Analysing and Predicting Outcome of IPL Cricket Data

Prof. Rajitha P R[1], Anjali S Kumar[2] and Dr. Mahalekshmi T[3]

Assistant Professor, Master of Computer Application[1]
Student, Final Year Master of Computer Application[2]
Principal, Sree Narayana Institute of Technology[3]
Sree Narayana Institute of Technology, Kollam, Kerala

Abstract: *Fundamental concepts of analytics and predictive modeling to IPL cricket matches will be applied to get meaningful information and predictions. In this paper, the past seven years data of IPL containing the players details, is taken and analyzed to draw various conclusions which help in the improvement of a players performance. Various other features like how the venue or toss decision has influenced the winning of the match in last seven years are also predicted. Various machine learning and data extraction models are considered for prediction are Linear Regression, Decision Tree, K-means, Logistic Regression. The cross validation score and the accuracy are also calculated using various machine learning algorithms. Before prediction we have to explore and visualize the data because data exploration and visualization is an important stage of predictive modeling.*

Keywords: Machine Learning, Cricket, Prediction

I. INTRODUCTION

The Indian Premier League (IPL) is a Twenty-20 cricket tournament league established with the objective of promoting cricket in India and thereby nurturing young and talented players. The league is an annual event where teams representing different Indian cities compete against each other. The teams for IPL are selected by means of an auction. Players' auctions are not a new phenomenon in the sports world. However, in India, selection of a team from a pool of available players by means of auctioning of players was done in Indian Premier League (IPL) for the first time. Due to the involvement of money, team spirit, city loyalty and a massive fan following, the outcome of matches is very important for all stake holders. This, in turn, is dependent on the complex rules governing the game, luck of the team (Toss), the ability of players and their performances on a given day. Various other natural parameters, such as the historical data related to players, play an integral role in predicting the outcome of a cricket match.

A way of predicting the outcome of matches between various teams can aid in the team selection process. However, the varied parameters involved present significant challenges in predicting accurate results of a game. Moreover; the accuracy of a prediction depends on the size of data used for the same. The tool presented in this paper can be used to evaluate the performance of players. This tool provides a visualization of player's performance. Further, several predictive models are also used for predicting the result of a match, based on each player's performance as well as some match related data[1][3][5][7].

II. BACKGROUND

Technologies are used:

- **Tkinter**: Python offers multiple options for developing GUI .out of all the GUI methods, tkinter is the most commonly used method. It is a standard python interface to the Tk GUI toolkit shipped with Python. Python with tkinter is the fastest and easier way to create the GUI application. Tkinter is the standard GUI library for python.Tkinter provides a powerful object oriented interface to the Tk GUI toolkit. Tk is called Tkinter in Python, or to be precise, Tkinter is the Python interface for Tk. Tkinter is an acronym for "Tk interface". Tk was developed as a GUI extension for the Tcl scripting language by John Ousterhout. The first release was in 1991. Tk proved as extremely successful in the 1990's, because it is easier to learn and to use than other toolkits. Tkinter is Python's de-facto standard GUI (Graphical User Interface) package. It is a thin

objectoriented layer on top of Tcl/Tk. Tkinter is not t he only GuiProgramming toolkit for Python. Python has a lot of GUI frameworks, but Tkinter is the only framework that's built into the Python standard library.

- **Anaconda** is a free and open-source distribution of the Python and R programming languages for scientific computing (data science, machine learning applications, large-scale data processing, predictive analytics, etc.), that aims to simplify package management and deployment. The distribution includes data-science packages suitable for Windows, Linux, and macOS. It is developed and maintained by Anaconda, Inc., which was founded by Peter Wang and Travis Oliphant in 2012.As an Anaconda, Inc. product, it is also known as Anaconda Distribution or Anaconda Individual Edition, while other products from the company are Anaconda Team Edition and Anaconda Edition, which are both not free. Package versions in Anaconda are managed by the management system conda. This package manager was spun out as a separate open-source package as it ended up being useful on its own things than Python. There is also a small, bootstrap version of Anaconda called Miniconda includes only conda,

III. EXISTING SYSTEM

In the existing system, a number of manual techniques and features sets have been used to development and implement IPL results prediction systems. There will be variations in the system based on the expertise and manual analysis of the person who involves in this.

