Predicting BPL Match Winners: An Empirical Study Using Machine Learning Approach

Conference Paper · July 2023

DOI: 10.1109/ICCCNT56998.2023.10306698

2 authors:

Bornita Adhikari
Daffodil International University
1 PUBLICATION **0** CITATIONS
SEE PROFILE

Md. Sazzadur Ahamed
Daffodil International University
24 PUBLICATIONS **17** CITATIONS
SEE PROFILE

Predicting BPL Match Winners: An Empirical Study Using Machine Learning Approach

Bornita Adhikari
Department. of Computer Science and Engineering
Daffodil International University
Dhaka,Bangladesh
bornita15-14068@diu.edu.bd

Md. Sazzadur Ahamed
Department. of Computer Science and Engineering
Daffodil International University
Dhaka,Bangladesh
sazzad.cse@diu.edu.bd

***Abstract*— With the evolution of computer science, every company is implementing the newest technologies to survive in market with better decision-making capabilities, better communication and customer satisfaction. The only means of fulfilling all these criteria's is to perform data analysis that is more accurate and pure. In cricket, where no one can guess which team will win until the last ball of the last over, machine learning can help by predicting the results of the games. Match outcome prediction models have a lot of financial incentive because cricket is a multi-billion-dollar industry. The goal of this study is to identify the most accurate machine learning model that can accurately predict the winner given the data from the Bangladesh Premier League. For this analysis five ML models XGBoost, Gradient Boosting, KNN, Decision Tree, Random Forest has been tested for the purpose of model building despite that our proposed model is XGBoost. To get access to BPL dataset web scrapping has been done, the dataset contains 15 columns and 3239 values and 8 team was available in each season from 2018 to 2023. We use cutting-edge machine learning techniques based on the use of numerous models, feature selection, and data separation techniques. Finally, by structuring every line of action, the forecast accuracy is attained.**

Keywords*— *BPL, Cricket, Prediction, XGBoost, Visualization, Classification

I. INTRODUCTION

Cricket is a well-liked sport that is played and enjoyed by millions of people around the globe. It is especially popular in countries such as India, Pakistan, Australia, England, South Africa, Sri Lanka, Bangladesh, and the West Indies, where it is considered a national pastime. The reason for the popularity of cricket is it offers excitement and drama. The sport is known for its high-scoring matches, close finishes, and the individual brilliance of its players. The fast-paced and dynamic nature of Twenty20 cricket has made it particularly popular in recent years, as it offers a shorter and more action-packed version of the game. Nowadays Cricket is being benefited from the growth of technology and the media, which has made it more accessible to fans. Live streaming of matches, social media, and mobile apps have made it easier for fans to stay up-to-date with the latest news and scores.

A professional Twenty20 cricket league is called the Bangladesh Premier League (BPL) that was launched in 2012 and it operates on franchise-based business model. BPL has become a t-twenty blast investing big amount of money. There are five individual winning team in BPL history, with the Dhaka Dynamites winning the most titles (3). Teams select their players based on draft system that says players should be selected based on their performances in the previous season. Every team want to give their best and for this purpose ML prediction can play significant role by handling any uncertainty and can make an impact in predicting the winner of matches using present data in several ways. ML models can analyze the past data and predict how weather condition and a particular pitch or ground may affect the match's outcome. By analyzing the factors like team performance, team ranking, head-to-head records and the recent form of players, probability of winning the game at particular venue, score impact on field/bat first after winning the toss machine learning models can predict which team is likely to win.

In our analysis, the performance of each model and future directions are discussed with the goal of predicting the outcome of BPL matches. With the growth of T20 leagues and technological advancements the ask for cricket winner prediction models is anticipated to expand in the upcoming years, as more fans and teams recognize the value of data-driven insights and the potential competitive advantages they can provide. Our experimental conclusion can help to optimize team strategies and increase their chances of winning. The prediction was convey using five machine learning classifiers XGBoost, Gradient Boosting(GB), k-nearest neighbors(KNN), Decision Tree(DT), Random Forest(RF) where all the models in our research has shown outstanding accuracy. According to investigational result, XGBoost indicates better prediction of 93%.We described the benefits of cricket outcome prediction modeling with brief introduction of this game. In section 2, some of the related works about cricket outcome prediction is shown. The section 3, represent the workflow of this analysis, Section 4 comes with the result formulation and conclusion section 5 of this work provides a description.

