IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

T20 CRICKET WORLD CUP 2024 PREDICTION USING MACHINE LEARNING

[1]Sudarsanan D, [2]Mohammed Mafaaz Chandwale, [3]Ryan Ahmed, [4]Harsh Nath Mishra

[1] Assistant Professor, [2] Student, [3] Student, [4] Student, [5] Student

Department of Information Science and Engineering,
Cambridge Institute of Technology, Bangalore, India

Abstract: This study applies machine learning (ML) techniques to predict Cricket World Cup winners, using historical data, team performances, and player stats. Comprehensive datasets from past tournaments are analyzed with algorithms like Random Forests and Logistic Regression, enhanced through cross-validation. Models are trained on diverse match scenarios and team performance data, aiming to forecast the champion accurately.

Index Terms **- Component, formatting, style, styling, insert.**

I. INTRODUCTION

In the realm of international cricket, the ICC Cricket World Cup holds a special place. It's a stage where nations pour their heart and soul into the game, igniting the fervour of millions of fans worldwide. The ICC 2023 Cricket World Cup, hosted in the cricket-crazy nation of India, promises to be an unforgettable spectacle. It is a time when the hopes and dreams of fans, players, and teams collide. However, amidst the anticipation and excitement, lies a challenge – predicting the tournament's winner with precision. This is where the world of sports and technology converge.

With the advent of machine learning, data analysis, and web scraping, predicting the winner of such a prestigious event has evolved into a data-driven quest. It's not just about gut feelings or intuition; it's about harnessing the power of data and technology to make informed predictions. In this essay, we will delve into the fascinating journey of using data as a crystal ball to foresee the outcome of the ICC 2023 Cricket World Cup, revealing the captivating synergy between the grandeur of sports and the precision of modern technology.

At the heart of this predictive endeavour lies the Random Forest (RF) classifier, a powerful machine learning algorithm renowned for its ability to handle complex datasets and make accurate predictions. By training the RF classifier on historical cricket match data, including team performance, match venues, and other relevant features, we aim to create a robust predictive model capable of forecasting the tournament's outcome.

However, before feeding the data into the RF classifier, a crucial step is data cleaning. This involves preprocessing the dataset to handle missing values, remove duplicates, and standardize the format for consistency. Additionally, irrelevant features are filtered out to focus only on those that contribute significantly to the prediction task. By ensuring the integrity and quality of the data, we lay the foundation for a reliable predictive model.

Once the data is cleaned and pre-processed, it is fed into the RF classifier to train the model. The classifier learns from past patterns and trends in the data to make predictions about future outcomes. Through iterative training and validation, the model fine-tunes its parameters to maximize predictive accuracy.

To interact with and visualize the predictions made by the RF classifier, an intuitive interface is developed. This interface provides users with real-time updates on match fixtures, predicted outcomes, and tournament standings. Users can explore various scenarios, simulate matches, and analyse the model's predictions to gain insights into the tournament dynamics.

In conclusion, the fusion of sports and technology in predicting the outcome of the ICC 2023 Cricket World Cup exemplifies the evolving landscape of predictive analytics in the world of sports. By leveraging the power of machine learning, data analysis, and web scraping, we embark on a data-driven quest to unravel the mysteries of cricket's most prestigious event. It's a journey that not only showcases the enchanting fusion of the fervour of sports and the precision of contemporary technology but also underscores the transformative potential of predictive analytics in shaping the future of sports.

II. METHODOLOGY

2.1Population and Sample

Our population consists of all participating teams in the ICC T20 Cricket World Cup. This encompasses a diverse range of cricketing nations, each bringing its unique style and strengths to the tournament. To construct our sample, we select specific matches or teams within the competition. These matches serve as representative examples for analysis, allowing us to draw conclusions about broader trends in the tournament.

2.2 Data and Sources of Data

For this study secondary data has been collected. From the website of cricksheet for the ball-by-ball sample data and their results of the all the T20 International matches played. From the website of crickinfo the data for the fixtures of the tournament is collected.

