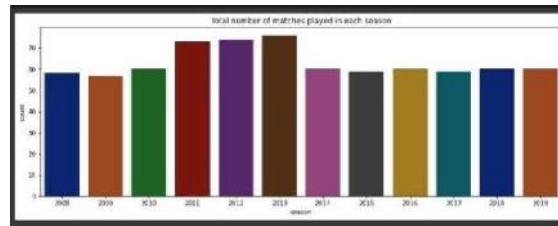
The methodology used in these models is classification algorithm. Alongside, different machine learning classification algorithms like Naïve Bayes, SVM, k- Nearest Neighbour, Random Forest, Logistic Regression, ExtraTreesClassifier, XGBoost were used to train the models for predicting the winner. The dataset contains data of IPL matches from 2008 to

2019. Which we are going to predict for

2020. First of all import all necessary packages and libraries. And then we have to train, the data of previous year's IPL matches. Check if any null values present inside the dataset. The presence of null values decrease the accuracy rate of our



1) Prediction Based On Previous Data



model. The dataset contains the null values may decrease the accuracy, so it is better to remove the null rows or we have to fill them some other related data. I'm going to fill the city null values with Abu Dhabi and the winner null values with draw.

we have to visualize the data that the total number of matches played in each season and then the total number of matches won by each team.

Another thing that the team names are larger size for easy training and test data set we replace them in a short form. Also, the we have change the names of the teams and different teams have participated in different years which we have to remove that.

And finally we have change the text to numerical using Label Encoder, and then train the model which gives the most accuracy.

III. METHODOLOGY

Our project's effort is focused on two models. They are:

- 1.Descriptive model
- 2.Predictive model

1.DESRIPTIVE MODEL:

The descriptive model focused mainly on two aspects:

It describes the data and statistics of the previous year's information of the teams that played.

It gives the information of the matches played by the IPL teams.

PREDICTIVE MODEL:

In this model we have to predicting the winning percentage of the team. Based on the previous year data, split the data into training and testing like 80% training data and 20% testing data.

After completion of training the data we have to test the data using Classification Algorithms like Random Forest , Decision Tree and k nearest Neighbour



,logistic regression, SVM, XGBoost for predicting the final result of the match, Out of these Decision tree gives the Accuracy is 81.34

```
knn_model = KNeighborsClassifier()
knn_model.fit(X,y)
print("KNeighbor Classifier accuracy", (knn_model.score(X,y))*100)

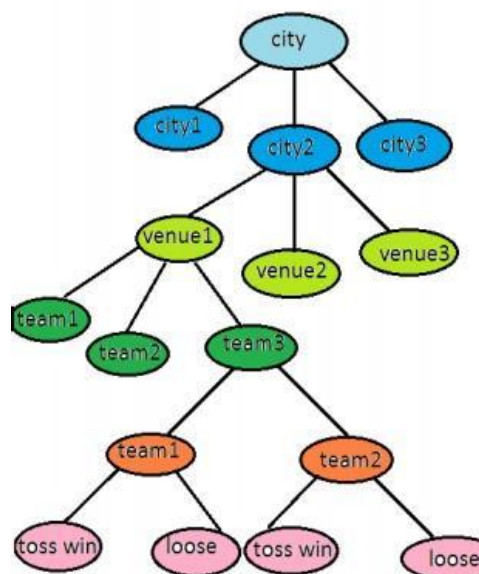
Neighbor Classifier accuracy 62.56613756613757
/usr/local/lib/python3.7/dist-packages/sklearn/neighbors/_classification.py:198
return self._fit(X, y)

from sklearn.tree import DecisionTreeClassifier

decision_model = DecisionTreeClassifier()
decision_model.fit(X,y)
print("Decision Tree Classifier accuracy: ", (decision_model.score(X,y))*100)

Decision Tree Classifier accuracy: 81.48148148148148
```

Random forest gives the 81.34 and remaining all algorithm gives the lowest accuracy. So, we can choose our model which gives the most Accuracy i.e Decision tree and Random Forest. A decision tree is built upon using the top down approach in this tree contain the root node ,sub node ,leaf node. This tree is built according to the prominent factors like city, venue, teams, toss decisions are considered in the match

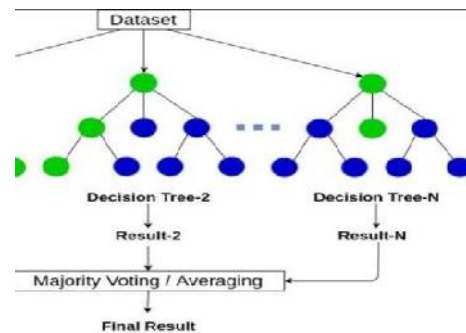


And we can approach Random forest Algorithm to predict the outcome of a match.



```
Classifier()  
: ", (Random_model.score(X,y))*100)  
-packages/ipykernel_launcher.py:2: D  
8148148148148
```

, Paper 50 August 2022



one of the major advantage of random forest is ability is used in both in classification and regression and it is formed as combining of multiple trees to give better accuracy and it decrease the overfitting problems in order to give best accuracy

IV. RESULT

By using decision tree classifier we can find the outcome of a match based on the previous year's data ,and this helps the people who are predicting the outcome of a match.



CONCLUSION

Our main Aim in this paper is to develop a model that predict the outcome of an IPL matches before the game is going to start. We used the previous year's data from 2000 to 2018 in order to design the model. This knowledge will help's us to analysis the matches and gives most accurate prediction.

REFERENCES :-

- [1] Mrs.J.Jayanthi Professor at Kalasalingam University, Department of Computer Science and Engineering.
- [2] jssm.org-Predicting the Match Outcome in One Day [Online]. <http://www.jssm.org/volume05/iss4/cap/jssm-05-480.pdf&p=DevEx.LB.1,5063.1&p=DevEx.LB.1,5063.1>
- [3] Optimal batting orders in one-day cricket, T. B. Swartz, P. S. Gill, D. Beaudoin, and B. M. Desilva, Computers and Operations Research, vol. 33, no. 7, pp. 1939–1950, 2006.
- [4] Data Analytics based Deep Mayo Predictor for IPL-9, C. Deep, C. Patvardhan, and C. Vasantha, International Journal of Computer Applications, vol. 152, no. 6, pp.6–11, 2016.
- [5] Role of External Factors on the Outcome of a One Day International Cricket (ODI) Match and Predictive Analysis, Mehvish Khan and Riddhi Shah, International Journal of Advanced Research in Computer and Communication Engineering, vol. 4, no. Issue 6, pp. 192– 197, Jun. 2015.