II. LITERATURE REVIEW

Cricket has gained a lot of attention as it has progressed among sports commentators. Cricket has been the subject of an increasing amount of research, but because this dataset is brand-new and private, it has not yet been utilized in any research articles. Vistro et al. [1] conducted a study which aimed to foresee the triumphant team in cricket matches utilizing machine learning and data analytics methods. The study incorporated a range of features related to team and player performance, venue, and other match-specific variables to train their models. The study's findings showed that the suggested method able to foresee the winner of cricket matches with an accuracy of more than 70%. The study accentuated the potential of machine learning and data analytics in predicting the winner of cricket matches. In the study of Awan et al. [2] The team scores were predicted using

a machine learning linear regression model, and the accuracy was assessed using the big data framework Spark ML., RMSE, MSE, MAE on ODI dataset and got the outstanding accuracy 95%. To predict the results of IPL matches, Pallavi Tekade er al. [3] used several ML models, including Decision Tree Regression, Random Forest Regression, Naive Bayes, and Logistic Regression. They briefly described the key factors, such as pitch conditions, temperature, humidity, and precipitation, that directly affect the match scores. To illustrate their methods, they used a number of diagrams, and they recorded a 90% accuracy rate. With 70% accuracy, Vignesh Vapors Sankaranarayanan et al. [4] created a model to forecast the results of 125 ODI cricket matches. They also discuss how well the model performed. The difference between projected and actual total home runs, the overall score error distribution, and an examination of betting market sensitivity are all being shown using various data mining approaches. Mittal and others [5] Kumar conducted a thorough analysis and comparison of some of the most widely used machine learning algorithms, Naive Bayes, Logistic Regression Neural Networks, and Random Forest, among others, which virtually exactly predict the outcome of a match. Based on player performance and previous form, as well as how many runs each batsman is expected to score and how many wickets each bowler is expected to take, Kalpdrum Passi and Niravkumar Pandey [7] attempt to forecast the outcome of the game. This document provides classification reports of algorithms and a detailed computation of batting and bowling attributes. In their investigation, the Random Forest classifier produced the highest accurate results. Kamble, R. R. [8] developed a system combining 2 models which can predict the score a team and prediction of the win percentage of both teams. This is an effective prediction system for cricket developed with 5 years of ODI data based on Linear Regression, Naïve Bayes algorithm in their analysis, Passi, Kalpdrum, and Niravkumar Pandey [9] used naive bayes, random forest, multiclass SVM, and decision tree to create prediction models that could forecast how many runs a player would score and how many wickets a bowler would take by analyzing attributes from ODI cricket matches. The application of machine learning to forecast the results of English county twenty over cricket matches is explored in the study by Kampakis and Thomas (2015).The goal of this work by Lamsal and Choudhary [11] was to use machine learning to predict the results of cricket matches in the Indian Premier League.They put out a revolutionary strategy that blended various machine learning algorithms and included both simple and intricate features concerning team and player performance, weather patterns, and other match-specific aspects. The suggested model outperformed the benchmark model and had a prediction accuracy for matches of 64.1%. The authors came to the conclusion that their strategy was an effective way to forecast cricket match results in the IPL and that it could also be used for other sports. Overall, this study adds to the body of knowledge on picking winners and losers in cricket matches.

Using machine learning methods, Srikantaiah, K. C. et al. [12] sought to forecast the results of Indian Premier League (IPL) matches. Authors use a variety of features to train their models, including batting and bowling statistics, location, and team performance. The findings demonstrated that ensemble approaches, such as Random Forest and Gradient Boosting, outperformed individual models, yielding match outcome predictions that were over 60% accurate. In Wickramasinghe's study [13], machine learning techniques were used to categorize all-rounders in One Day International (ODI) cricket. The authors chose the fielding, bowling, and batting statistics-related variables to train the algorithms. The suggested method outperformed conventional statistical methods in categorizing all-rounders with high accuracy. The study helps estimate cricket player performance using machine learning techniques. Overall, the study shows how machine learning may be used to categorize and forecast cricket player performance. To find the ideal cricket match winning strategy, Srivastava et al. [17] suggested a hybrid machine learning-clustering-association rule architecture. In order to train their models, the authors took advantage of a variety of team composition, batting and bowling performance, and pitch characteristics factors. The best players and tactics for winning a match could be found using the suggested framework. The study adds to the body of knowledge regarding the application of machine learning to cricket strategy. In order to forecast player performance and team recommendations in the sport of cricket, Biswas et al. [16] conducted a poll. The authors analyzed a number of research that employed machine learning strategies to forecast cricket player performance, team makeup, and game outcomes. The poll brought attention to how machine learning could enhance cricket team selection, training, and strategic decision-making.