IV. PROPOSED SYSTEM

The proposed prediction model makes use of multivariate Regression to calculate points of each player in the league and compute the overall strength of each team based on the past performance of the players who appeared most for the team. The developed models can help decision makers during the IPL matches to evaluate the strength of a team against another.

A) Description

- This methodology consists of 4 main stages processing, Data Cleaning, Data Preparation, Encthe data. Initially, the seven IPL real taken in CSV format.
- In data pre-processing phase, and correct the inconsistencies.
- In data cleansing phase, data validation is done by maintaining consistency across the dataset and data enhancement id done by adding related information to the dataset.
- The data preparation is significant for achieving optimal results[9].

B) Advantages

- Can be used by the team management to effective decision for selecting team based on the form of the individual players.
- To provide the statistical analysis of players based on different characteristics.
- To predict the performance of a team depending on individual player statistics.
- To predict successfully predict the outcome of IPL matches[9].

C) Algorithm Used

Different multi-classification algorithms such as Linear Regression, Decision Tree, K-means, and Logistic Regression[14] are implemented to predict the accuracy and cross-validation score. Before prediction we have to explore and visualize the data because data exploration and visualization is an important stage of predictive modeling[12][13].

Generic function for making a classification model and accessing performance:

```
def classification_model(model, data, predictors, outcome):
    model.fit(data[predictors],data[outcome])
```

```
predictions = model.predict(data[predictors])
accuracy=metrics.accuracy_score(predictions,data
print('Accuracy : %s' % '{0:.3%}'.format(accuracy))
kf = KFold(data.shape[0], n_folds=5)
error = []
for train, test in kf:
train_predictors=(data[predictors].iloc[train,:])
train_target=data[outcome].iloc[train]
model.fit(train_predictors, train_target)
error.append(model.score(data[predictors].iloc[test,:],
data[outcome].iloc[test]))
print('CrossValidationScore:%s'%'{0:.3%}'.format(np.mean(
error)))
model.fit(data[predictors],data[outcome])
```

