ICC T20 Cricket World Cup 2023 Winner Prediction Using Machine Learning Techniques

Anupam Tiwari
anupam.tiwari@spit.ac.in
Master of Computer Applications

'Sardar Patel Institute of Technology-Andheri(W),Mumbai:400058,India'

Abstract- This paper predict who will win the upcoming **ICC T20 world cup 2023**.Cricket is one of the most popular team games in the world.Machine learning has served to amplify this magic and uncover the mystery. It has also served in the fields of sports. Billions of people around the world are big fans of Cricket and wait for the results eagerly.We have compared popular machine learning techniques for the prediction of the T20 cricket world cup winner. Among the developed models, Decision Tree Algorithm proved to be the best machine learning algorithm using a custom accuracy metric. It obtained a custom accuracy of 98.13%. In this analysis, India emerged as the winner of the T20 world cup 2023. For this purpose, the ESPN Cricinfo dataset has been used.

Keywords- Machine learning, T20 world cup 2023, winner prediction, XGBRegression,Decision tree, random forest algorithm,KNN Algorithm

I. INTRODUCTION

Machine Learning has been used in sports for decades and has contributed significantly to the success of the field.Cricket is one of the most popular sports in the world.Various[1] natural factors affecting the game ,enormous media coverage, and a huge betting market have given from various perspective.However,the complex rules governing the game,the ability of players and their performances on a given day, and various other natural parameters play integral role in affecting the outcome of a cricket match.This presents significance challenge in predicting the accurate results of a game.The game of cricket is played in three formats -Test matches ODIs and T20,the most popular format of the game.To predict the outcome[3] of T20 cricket matches,I will propose an approach where I first estimate the batting and bowling potentials of the 22 players the match using their career statistics and active participation in recent games.I will use these player potentials to render the relative dominance one team has over the other.Taking two other base features into account,namely,toss decision and the venue of the match, along with the relative team strength ,I adopt supervised learning algorithms to predict the winner of the match.

II. LITERATURE REVIEW

Cricket is one of the spectator games at this time. Prediction of match results before the tournament is an interesting matter to the spectator, player, and sponsor.[2] Several studies have been conducted to predict match results or tournament results. A study by Saqlain et al [5] is organized to predict the ICC T20 cricket world of 2020 through using the TOPSIS technique. In the TOPSIS method, they used different factors such as pitch and weather conditions, player performance (Century, catch drop, fielding miss), team performance (series win, series loss).From their study, they concluded that India had the highest chance of winning the T20 world cup and Afghanistan had the lowest chance of winning.

III. PROPOSED METHODOLOGY

.

For our experiment, I have used kaggle datasets..In this datasets, I scrapped data from 2010 May to 2021 May. Our first data set consisted of thirteen attributes like the team name, playing span, the total match played, won, lost, tied, average, Run per over, innings played total, highest run total, lowest run total,win-loss ratio, and last on is the class value. Where i divided the win/lost ratio and get the number.The datasets which we will be used for the initial entries of the training of the system is assumed to be the correct input for the system. The

training dataset is taken from the online repositories such as Kaggle.

Here,I have followed following steps on T20 datasets:

1) Collecting Datasets
2) Pre-Processing
3) Training the datasets
4) Testing the datasets
5) Model Evaluation
6) Predicts Output

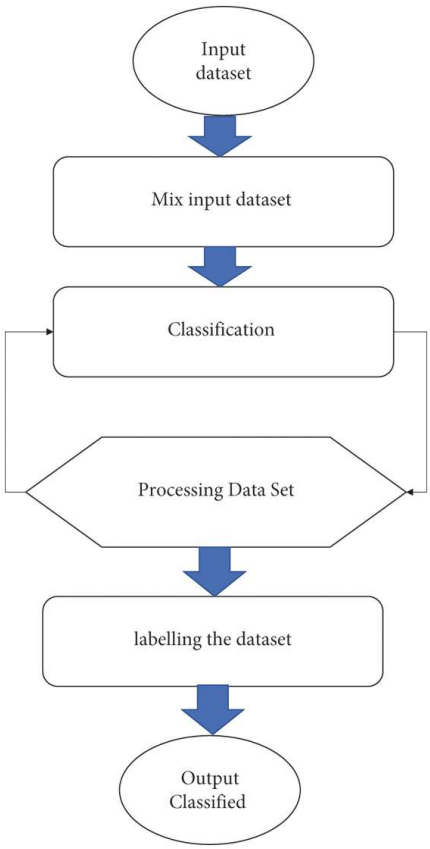

Fig 3. Proposed Methodology

Following are the Algorithms used for training the datasets.

1) Random Forest Algorithm

Random Forest is a popular machine learning algorithm that belongs to the supervised leaning technique.It can be used for both Classification and Regression Problems In ML.It is based on concept of ensemble learning,which is a process of combining concept of ensemble learning ,which is a process of combing multiple classifiers to solve a complex problem and to improve the performance of the model.Random Forest Algorithm has predicated accuracy approx.97.96 % by using T20 World Cup datasets

2) Decision Tree Algorithm

Decision Tree is a supervised learning technique that can be used for both classification and Regression problems,but mostly it is preferred for solving classification problems.It is a Tree-structured classifier,where internal nodes represents the features of a datasets ,branches represents the decision rules and each leaf node represents the outcome.Decision Tree Algorithm has predicated accuracy approx.98.13 % by using T20 World Cup datasets.

