Performance Analysis of Classification Algorithms for Outcome Prediction of T20 Cricket Tournament Matches

Md. Aktaruzzaman Pramanik
Dept. of CSE
Daffodil International University
Dhaka, Bangladesh
a.pramanikk@gmail.com

Md. Mahmudul Hasan Suzan
Dept. of CIS
Daffodil International University
Dhaka, Bangladesh
mahmudul16-410@diu.edu.bd

Al Amin Biswas
Dept. of CSE
Daffodil International University
Dhaka, Bangladesh
alaminbiswas.cse@gmail.com

Mohammad Zahidur Rahman
Dept. of CSE
Jahangirnagar University
Dhaka, Bangladesh
rmzahid@juniv.edu

A. Kalaiarasi
Dept. of EEE
Sri Shakthi institute of engineering and technology
Tamil Nadu, India
kalaiarasi.a1406@gmail.com

2022 International Conference on Computer Communication and Informatics (ICCCI) | 978-1-6654-8035-2/22/$31.00 ©2022 IEEE | DOI: 10.1109/ICCCI54379.2022.9740867

***Abstract*—Cricket is the most popular sport in south Asian countries and the second most popular sport globally. And T20 cricket is the most attractive version of this game. Businesses have grown enormously based on cricketing sports events from the last decade. A large number of research have been done to predict the winner of cricket matches or to analyze game statistics. These studies are helping investors and franchisees to decide on which team they can invest in to gain more profit. Also, coaches, sports analysts, and technicians get game facts and ideas about other teams, which help them make decisions and change plans accordingly. In this study, we have shown a comparative analysis of different non-ensemble and ensemble machine learning classifiers. We have collected data of all seasons of the Bangladesh Premier League(BPL) T20 tournament. We have prepared two datasets: one contains only pre-match information, and the second includes post-match information. We have applied five base classifiers and five ensemble classifiers for the analysis. We found that K-Nearest Neighbor(KNN) performed best compared to other algorithms while predicting match results before starting the game. The Gradient Boosting classifier seemed more robust than the different classifiers for predicting match outcomes considering all features.**

***Index Terms*—machine learning, prediction, cricket, t20, BPL**

I. Introduction

In the South Asian region, cricket is viewed and played by the maximum number of people compared to other sports [1]. After football, it is considered as the second most liked sport in the world [2]. Due to the vast popularity and the success of South Asian countries in the game of cricket, the economic impact of this game also reached its peak. Even in India, the study on this game has been increased significantly. As a result, it got more academic attention than ever, and research culture is expected to get boosted soon [3]. Since South Asian countries got massive success in this game, a growing number of stakeholders are getting involved in cricket-centric sports events. There are three formats in the game of cricket. The first one is test cricket which is played through 5 days in a row; the second one is called One Day international cricket, in which each side bats for 50 overs(1 over consists of 6 balls) and bowls for 50 overs. The third and most recent version of world cricket is T20 cricket, where each side got 20 overs for bowling and 20 overs for batting; this is the most exciting and popular version of the game of cricket. Because of the increasing popularity of T20 cricket, India, Pakistan, Bangladesh, and Sri Lanka are arranging their own domestic T20 cricket tournaments for the last few years. India is playing a leading role in this case. They started their domestic franchise-based T20 league in 2008. Bangladesh also started its domestic T20 league named BPL T20 in 2012. During these tournaments, the sports-loving people feel festive, and they enjoy every game either physically coming to the stadium or on TV and live streams. As a result, the business has grown significantly based on these tournaments in South Asian countries and Bangladesh.

A significant amount of analysis and research is going on how this game can be made more exciting and gain more profit. Most studies have been conducted on historical game data or data collected during the game. The researchers have considered either post-match features [4]–[8] for analysis game facts and predicting winning team or using pre-match features [9]–[12]. We have tried to contribute to this case. We have collected all the match data of all BPL seasons and performed a thorough machine learning-based analysis on pre-match and post-match features. The main goal of this study is to help franchisees on which team they should invest in making it economically beneficial by using the results from the historical data of BPL, which includes post-match features. Also, it will help coaches, technicians, and sportspersons during the game because predicting match results before starting the game can help them make changes in their strategies during and before the game. Our main contributions include-

- We have collected the historical data of all the games played in BPL T20 from 2012 to 2019, containing 590 entries and 20 attributes.
- We experimented with machine learning-based classifiers with two different approaches: predicting the match result before starting the game by considering only pre-match features and predicting the match result based on all the match information, where the dataset included post-match features.
- We have shown the comparative analysis of individual and ensemble machine learning classifiers and demonstrated the efficacy.