V. SYSTEM STRUCTURE
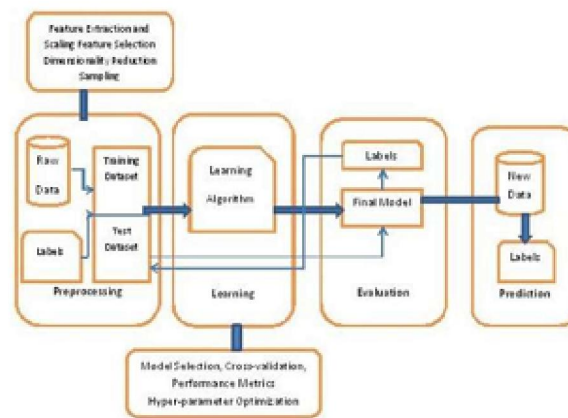
Figure: System Structure

VI. RESULTS AND DISCUSSIONS

Figure: Prediction

Figure 3: Match Understanding

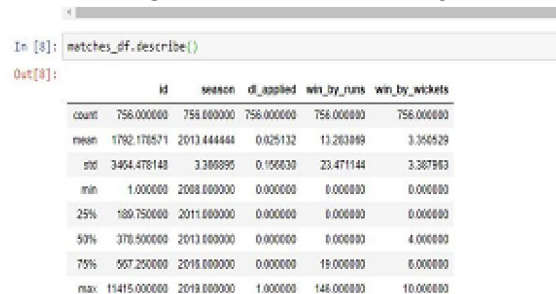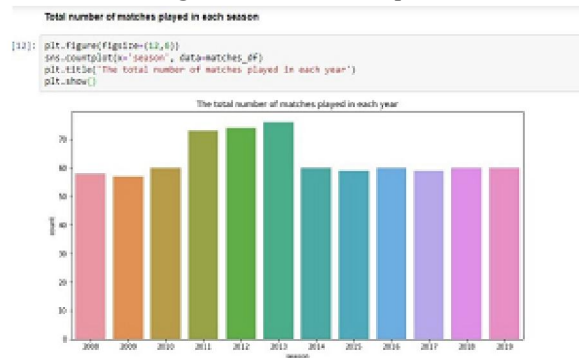**Figure 4:** Match Description

Figure 5: Visualization

VII. CONCLUSION

Selection of the best team for a cricket match plays a significant role for the team's victory. The main goal of this paper is to analyze the IPL cricket data and predict the players' performance. Here, three classification algorithms are used and compared to find the best accurate algorithm. The implementation tools used are Anaconda navigator and Jupiter. Random Forest is observed to be the best accurate classifier with 89.15% to predict the best player performance. This knowledge will be used in future to predict the winning teams for the next series IPL matches. Hence using this prediction, the best team can be formed[1][8][11].

REFERENCES

[1]. Passi, Kalpdrum & Pandey, Niravkumar. (2018) "Predicting Players' Performance in One Day International Cricket
[2]. Matches Using Machine Learning" 111-126. 10.5121/csit.2018.80310.

[3]. I.P. Wickramasinghe et. al, "Predicting the performance of batsmen in test cricket," Journal of Human Sport & Exercise", vol. 9, no. 4, pp. 744-751, May 2014.

[4]. R. P. Schumaker, O. K. Solieman and H. Chen, "Predictive Modeling for Sports and Gaming" in Sports Data Mining, vol. 26, Boston, Massachusetts: Springer, 2010.

[5]. J. McCullagh, "Data Mining in Sport: A Neural Network Approach," International Journal of Sports Science and Engineering, vol. 4, no. 3, pp. 131-138, 2012.

[6]. Bunker, Rory & Thabtah, Fadi. (2017) "A Machine Learning Framework for Sport Result Prediction. Applied Computing and Informatics", 15. 10.1016/j.aci.2017.09.005.

[7]. Ramon Diaz-Uriarte and Sara, "Gene selection and classification of microarray data using random forest, BMC Bioinformatics", doi:10.1186/1471-2105-7-3

[8]. Rabindra Lamsal and AyeshaChoudhary, "Predicting Outcome of Indian Premier League (IPL) Matches Using Machine Learning" Akhil Nimmagadda et. Al, "Cricket score and winning prediction using data mining", IJARnD Vol.3, Issue3.

[9]. Ujwal U J et. At, "Predictive Analysis of Sports Data using Google Prediction API" International Journal of Applied Engineering Research", ISSN 0973-4562 Volume 13, Number 5 (2018) pp. 2814-2816.

[10]. Rameshwari Lokhande and P.M.Chawan, "Live Cricket Score and Winning Prediction", International Journal of Trend in Research and Development, Volume 5(1), ISSN: 2394-9333.

[11]. Abhishek Naiket. Al, "Winning Prediction Analysis in One-Day-International (ODI) Cricket Using Machine Learning Techniques", IJETCS, vol. 3, issue 2, ISSN:2455-9954, April 2018.

[12]. Esha Goel and Er. Abhilasha, " Random Forest: A Review", IJARCSSE, Volume 7, Issue 1, DOI: 10.23956/ijarcsse/V7I1/01113, 2017.

[13]. Amit Dhurandhar and Alin Dobra, " Probabilistic Characterization of Random Decision Trees", Journal of Machine Learning Research, 2008.

[14]. H. Yusuff et. Al, "BREAST CANCER ANALYSIS USING LOGISTIC REGRESSION", IJRRAS, Vol. 10, Issue 1, 2012.