Using machine learning techniques, Ahmed et al.'s [14] analysis of the Pakistan cricket team's performance. To forecast game results and pinpoint variables that affect a team's success or failure, the authors examined several performance indicators and player information. According to the study, the performance of the Pakistani team can be influenced by a number of variables, including player form, pitch conditions, and team composition. The study's conclusions show the potential of machine learning for cricket analytics and have practical ramifications for raising the performance of the Pakistan side. This study adds to the body of knowledge on cricket performance analysis and strategic decision-making using machine learning. n their 2017 study, Mustafa et al. sought to determine whether machine learning techniques could accurately forecast cricket match results based on user sentiment on social media. To find the most effective method for prediction, the authors compared various algorithms, including Random Forests (RF), K-Nearest Neighbors (KNN), and Support Vector Machines (SVM). With a 67% accuracy rate, the data demonstrated that the RF algorithm performed better than the others. The study emphasizes the potential of using online social network crowdsourcing to forecast sporting events. The investigation made by Mustafa et al. [6] (2017) aimed to investigate the effectiveness of machine learning methods in predicting cricket match outcomes employing social network user evaluations. The authors compared different algorithms such as Random Forests (RF), K-Nearest Neighbors (KNN), and Support Vector Machines (SVM) to determine the best approach for prediction. The findings revealed that the RF algorithm outperformed the others with an accuracy rate of 67%. The study highlights the potential of using crowd opinions on social networks for predicting sports outcomes.

III. METHODOLOGY

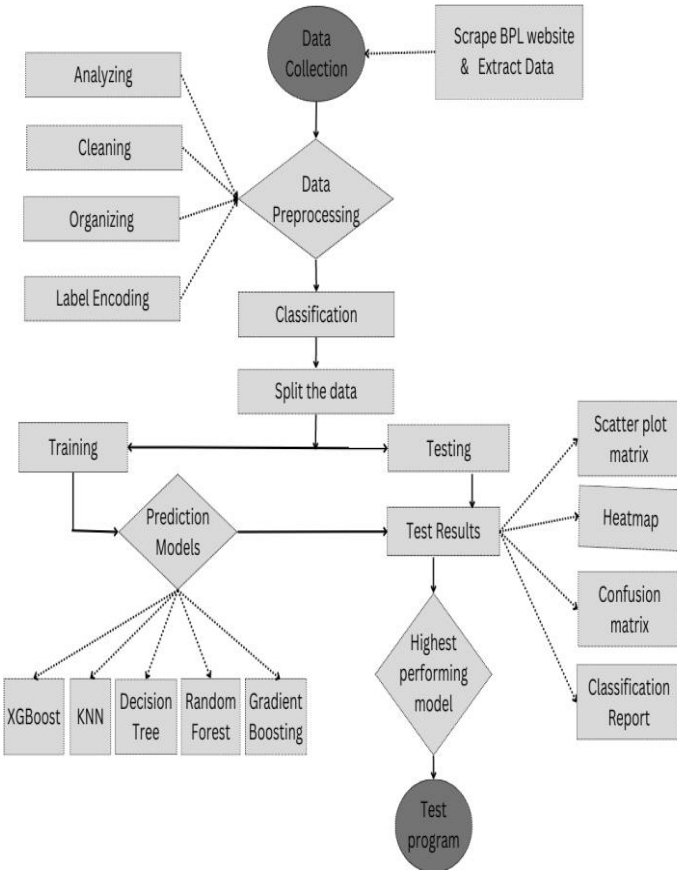

Fig. 1. Workflow of Methodology

The workflow of this research are briefly elaborated in Figure 1. The prime decision of this analysis was select Bangladesh Premier League as our study work. First and foremost, data extraction was done. Before classification a small group of data preprocessing was needed. After testing several models, the proposed model was found, that got highest accuracy.