The remainder of the paper is divided into the following sections: Section II is an overview of the available relevant literature. Our proposed methodology, including data collection and pre-processing, is described in Section III. Section IV demonstrates the brief descriptions of the applied classification algorithms. Section V explains the findings of the experiments as well as a full analysis. Finally, section VI finishes the paper by laying out recommendations for future.

II. Literature Review

Machine Learning-based classifiers were used for outcome prediction of Indian Premier League(IPL) T20 matches, and many researchers [4]–[6] included post-match features while Tripathi et al. [9] used only pre-match features for the classification task.

Kapadia et al. [4] used data of 10 IPL seasons(2008-2017). They applied statistical and probabilistic classification algorithms; though post-match features were included in their dataset, they found the highest classification accuracy from KNN of 62.00%. While Lamsal and Choudhary [5] got some improved performance(71.66%) using Multilayer Perceptron compared to other machine learning algorithms. Tripathi et al. [9] excluded all the post-match features from their dataset; they included the features till the toss happened and the player's historical career information, team strength, team ranks, winning rate, current points of the team, consistency level of the team, etc. They applied ensemble classifiers like Random Forest, XGBoost, AdaBoost, and ExTraTreesClassifier, and other statistical and probabilistic classification algorithms. They achieved the highest 60.043% prediction accuracy using Random Forest classifier. Mahmood et al. [13] collected Pakistan Super League(PSL) data of 4 seasons containing 115 entries only. The dataset includes post-match features like batting and bowling performance of both teams, scores of both teams. Their proposed model achieved 82.00% classification accuracy. Where Singh et al. [14] collected historical data of international T20 games from ESPN and Cricbuzz website and predicted the winning team of the upcoming World Cup T20 tournament. They applied several ensembles and non-ensemble classifiers for the prediction.

A significant amount of researches have been conducted to predict the result of One Day International(ODI) games. Some works [10]–[12] have only used pre-match information for the prediction task. On the other hand, some researchers [7], [8] included post-match features for their respective works where Weeraddana and Premaratne [15] used features that are collected when the match was in progress.

Pathak and Wadhwa [10] used only pre-match features and applied Naive Bayes, SVM, and Random Forest Classifier for predicting ODI match outcome. Naik et al. [11] also used only pre-match features. They have analyzed only for one game considering the player's previous performance and playing order, which may not be feasible for a larger scale. Kumar et al. [12] used Multilayer Perceptron Network(MLP) on the dataset containing only pre-match features, and the model performance was 57.4%.

Rahman et al. [16] collected data of matches played between Bangladesh and other teams from 2005 to 2017. They applied machine learning algorithms for predicting match results. They expected the game outcome before starting and after the end of the first innings. Using the SVM classifier, the got prediction accuracy of 63.63%. Mahtab et al. [17] used Bengali dataset ABSA containing five attributes and 1601 entries for their sentiment analysis using Support Vector Machine(SVM). The dataset comprises public opinions about cricket, which were available on social media platforms and online news portals. Their classification accuracy was 64.00%.

III. Methodology

Figure 1 shows every step of our proposed system for the prediction task. The next subsections go over each stage in further depth.

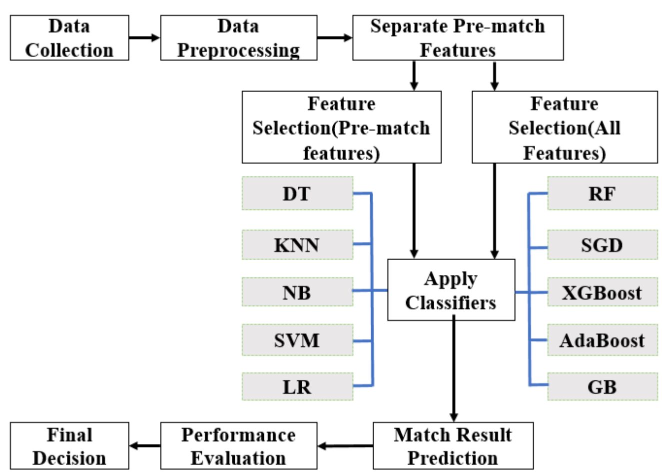

Fig. 1. Proposed system for predicting BPL match result

A. Data Collection

Seven Bangladesh Premier League(BPL) T20 seasons have been arranged from season 2011-12 to 2019-20(Season 2013-14 and 2014-15 were not held), and 8 teams have participated in it. We have manually collected data of all 7 BPL seasons from ESPNcricinfo[1] and Timeanddate[2]website. The team statistics found from our dataset are shown in table I.