3) KNN Algorithm

The K-nearest neighbors classifier(KNN)is non-parametric supervised machine learning algorithm.It's distance-based. It classifies objects based on their approximate neighbors' classes.KNN is most often used for classification,but can be applied to regression problems as well.KNeighbors Classifier Algorithm has predicated Accuracy approx. 79.04 % by using T20 World Cup datasets.

4) XGBRegressor Algorithm

XGBoost is a powerful approach for building supervised regression models.The validity of this statement can be inferred by knowing about its (XGBoost) objective function and base learners.The objective function and a regularization term.It tells about the difference between actual values and predicated values,i.e. how far the model results are from the real values.XGBRegressor Algorithm has predicated Accuracy approx. 94.82 % by using T20 World Cup datasets.

5) SVM Algorithm

Support Vector Machine or SVM is one of the most popular Supervised Learning algorithms, which is used for Classification as well as Regression problems. However, primarily, it is used for Classification problems in Machine Learning.The goal of the SVM algorithm is to create the best line or decision boundary that can segregate n-dimensional space into classes so that we can easily put the new data point in the correct category in the future.SVM Algorithm has predicated Accuracy approx. 83.45 % by using T20 World Cup datasets.

IV. RESULT

Fig. 4. displays the comparative performance of five Machine learning algorithms (Randome Forest, Decision Tree, KNN,RGBRegressor and SVM).

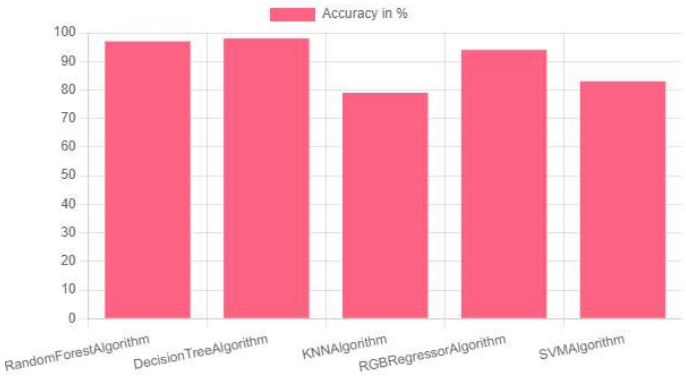

Fig4.Accuracy Comparision of Prediction Algorithms

From the above Comaprision,Decision Tree algorithm proved to be the best among five selected machine learning algorithm.Decision Tree algorithm achieved 98.13% custome accuracy.

V. CONCLUSION

In conclusion, This study is an exploratory study aimed to predict the winner of the upcoming ICC T20 World Cup by utilizing a few prominent parameters associated with T20 previous matches.It is apparent from the accuracy comparison models that the Decision Tree algorithm gives the highest prediction as it is able to find the most important feature from the selected feature set and is less prone to overfitting.

VI. REFERENCES

[1] N. Pathak and H. Wadhwa, “Applications of Modern Classification Techniques to Predict the Outcome of ODI Cricket,” Procedia Comput. Sci., vol. 87, pp. 55–60, 2019, doi: 10.1016/j.procs.2016.05.126.

[2] S. Kampakis and W. Thomas, “Using Machine Learning to Predict the Outcome of English County twenty over Cricket Matches,” pp. 1–17, 2015, [Online]. Available: http://arxiv.org/abs/1511.05837.

[3] K. Passi and N. Pandey, “Increased Prediction Accuracy in the Game of Cricket Using Machine Learning,” Int. J. Data Min. Knowl. Manag. Process, vol. 8, no. 2, pp. 19–36, 2018, doi: 10.5121/ijdkp.2018.8203.

[4] M. Yasir, L. I. Chen, S. A. Shah, K. Akbar, and M. U. Sarwar, “Ongoing Match Prediction in T20 International,” vol. 17, no. 11, pp. 176–181, 2019.

[5] P. The et al., “Predicting The Cricket Match Outcome Using Crowd Opinions On Social Networks: A Comparative Study Of Machine Learning Methods. pp 63-76,” vol. 30, no. 1, pp. 63–76, 2019.

[6] G. S. K. Paramesha, “A Perspective on Analyzing IPL Match Results using Machine Learning,” vol. 7, no. 03, pp. 1476–1479, 2019.

[7] V. Q. L. N. Hdvhu et al., “Player ’ s Performance Prediction in ODI Cricket Using Machine Learning Algorithms.”

[8] P. The et al., “Predicting The Cricket Match Outcome Using Crowd Opinions On Social Networks: A Comparative Study Of Machine Learning Methods. pp 63-76,” vol. 30, no. 1, pp. 63–76, 2021.

[9] M. G. Jhanwar and V. Pudi, “Predicting the outcome of ODI cricket matches: A team composition based approach,” CEUR Workshop Proc., vol. 1842, no. September, 2016.

[10] K. Desai, “Predicting Outcome of ODI Cricket Games,” vol. 3, no. 01, pp. 1–5, 2020.