A. Data Description

To conduct the analysis, we sourced data from cricbuzz.com that pertains to matches played in the BPL Twenty20 cricket league from 2018 to 2023. Each season comprises 46 matches and the data includes various attributes such as the city, date, teams, toss winner, toss decision, winner team, win-by-runs, result, win-by-wickets, and venue. To ensure high accuracy and better outcomes when applying prediction models, a significant amount of data was extracted, resulting in a dataset that contains 15 columns and 3239 entries.

B. Data Pre-Processing

Preparing data for analysis involves several crucial steps, collectively known as data pre-processing. These include cleaning, transforming, and organizing raw data to make it suitable for analysis. This step is crucial because the accuracy and effectiveness of any analysis or modeling depend heavily on the quality of the data. The various steps involved in pre-processing are discussed in detail below.

- *Removing Unwanted Columns*

We eliminated null values and unnecessary columns like ID, Season, and Man of the Match from this analysis in order to focus on the columns that our forecast is founded on and rely upon.

- *Tackle Data Volume*

To tackle the low volume of data causing lower performance of algorithms, the data had to be volume up by duplicating data rows.

- *Label Encoding*

Label encoding is the process of giving several levels or categories of a variable or attribute a numerical value or code. The pre-processing step of categorical data representation is crucial for statistical modeling and data analysis. Data for ML models must only be in numerical format. Therefore, label encoding has been used to transform every entity into a numerical representation. Before label encoding, Table 1 displays all the properties and values of our dataset. The idea of label encoding is displayed in Table 2.

TABLE I. BEFORE LABEL ENCODING

City	date	team1	Team2	toss-winner	Toss-decision	result	winner	Win-by-runs	Win-by-wickets	value
Chattogram	20/12/2019	Chattogram Challengers	Comilla Warriors	Comilla Warriors	Field	Normal	Chattogram Challengers	16	0	Zahur Ahmed Chowdhury Stadium
Dhaka	11/1/2020	Dhaka Platoon	Khulna Tighers	Khulna Tighers	Field	Normal	Khulna Tighers	0	8	Sher-e-Bangla National Cricket Stadium
Dhaka	11/1/2020	Dhaka Platoon	Khulna Tighers	Khulna Tighers	Field	Normal	Dhaka Platoon	0	4	Sher-e-Bangla National Cricket Stadium
Dhaka	4/2/2018	Chattogram Challengers	Dhaka Platoon	Chattogram Challengers	Bat	Normal	Dhaka Platoon	0	7	Sher-e-Bangla National Cricket Stadium
Dhaka	4/2/2023	Chattogram Challengers	Comilla Warriors	Chattogram Challengers	Bat	Normal	Comilla Warriors	0	6	Sher-e-Bangla National Cricket Stadium

TABLE II. AFTER LABEL ENCODING

City	date	team1	Team2	toss-winner	Toss-decision	result	winner	Win-by-runs	Win-by-wickets	value
0	41	0	2	1	1	1	0	16	0	4
2	8	2	5	4	1	1	6	0	8	1
2	79	2	5	4	1	1	2	0	4	2
2	75	0	3	0	0	1	2	0	7	2
2	76	0	2	0	0	1	1	0	6	2

C. *Visualize dataset*

- scatter plot matrix

In machine learning research, data visualization plays a crucial role in comprehending the connections between variables, recognizing patterns, and gaining insights into model performance. One such visual representation is the scatter plot matrix of features in xtrain, which displays the correlations between pairs of features (or variables) in the training dataset. To predict the winner of BPL matches, this research has selected five columns - city, date, team1, team2, and toss winner - to plot on the two axes, X and Y in a grid of scatter plots. Each plot within the grid demonstrates the relationship between two features, while the diagonal of the grid showcases the distribution of each feature. The correlations observed in the dataset is shown in Figure 2