The dataset contains 590 entries and 20 attributes. Among 20 features, 16 are pre-match features that can be collected at least 15 minutes before starting the game. Pre-match features are: team 1, team 2, day or night match, match type, matchday is a holiday or not, nationality of captain, no. of batters, no. of bowlers, no. of allrounders, no. of batters of opponent team, no. of bowlers of opponent team, no. of allrounders of opponent team, toss win, toss decision, ground, weather. And four post-match are batting score of team 1, player of the match team, player type of player of the match, match result. Description of attributes and their frequency distribution is presented in table II.

TABLE I
TEAM STATISTICS OF PARTICIPATING TEAMS

Team name	No. of games played	No. of wins	No. of losses	Win percentage
Dhaka	92	55	37	59.78%
Barisal	49	24	25	48.98%
Chittagong	86	40	46	46.51%
Cumilla	64	38	26	59.38%
Khulna	75	34	41	45.33%
Rajshahi	78	40	38	51.28%
Rangpur	77	40	37	51.95%
Sylhet	69	24	45	34.78%
Total	**590**	**295**	**295**	

B. Data Preprocessing

Most of the attribute values are of string data type. So firstly converted the string values into numeric values. We have mapped the batting score values into four categories and assigned numeric values for each range, such as 0 for the scores below 101, 1 for the scores between 101 to 150, 2 for the 151 to 200, and 3 for more than 200. Then we have replicated the whole dataset, and in the second one, we just kept pre-match features in it, and three post-match features were discarded from it except the target attribute(Match Result). So, the first dataset contained all 20 features, and the second dataset contained 17 pre-match features(including the target attribute).

To minimize the number of features, we omitted heavily correlated features from both the datasets. After reducing correlated features, the number of features remained in the first and second datasets, respectively 17 and 14. Stratified k-fold cross-validation ensures the same ratio of dependent and independent variables for each fold. So, we applied five-fold stratified cross-validation in both datasets while training with our chosen algorithms. We considered the five folds' mean accuracy and other performance metric values.

[1]https://stats.espncricinfo.com/ci/engine/records/team/series$_results.html?id=159;type=trophy$

[2]https://www.timeanddate.com/

IV. CLASSIFICATION ALGORITHMS

Classification is the process of identifying a discrete category of a query observation by training a set of data points [18]. Classification algorithms attempt to predict the target class with the highest legibility. These algorithms find out the connection between input attributes and the target attribute and build the models [19].

$$y = f(x, \theta), y \in Z \tag{1}$$

In equation 1, x is a feature vector which is the query observation, y is the predicted category for the query observation, the trained classification function is represented by f(.), θ is the set of parameters of the classification function, and Z is the set of class labels of the target attribute.

A. Base Classifiers

We have applied five base classification algorithms to our dataset, briefly described in the following subsections.

1) Decision Tree (DT): It is a vastly used tree-based supervised classification technique. It is a flowchart-like structure where attributes, the goal of attribute test, and the label of the class are represented by internal nodes, branches of the tree, and leaf nodes, respectively [20]. Branches of the tree are split repeatedly until its reaches the last level. The splits represent tests on data attributes, and through this process, the classification is achieved.

2) K-Nearest Neighbor (KNN): KNN uses a simple concept for classification, but it could be effective in many cases [21]. While classifying a dataset, it considers k nearest data points of the query data point and applies majority voting into those selected neighbors to decide the target class label. It may or may not consider distance-based weighting.

3) Naive Bayes (NB): It is a supervised probabilistic classification technique. This method determines the optimum mapping between a new data point and a collection of classifications for a given problem domain by performing some mathematical manipulation for transforming joint probability into the multiplication of prior and conditional probability [22].

4) Support Vector Machine (SVM): SVM uses a surface for maximizing the margin between two or more classes of the training dataset[6]. A maximum margin hyperplane splits positive and negative training data points in our training dataset labeled as positive and negative (win or loss). A hyperplane is chosen, which separates the sample as rigorously as possible if no hyperplane exists that can split the positive and negative samples.