2.3 System Architecture

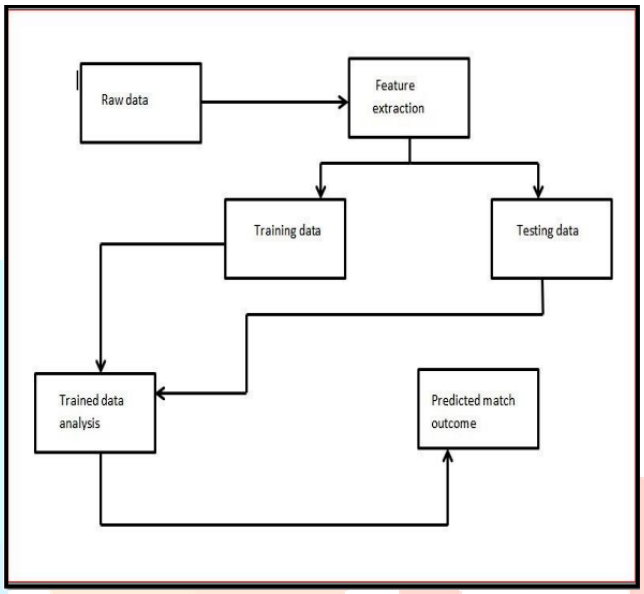

Fig 2.1: System Architecture

The system architecture for predicting match outcomes in the ICC T20 Cricket World Cup begins with the collection of raw data from various sources including historical match data, player statistics, team compositions, pitch conditions, and weather forecasts. Once collected, the raw data undergoes feature extraction, where relevant variables and metrics are identified and transformed into meaningful features that may influence match outcomes. Following feature extraction, the data is prepared for training by undergoing processes such as cleaning, normalization, encoding categorical variables, and splitting into training and testing datasets. Most of the data is allocated to the training set, while the remainder is reserved for testing. With the data ready, various machine learning algorithms are applied to the training set to build predictive models. These models are trained to minimize prediction errors and adjust their parameters accordingly.

Once trained, the models undergo evaluation using the testing dataset to assess their performance and generalization capabilities. Performance metrics such as accuracy, precision, recall, F1-score, and confusion matrix are computed to measure the effectiveness of the models in predicting match outcomes. This analysis helps in identifying the strengths and weaknesses of different models, enabling fine-tuning of parameters for improved performance. Following evaluation, the trained models are deployed to predict match outcomes for upcoming matches in the ICC T20 Cricket World Cup. These predictions are based on input features such as team compositions, player form, pitch conditions, and weather forecasts. The output of the models provides valuable insights for cricket enthusiasts, analysts, and stakeholders, aiding in decision-making and strategy formulation for the tournament.

2.4 Model Used

A Random Forest Algorithm is a supervised machine learning algorithm that is extremely popular and is used for Classification and Regression problems in Machine Learning. We know that a forest comprises numerous trees, and the more trees it will be robust. Similarly, the greater the number of trees in a Random Forest Algorithm, the higher its accuracy and problem-solving ability. Random Forest is a classifier that contains several decision trees on various subsets of the given dataset and takes the average to improve the predictive accuracy of that dataset. It is based on the concept of ensemble learning which is a process of combining multiple classifiers to solve a complex problem and improve the performance of the model.