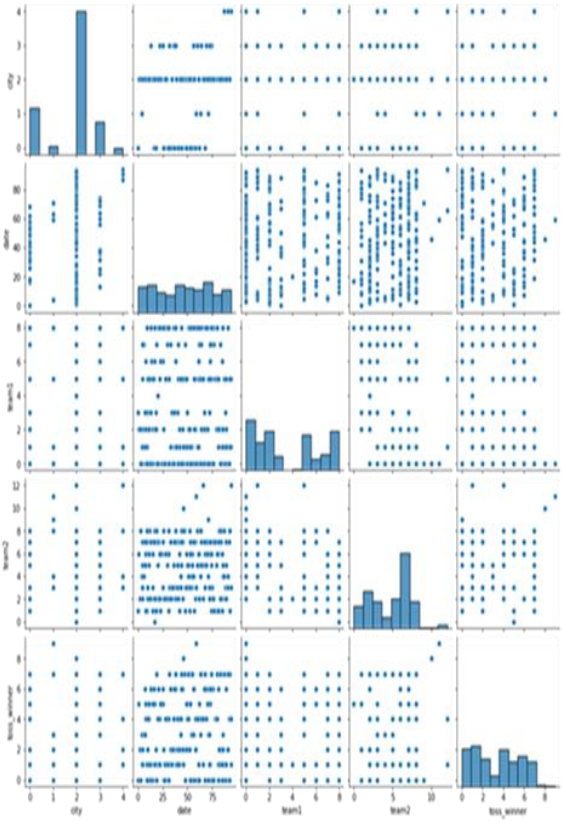

Fig. 2. Scatter plot matrix

- visualize the correlation matrix Heatmap

To view the correlation matrix Heatmap, one needs to visualize a table that displays the correlation coefficients between pairs of variables in a dataset. These coefficients are statistical measures that indicate the degree of correlation between two elements. Correlation coefficient values vary from -1 to 1, where -1 denotes a fully negative correlation, 1 denotes a fully positive correlation, and 0 denotes no association. The plot that results from this visualization is a matrix that is color-coded, with red hues representing positive relationships and blue tones representing negative correlations. This display is shown in Figure3.

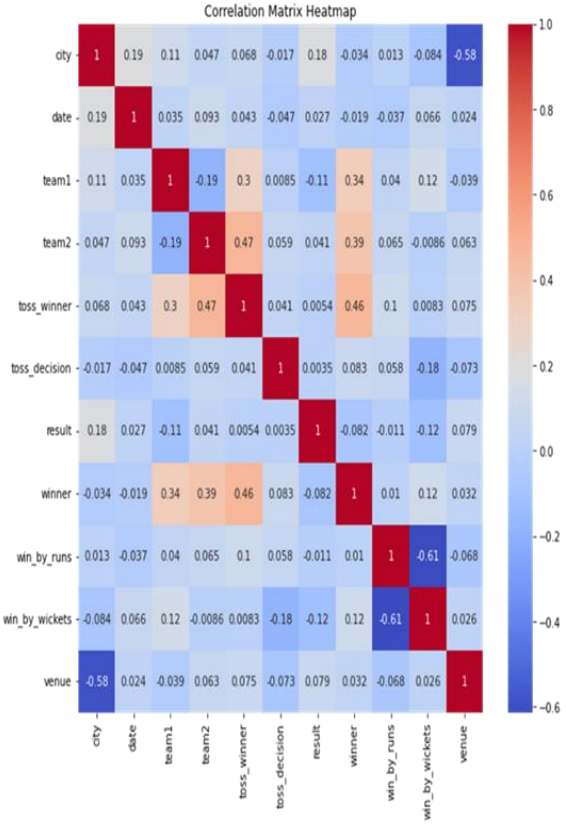

Fig. 3. Correlation matrix Heatmap

D. *Classification models training*

To prepare the models and preprocess the data, Google Colaboratory's cloud GPU, along with essential libraries such as pandas, numpy, matplotlib, and seaborn, have been employed. The data set includes a substantial number of entities, and the relevant variables that are likely to aid in predicting the target variable (winner) have been selected based on the data description. Following the selection of algorithms, the input data with corresponding target values was fed into the model training. To increase the volume of data, the data was duplicated by a factor of two, and during model training, the data was randomly divided into testing and training sections. The training component used 80% of the data, and the testing portion used the remaining 20%. Brief descriptions of the tested models are provided below.

- KNN

The K-Nearest Neighbors (KNN) is a straightforward machine learning technique that is utilized for both regression and classification assignments. The algorithm functions by finding the K closest data points to a provided input data point within the training set and then utilizing their values or labels to predict an outcome.

- Decision Tree

The Decision Tree algorithm functions by dividing data into smaller, similar subsets based on its defining features. To make predictions, the input data is moved through the tree from the top node to the leaf node. The algorithm determines which branch to follow at each internal node by evaluating the input feature's value. This process continues until the leaf node is reached, which provides a forecast for the input data. No information has been left out in the paraphrased text.