5) Logistic Regression (LR): This model has been used for classifying our dataset. LR works by estimating the probability of a particular event occurring and tries to find the relationship between attributes [23].

B. Ensemble Classifiers

Ensemble classifiers combine several homogeneous weak classifiers and apply several merging method [24] and often

TABLE II
DESCRIPTION OF THE DATASET AND FREQUENCY DISTRIBUTION

Attributes	Description	Frequency Distribution
Team 1	Query Team	Barisal(49), Chittagong(86), Cumilla(64), Dhaka(92), Khulna(75), Rajshahi(78), Rangpur(77), Sylhet (69)
Team 2	Opponent team	Barisal(49), Chittagong(86), Cumilla(64), Dhaka(92), Khulna(75), Rajshahi(78), Rangpur(77), Sylhet (69)
Day/Night	Match conduction time(day or night)	Day(276), Night(314)
Match Type	Group stage or knockout game	Group Stage(536), Knock Out(54)
Holiday Match	Public holiday or not	Yes(100), No(490)
Captain Nationality	Nationality of Team 1 Captain	BD(453), Others(137)
No. of Batters	No. of batters of Team 1	6(38), 7(263), 8(235), 9(50), 10(4)
No. of Bowlers	No. of bowlers of Team 1	1(4), 2(50), 3(235), 4(263), 5(38)
No. of Allrounders	No. of allrounders of Team 1	0(7), 1(31), 2(167), 3(184), 4(129), 5(61), 6(9),7(1), 8(1)
No. of Oppo. Batters	No. of batters of Team 2	6(38), 7(263), 8(235), 9(50), 10(4)
No. of Oppo. Bowlers	No. of bowlers of Team 2	1(4), 2(50), 3(235), 4(263), 5(38)
No. of Oppo. Allrounders	No. of allrounders of Team 2	0(7), 1(31), 2(167), 3(184), 4(129), 5(61), 6(9),7(1), 8(1)
Toss Win	Team 1 won the toss or not	Yes(295), No(295)
Toss Decision	Team 1 choosed bat or bowl	Bat(295), Bowl(295)
Ground	Where the stadium situated	Dhaka(392), Chittagong(138), Khulna(16), Sylhet(44)
Weather	Weather condition during game time	Broken Clouds(6), Clear(38), Fog(204), Haze(182), Light Rain(4), Mostly Clouds(2), Overcast(4), Partly Clouds(2), Partly Sunny(12), Passing Clouds(46), Scattered Clouds(52), Sunny(38)
Player of the Match Team	Team of the player of the match	Barisal(48), Chittagong(80), Cumilla(76), Dhaka(110), Khulna(68), Rajshahi(80), Rangpur(80), Sylhet(48)
Player of the Match Type	Playing role of player of the match	Allrounder(202), Batsman(260), Bowler(128)
Score	Final score of Team 1	Below 101(45), 101-150(272), 151-200(248),Above 200(25)
Result	Match result for Team 1	Win(295), Loss(295)

performs comparatively better than single or weak classifier [25]. That is why we have experimented with ensemble classifiers on our dataset, intending to get better prediction accuracy.

1) Random Forest (RF): We applied Random Forest classifier in our dataset, which is comprised of multiple decision trees [26]. And predictions from individual decision trees are combined for predicting the final class label.

2) Stochastic Gradient Descent (SGD): It is a type of Gradient Descent Optimization algorithm [27]. Instead of taking all the data points, SGD randomly selects some data points for every iteration. Though SGD takes more iterations, but it is still computationally inexpensive than Gradient Descent. We set the maximum number of iterations[3] for our dataset to 100.

3) XGBoost: XGBoost is an ensemble classifier that primarily combines boosted decision trees. It gradually trains the dataset by building subtrees, and in each iteration, it updates the weights in such a way that it provides less error than the previous iteration [28]. We set the number of trees or the n_estimators[4] value to 200 while training our dataset.

4) AdaBoost: Ada-Boost, also known as Adaptive Boosting ensemble boosting classifier, combines weak classifiers and reduces error. It follows an iterative approach, and in each iteration, it readjusts weights of classifiers [29]. So the error rate decreases significantly in each iteration. We set the number of iteration to 300 while training our dataset.