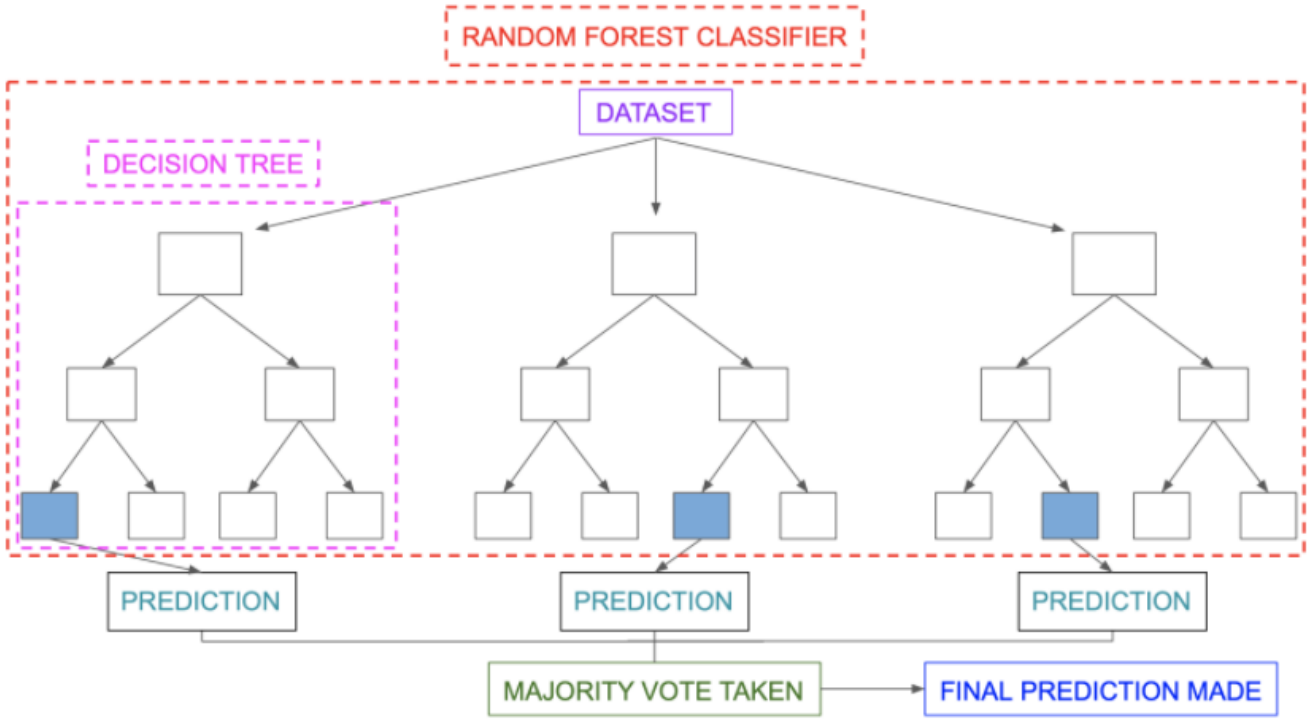

Fig 2.2 Random Forest Algorithm

III. IMPLEMENTATION

Predicting the outcome of the T20 Cricket World Cup 2024 using machine learning involves several essential steps. Initially, historical data from previous tournaments, encompassing match results and venue specifics, needs to be collected. This data is then preprocessed to handle missing values, encode categorical variables, and scale numerical features, ensuring it's ready for model training. Feature selection follows, identifying the most influential factors using techniques like correlation analysis and feature importance ranking. Subsequently, the machine learning model Random Forest was chosen for prediction. The model is trained on preprocessed data, employing techniques like cross-validation for optimal performance. Once trained, the models are utilized to forecast match outcomes for the upcoming T20 Cricket World Cup matches. However, it's important to acknowledge the inherent uncertainty in predicting sports outcomes, particularly in cricket, necessitating ongoing refinement and adaptation of the predictive models.

IV. RESULTS AND DISCUSSION

Interface: The interface consists of three screens, each providing a different perspective on the T20 World Cup tournament. Screen 1 focuses on the Group Stage of the T20 World Cup. It presents an overview of the group stage matches, displaying the results predicted by our model. Users can view the participating teams in each group, along with match fixtures, dates, and times. Predicted outcomes for each match are showcased based on our prediction model. Real-time updates on match results are provided, allowing users to stay informed about the latest developments. Additional statistics and insights about the teams' performances in the group stage may also be included to enrich the user experience.

Moving to Screen 2, attention shifts to the Super Eight Group stage. Here, users can explore the top-performing teams that have advanced from the group stage. Similar to the previous screen, match fixtures, dates, and times for the Super Eight matches are displayed. Predicted outcomes for each match are provided based on our model's predictions. Real-time updates on match results and the standings in the Super Eight group keep users engaged and informed. Comparative analysis of team performance between the group stage and the Super Eight stage may also be available for deeper insights.