- Random Forest

One sort of ensemble learning that uses the Random Forest method utilizes numerous decision trees to predict outcomes. In order to make a prediction using this algorithm, the input data is processed through each decision tree from the root node to a leaf node. Once the data has reached the leaf node of each tree, the algorithm produces a prediction based on either the majority class or the average prediction value of all the decision trees within the Random Forest.

- Gradient Boosting

Gradient Boosting is a technique that involves combining several weak learners to form a powerful one. To make predictions, the input data is first fed into the weak learner, and the errors in the initial prediction are determined. To improve the accuracy of the prediction, a new weak learner is then trained to correct the previous learner. The final prediction is obtained by adding up the predictions from all the weak learners.

- XGBoost

XGBoost is a well-known machine learning model that is widely used for regression, classification, and ranking tasks. It is an ensemble model that consists of multiple decision trees and leverages the errors of previous trees to enhance its predictions. During the prediction process, the input data is processed through various decision trees, and the scores generated by each tree are merged to produce a final prediction. Finally, a non-linear function like the sigmoid or softmax function is applied to convert the output into a probability or class label. No information has been omitted in the paraphrased text.

E. *Making predictions*

After testing several model / machine learning algorithms, it is seen that the models are getting trained conveniently. But the testing capability is not performed yet. To find out how the algorithms are performing on each test portions of the dataset, we had to evaluate the model's performance through a step of prediction performed from the trained models separately and the testing results indicate XGBoost model to be the highest performing one.

IV. RESULT AND DISCUSSION

The accuracy table is helpful for assessing the effectiveness of models and pinpointing potential improvement areas. By emphasizing particular sorts of model flaws, the presentation of the accuracy table can help reveal areas that need work. The proportion of correctly classified observations in the dataset after testing several models is shown in Table 3 and Figure 4 is the proper visualization of the result.

TABLE III. MODEL ACCURACY

Models	Accuracy
XGBoost	93%
Gradient Boosting	92%
Random Forest	86%
Decision Tree	86%
KNN	79%