5) Gradient Boosting (GB): Gradient Boosting is an ensemble learning algorithm. This algorithm iteratively combines weak learners(decision trees) and gradually reduces error gradient in each iteration. We set the iteration number to 250 for training our dataset using Gradient Boosting.

[3]https://realpython.com/gradient-descent-algorithm-python/

[4]https://machinelearningmastery.com/tune-number-size-decision-trees-xgboost-python/

V. RESULT AND ANALYSIS

A. Performance Evaluation

We have implemented the supervised classification algorithms and calculated the Recall, Precision, and F1-score to evaluate the algorithms' performance. We have performed a comparative analysis among the applied models using these performance metrics.The formulas for the aforementioned evaluation metrics are shown in the equations below:

$$Precision = \frac{True\ Positive}{True\ Positive + False\ Positive} \tag{2}$$

$$Recall = \frac{True\ Positive}{True\ Positive + False\ Negative} \tag{3}$$

$$F1 - Score = \frac{2 * (Precision * Recall)}{Precision + Recall} \tag{4}$$

B. Performance Analysis of Classifiers

The performances of all the classifiers in the case of both pre and post-match winner prediction tasks have been presented in Table III. Among non-ensemble classifiers, KNN performed best on the dataset containing pre-match features with precision, recall, and f1-score values, respectively 0.63358, 0.69153, and 0.66118. While Naive Bayes, SVM, and Logistic Regression performed very close to KNN. But Decision Tree classifier was the worst-performing algorithm on the pre-match dataset. On the other hand, SVM provides the best

TABLE III
PERFORMANCE MEASURE OF DIFFERENT CLASSIFICATION ALGORITHMS

Model		Pre-Match Features			All Features		
		Precision	Recall	F1-score	Precision	Recall	F1-score
Base Classifiers	Decision Tree	0.55415	0.53559	0.53304	0.66528	0.67119	0.69033
	KNN	**0.63358**	**0.69153**	**0.66118**	0.89224	0.86102	0.87616
	Naïve Bayes	0.64614	0.63729	0.64068	0.66795	0.69153	0.67826
	SVM	0.62731	0.66102	0.64277	0.88936	0.90169	0.89455
	Logistic Regression	0.62515	0.64068	0.63251	0.66207	0.68136	0.67069
Ensemble Classifiers	Random Forest	0.63793	0.61017	0.63123	0.84488	0.83051	0.83962
	SGD	0.61485	0.62661	0.62319	0.63026	0.68814	0.62445
	XGBoost	0.62363	0.66441	0.64252	0.93267	0.92542	0.92874
	AdaBoost	0.62582	0.61695	0.62123	0.67597	0.68136	0.67666
	Gradient Boosting	0.61950	0.63051	0.62966	**0.93757**	**0.92542**	**0.93271**

classification performance with precision = 0.88936, recall = 0.90169, and f1-score = 0.89455 while performing on the dataset containing all features. Though KNN was closer to SVM, the other three base classifiers' performance fell drastically in the dataset containing all features.

Ensemble classifiers performed almost equally on the pre-match feature dataset, but XGBoost provides comparatively better performance with 0.62363, 0.66441, and 0.64252 recall, precision, and f1-score value, respectively. While performing on all features, Gradient Boosting provided the best, and XGBoost provided second-best performances. In a nutshell, among all the classifiers, KNN performed best on the pre-match features dataset, and Gradient Boosting outperformed other algorithms on the all-feature dataset.

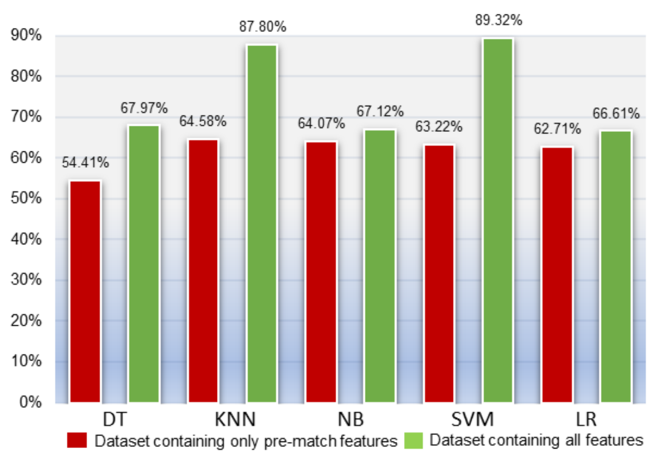

Fig. 2. Accuracy of base classifiers on both scenario

A comparison of classification accuracy of base classifiers is shown in figure 2 as bar chart representation. The red bars represent the accuracy found on the pre-match feature dataset, and the green bars represent the accuracy found on the dataset containing all features. We observed KNN gives the best classification accuracy of 64.58% on the pre-match dataset, and SVM gives 89.32% accuracy, the best among base classifiers.