Finally, Screen 3 encapsulates the Semi-Final and Winner stage of the tournament. It marks the culmination of the tournament journey, leading up to the crowning of the ultimate winner. Details of the semi-final matches, including participating teams and predicted outcomes, are presented. Real-time updates on semi-final match results are provided, along with anticipation-building elements for the final match. Once the final match concludes, the interface announces the tournament winner, showcasing the victorious team and any additional accolades or awards they may have achieved.

Overall, this interface offers users a comprehensive and engaging experience, providing insights, predictions, and real-time updates throughout the T20 World Cup tournament journey, from the group stage to the final match.

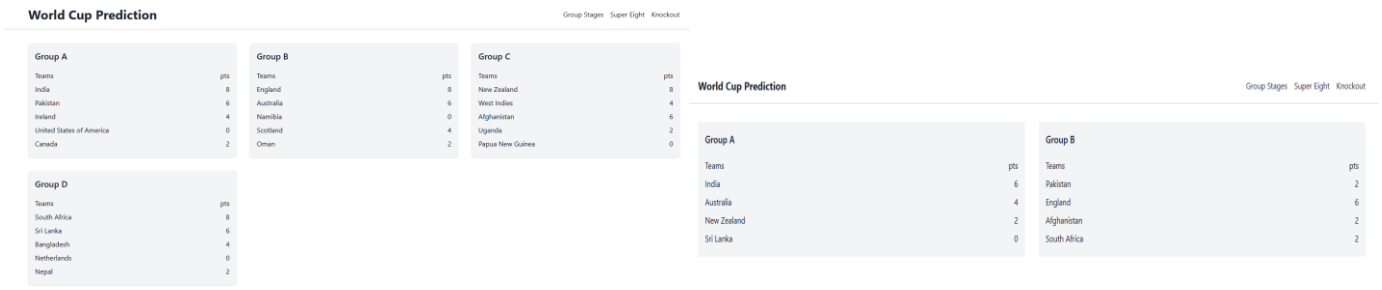

Fig 4.1 Prediction Interface

Accuracy: With an achieved accuracy of 0.6618705035971223, our model demonstrates a respectable level of performance in its predictions. This accuracy metric signifies the proportion of correct predictions made by the model compared to the total number of predictions. While it may not reach perfection, this level of accuracy suggests that the model is making informed decisions and capturing patterns within the data reasonably well. However, it's essential to consider other evaluation metrics and domain-specific factors to comprehensively assess the model's effectiveness. Fine-tuning and further refinement of the model may help improve its accuracy and enhance its predictive capabilities in future iterations.

Confusion Matrix: The results of the predictive analysis reveal the RF classifier's effectiveness in capturing patterns and trends within the CWC T20 2024 dataset. The confusion matrix analysis highlights the model's ability to correctly predict match winners, identify underperforming teams, and anticipate tournament dynamics. Insights drawn from the analysis contribute to refining the RF classifier's parameters and enhancing its predictive accuracy for future cricket tournaments.