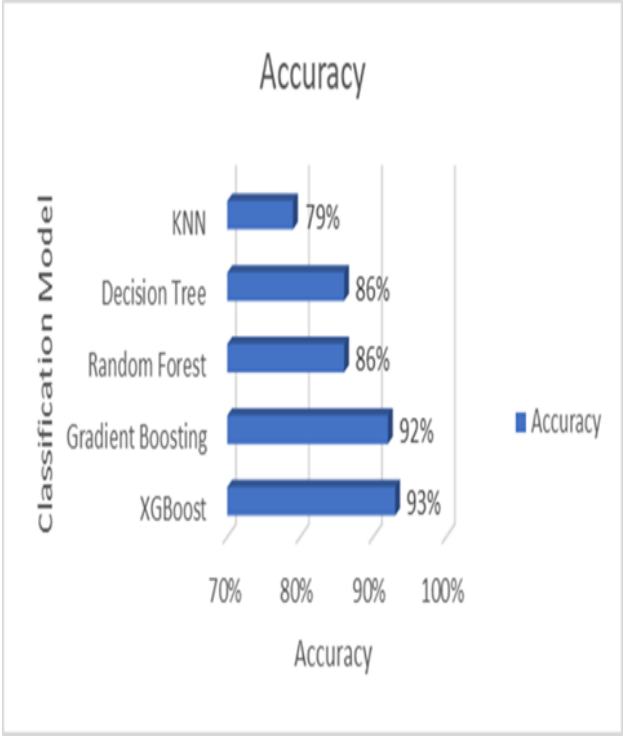

Fig. 4. Accuracy compares

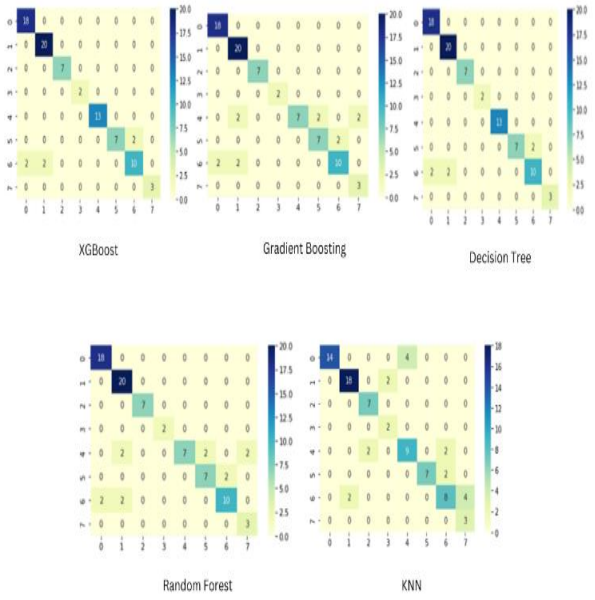

Fig. 5. Confusion Matrix

A. *Confusion Matrix:*

The counts of true positives, true negatives, false positives, and false negatives for each class in the dataset are shown in the confusion matrix. In case the model is producing a lot of false positives, we can attempt to enhance its performance by modifying the decision threshold or experimenting with different features. Figure 5 illustrates the confusion matrix of the models that were tested.

B. *Model Implementation –XGBoost Classifiers*

Extreme Gradient Boosting (XGBoost) is a highly effective machine learning algorithm known for its remarkable predictive accuracy. It is specially designed to be computationally efficient and capable of handling large datasets. The algorithm incorporates L1 and L2 regularization, as well as dropout regularization, to prevent overfitting and improve the model's generalization performance. Additionally, XGBoost generates feature importance scores, which help in identifying the most significant features in the data. This feature can be utilized for feature selection, feature engineering, and to gain insights into various machine learning tasks such as regression, classification, and ranking.

The flexibility and customization of the XGBoost algorithm is due to its support for various loss functions and evaluation metrics. To implement this algorithm, we utilized the BPL dataset from scikit-learn and split it into training and testing sets. Our XGBoost model was defined using the XGBClassifier class and was trained using the training data. Predictions were then made on the test data using the predict method. The evaluation of our model's performance was done using accuracy as the primary metric.

In machine learning, performance metrics are used to assess the effectiveness of different models. Accuracy is a useful metric that measures the proportion of correctly predicted instances in the dataset. Our XGBoost Classifier achieved an accuracy of 94.25%, indicating a satisfactory performance of our model.

C. *Classification report of XGBoost*

Table 4 presents a report that summarizes the performance of various classes or categories in the XGBoost dataset. The report includes key metrics such as Precision, which represents the proportion of predicted positives that are true positives, and Recall, which refers to the proportion of actual positives that are correctly identified. The F1 score is a combined measure of precision and recall, while Support indicates the number of observations in each class.

TABLE IV. AFTER LABEL ENCODING

Class	Precision	Recall	F1-Score	Support
Fortune Barishal	0.90	1.00	0.95	18
Khulna Tighers	0.91	1.00	0.95	20
Comilla Warriors	1.00	1.00	1.00	7
Chattogram Challengers	1.00	1.00	1.00	2
Dhaka Platoon	1.00	1.00	1.00	13
Sylhet Thunder	1.00	0.78	0.88	9
Rajshahi Royals	0.83	0.71	0.77	14
Rangpur Riders	1.00	1.00	1.00	3

V. CONCLUSION AND FUTURE WORK

Predicting the match winner of BPL data from season 2018 to 2023 was the primary goal of this study. We compared the performance of 5 different machine learning models to see which one best fits the dataset and The fields of data science that have merged include visualizations, pre-processing, implementing machine learning models, feature selection for winner prediction and analysis. By Level encoding into numerical format, removing the missing values from the dataset the preprocessing has been done By combining the target variable with the properties of the data, the best features were chosen. We have found outstanding results by performing. models for machine learning with specific attributes. Our goal was to predict the winner with highest accuracy, we achieved the highest accuracy on XGBoost and predicted the winner with 93% accuracy. In this process of predicting the winner KNN gave us the lowest accuracy which is 79%.

Because of different ways of data preprocessing, hyper parameter tuning, Sensitivity to initial conditions a single algorithm can produce different result. Therefore, we are currently attempting to decrease the prediction inaccuracy. Furthermore, we intend to predict fall of wickets, finding about the reasons behind most preferred decision of BPL after toss Field or Ball. Eventually, we aim to host a website where all this analysis will be available for audience.

REFERENCES

[1] Vistro, Daniel Mago, Faizan Rasheed, and Leo Gertrude David. "The cricket winner prediction with application of machine learning and data analytics." International Journal of Scientific & Technology Research 8, no. 09 (2019).