Figure 3 shows the accuracy of ensemble classifiers on two datasets. In this case, blue bars represent accuracy on the pre-match feature dataset, and yellow bars show the accuracy of

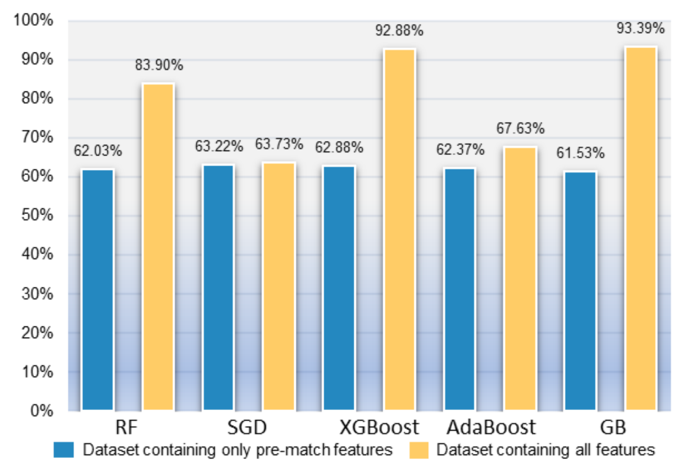

Fig. 3. Accuracy of ensemble classifiers on both scenario

all feature datasets. We found that Stochastic Gradient Descent gives the best accuracy on the pre-match dataset of 63.22%, and Gradient Boosting gives the best classification accuracy of 93.39% on all feature datasets.

So, overall, among all 10 algorithms, KNN was found best for predicting the game result before starting the game, and Gradient Descent performed best for predicting while post-match features are included.

A Receiver Operating Characteristic(ROC) curve is a two-dimensional plot representing a classification algorithm's performance. The rates of false positives and true positives are represented on the x and y axes, respectively, for the predictive test. The AUC score reflects how well the classifier differentiates between positive and negative categories. The better, the higher the AUC score.

We have plotted the ROC curve for analyzing the performances of our used classifier for both the pre-match feature [Figure 4]and all feature datasets [figure 5]. The left and right portions of figure 4 represent the ROC curve and AUC scores for non-ensemble and ensemble classifiers, respectively. We see, among base classifiers KNN (0.669). Among ensemble classifiers, AdaBoost(0.684) has the highest AUC scores, which are not excellent. Still, it's acceptable discrimination, and it is obviously quite good while predicting match results before starting the game compared to some existing works we