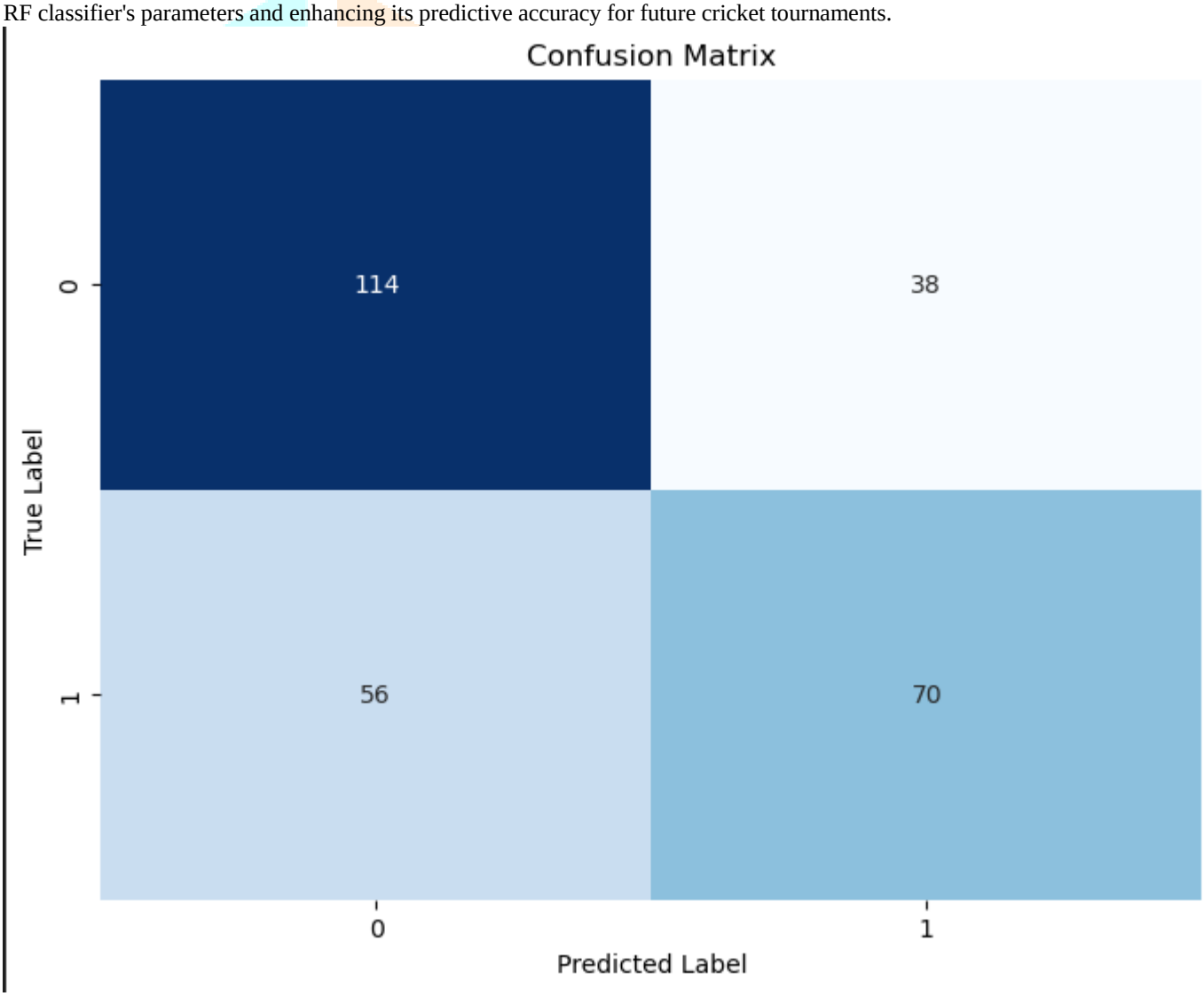

Fig 4.2 Confusion Matrix

.

REFERENCES

[1] S. Priya, A. K. Gupta, A. Dwivedi and A. Prabhakar, - Analysis and Winning Prediction in T20 Cricket using Machine Learning.

[2] A.P. Nirmala, B. Gogi, V. Asha, A. Naveen, A. Prasad and D.P. Reddy - Analysis and Predictions of Winning Indian Premier League match using Machine Learning Algorithms
[3] M. Sumathi, S. Prabu and M. Rajkamal - Cricket Players Performance Prediction and Evaluation using Machine Learning Algorithms
[4] M. Hatharasinghe and G. Poravi - Data Mining and Machine Learning in Cricket Match Outcome Prediction: Missing Links
[5] J. Kumar, R. Kumar and P. Kumar - Outcome Prediction of ODI Cricket Matches using Decision Trees and MLP Networks
[6] S. Priya, A. K. Gupta, A. Dwivedi and A. Prabhakar, "Analysis and Winning Prediction in T20 Cricket using Machine Learning," 2022 Second International Conference on Advances in Electrical, Computing, Communication and Sustainable Technologies (ICAECT), Bhilai, India, 2022, pp. 1-4
[7] H. Barot, A. Kothari, P. Bide, B. Ahir and R. Kankaria, "Analysis and Prediction for the Indian Premier League," 2020 International Conference for Emerging Technology (INCET), Belgaum, India, 2020.

.