[2] 2. Awan, Mazhar Javed, Syed Arbaz Haider Gilani, Hamza Ramzan, Haitham Nobanee, Awais Yasin, Azlan Mohd Zain, and Rabia Javed. "Cricket match analytics using the big data approach." Electronics 10, no. 19 (2021): 2350.

[3] 3. Tekade, Pallavi, Kunal Markad, Aniket Amage, and Bhagwat Natekar. "Cricket match outcome prediction using machine learning." International journal 5, no. 7 (2020).

[4] 4. Sankaranarayanan, Vignesh Veppur, Junaed Sattar, and Laks VS Lakshmanan. "Auto-play: A data mining approach to ODI cricket simulation and prediction." In Proceedings of the 2014 SIAM international conference on data mining, pp. 1064-1072. Society for Industrial and Applied Mathematics, 2014.

[5] 5. Mittal, Harsh, Deepak Rikhari, Jitendra Kumar, and Ashutosh Kumar Singh. "A study on machine learning approaches for player performance and match results prediction." arXiv preprint arXiv:2108.10125 (2021).

[6] 6. Mustafa, Raza Ul, M. Saqib Nawaz, M. Ikram Ullah Lali, Tehseen Zia, and Waqar Mehmood. "Predicting the cricket match outcome using crowd opinions on social networks: A comparative study of machine learning methods." Malaysian Journal of Computer Science 30, no. 1 (2017): 63-76.

[7] 7. Passi, Kalpdrum, and Niravkumar Pandey. "Increased prediction accuracy in the game of cricket using machine learning." arXiv preprint arXiv:1804.04226 (2018).

[8] 8. Kamble, R. R. "Cricket score prediction using machine learning." Turkish Journal of Computer and Mathematics Education (TURCOMAT) 12, no. 1S (2021): 23-28.

[9] 9. Passi, Kalpdrum, and Niravkumar Pandey. "Predicting players' performance in one day international cricket matches using machine learning." Computer Science & Information Technology (CS & IT) (2017).

[10] 10. Kampakis, Stylianos, and William Thomas. "Using machine learning to predict the outcome of english county twenty over cricket matches." arXiv preprint arXiv:1511.05837 (2015).

[11] 11. Lamsal, Rabindra, and Ayesha Choudhary. "Predicting outcome of Indian premier league (IPL) matches using machine learning." arXiv preprint arXiv:1809.09813 (2018).

[12] 12. Srikantaiah, K. C., Aryan Khetan, Baibhav Kumar, Divy Tolani, and Harshal Patel. "Prediction of IPL match outcome using machine learning techniques." In 3rd International Conference on Integrated Intelligent Computing Communication & Security (ICIIC 2021), pp. 399-406. Atlantis Press, 2021.

[13] 13. Wickramasinghe, Indika. "Classification of all-rounders in the game of ODI cricket: Machine learning approach." Athens Journal of Sports 7, no. 1 (2020): 21-34.

[14] 14. Dhonge, Nikhil, Shraddha Dhole, Nikita Wavre, Mandar Pardakhe, and Amit Nagarale. "IPL cricket score and winning prediction using machine learning techniques." Int Res J Modernization Eng Technol Sci 3 (2021): 1723-1730.

[15] 15. Ahmed, Waqar, Mahwish Amjad, K. Junejo, Tariq Mahmood, and A. Khan. "Is the performance of a cricket team really unpredictable? a case study on pakistan team using machine learning." Indian Journal of Science and Technology 13, no. 34 (2020): 3586-3599.

[16] 16. Biswas, Milon, Tajim Md Niamat Ullah Akhund, Md Kawsher Mahbub, Sikder Md Saiful Islam, Sadia Sorna, and M. Shamim Kaiser. "A survey on predicting player's performance and team recommendation in game of cricket using machine learning." In Information and Communication Technology for Competitive Strategies (ICTCS 2020) ICT: Applications and Social Interfaces, pp. 223-230. Springer Singapore, 2022.

[17] 17. Srivastava, Praveen Ranjan, Prajwal Eachempati, Ajay Kumar, Ashish Kumar Jha, and Lalitha Dhamotharan. "Best strategy to win a match: an analytical approach using hybrid machine learning-clustering-association rule framework." Annals of Operations Research (2022): 1-43.