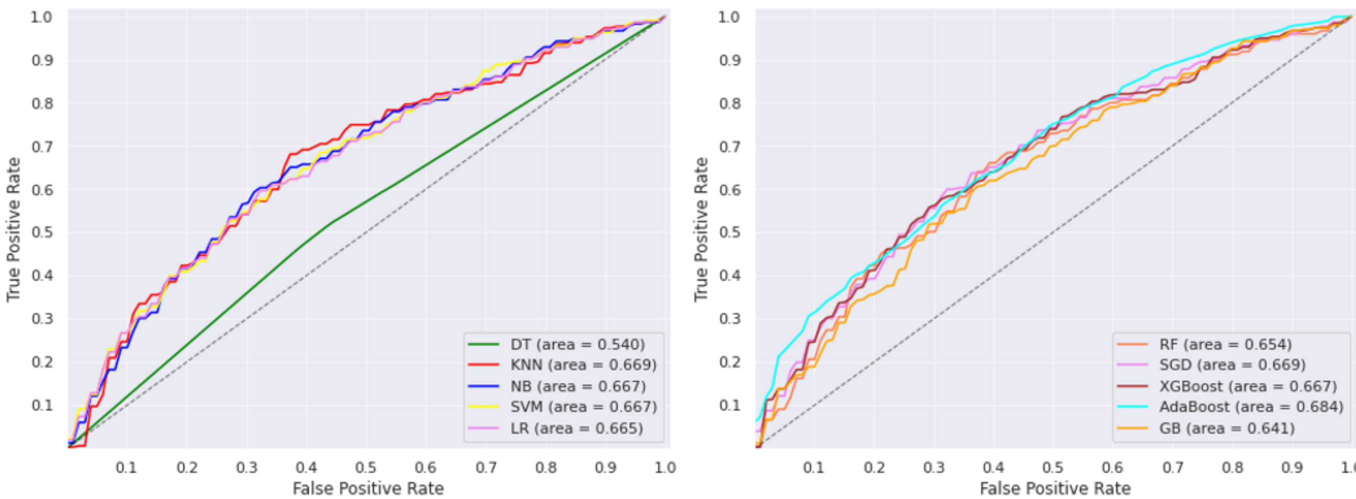

Fig. 4. Base and ensemble classifiers' ROC curves and AUC scores for pre-match features

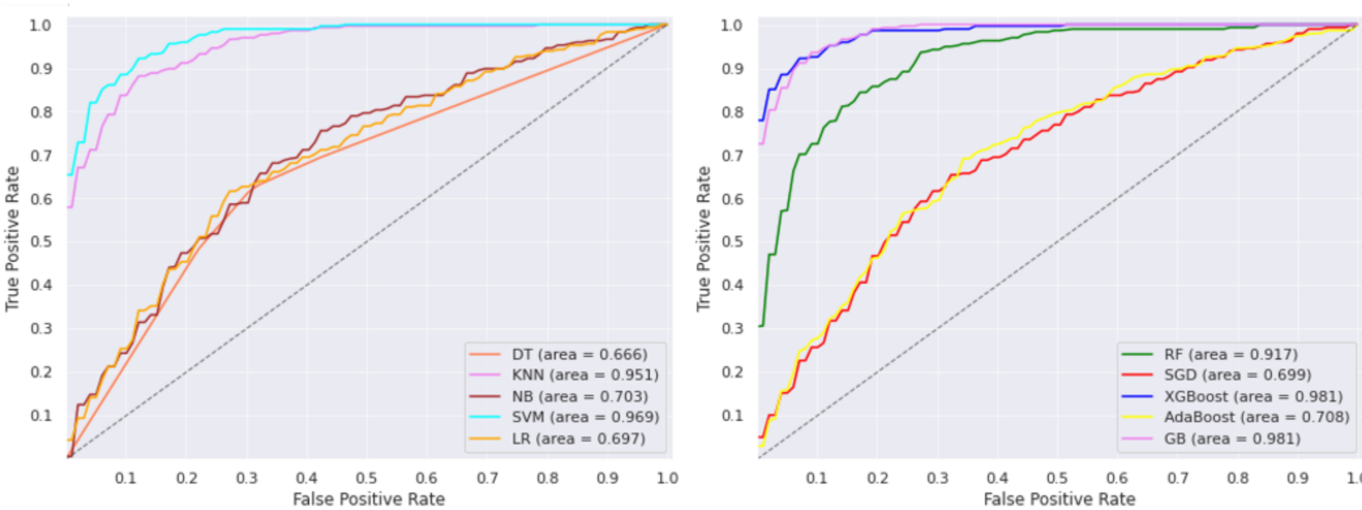

Fig. 5. ROC curve and AUC scores of base and ensemble classifiers for all features

have discussed in the related work section.

While in figure 5 we see SVM classifier shows the highest AUC score of 0.969 among base classifiers. And both XG-Boost and Gradient Boosting provide the highest AUC score of 0.981 and outperformed other ensemble classifiers while they have performed in the dataset containing all features. So we can conclude that, for all feature dataset, SVM, XGBoost, and Gradient Boosting provides outstanding discrimination in terms of AUC measure.

VI. Conclusion

In this work, we have applied several non-ensemble and ensemble classifiers to analyze their performances for the match outcome prediction of Bangladesh Premier League(BPL) T20 matches. We have predicted match results in two ways based on only pre-match features and with the inclusion of post-match features. We have prepared our dataset covering all the matches held from season 2011-12 to 2019-20 in the BPL T20 tournament. We have thoroughly analyzed and compared the performance of the applied classifiers. While predicting match results before starting the game, we received the highest of 64.58% accuracy from KNN, and 93.39% accuracy is found from the Gradient Boosting algorithm while predicting match results from all historical data, including post-match features. Our proposed approach could easily be used in all the cricket tournament formats and types to predict match results and analyze game facts. In the future, we will collect data in a similar approach but to a larger extent like ODI Cricket World Cup. We will follow the more robust process and apply advanced algorithms to increase classification accuracy.

References

[1] A. Abid, B. Hassan, K. Abid, U. Farooq, M. S. Farooq, and M. A. Naeem, "Sports culture in south asia: Effects of modern bowling action rules on cricket, an information technology perspective," *South Asian Studies*, vol. 33, no. 1, 2020.

[2] A. Kavoori, “Cricket, media and the nation: An autoethnographic exploration of three mediated moments in indian cricket,” 2021.

[3] D. Malcolm and S. Naha, “Cricket at the beginning of the long twenty-first century,” *Sport in Society*, pp. 1–7, 2021.

[4] K. Kapadia, H. Abdel-Jaber, F. Thabtah, and W. Hadi, “Sport analytics for cricket game results using machine learning: An experimental study,” *Applied Computing and Informatics*, 2020.

[5] R. Lamsal and A. Choudhary, “Predicting outcome of indian premier league (ipl) matches using machine learning,” *arXiv preprint arXiv:1809.09813*, 2018.

[6] D. M. Vistro, F. Rasheed, and L. G. David, “The cricket winner prediction with application of machine learning and data analytics,” *International Journal of Scientific & Technology Research*, vol. 8, no. 09, 2019.

[7] K. Passi and N. Pandey, “Increased prediction accuracy in the game of cricket using machine learning,” *arXiv preprint arXiv:1804.04226*, 2018.

[8] I. Wickramasinghe, “Naive bayes approach to predict the winner of an odi cricket game,” *Journal of Sports Analytics*, vol. 6, no. 2, pp. 75–84, 2020.

[9] A. Tripathi, R. Islam, V. Khandor, and V. Murugan, “Prediction of ipl matches using machine learning while tackling ambiguity in results,” *Indian Journal of Science and Technology*, vol. 13, no. 38, pp. 4013–4035, 2020.

[10] N. Pathak and H. Wadhwa, “Applications of modern classification techniques to predict the outcome of odi cricket,” *Procedia Computer Science*, vol. 87, pp. 55–60, 2016.

[11] N. Abhishek, P. Shivanee, N. Minakshee, and M. Sahil, “Winning prediction analysis in one-day-international (odi) cricket using machine learning techniques,” *International Journal of Emerging Technology and Computer Science*, vol. 3, no. 2, pp. 137–144, 2018.

[12] J. Kumar, R. Kumar, and P. Kumar, “Outcome prediction of odi cricket matches using decision trees and mlp networks,” in *2018 First International Conference on Secure Cyber Computing and Communication (ICSCCC)*, pp. 343–347, IEEE, 2018.

[13] T. Mahmood, M. Riaz, M. Nasir, U. Afzal, M. H. Siddiqui, *et al.*, “Psl eye: Predicting the winning team in pakistan super league (psl) matches,” *KIET Journal of Computing and Information Sciences*, vol. 4, no. 2, pp. 13–13, 2021.

[14] S. Singh, Y. Aggarwal, and K. Kundu, “Quantitative analysis of forthcoming icc men’s t20 world cup 2020 winner prediction using machine learning,” *International Journal of Computer Applications*, vol. 975, p. 8887, 2020.

[15] N. WEERADDANA and S. PREMARATNE, “Unique approach for cricket match outcome prediction using xgboost algorithms,” *Journal of Theoretical and Applied Information Technology*, vol. 99, no. 9, 2021.

[16] M. M. Rahman, M. O. F. Shamim, and S. Ismail, “An analysis of bangladesh one day international cricket data: a machine learning approach,” in *2018 International Conference on Innovations in Science, Engineering and Technology (ICISET)*, pp. 190–194, IEEE, 2018.

[17] S. A. Mahtab, N. Islam, and M. M. Rahaman, “Sentiment analysis on bangladesh cricket with support vector machine,” in *2018 International Conference on Bangla Speech and Language Processing (ICBSLP)*, pp. 1–4, IEEE, 2018.

[18] Y. Ren, L. Zhang, and P. N. Suganthan, “Ensemble classification and regression-recent developments, applications and future directions,” *IEEE Computational intelligence magazine*, vol. 11, no. 1, pp. 41–53, 2016.

[19] B. Charbuty and A. Abdulazeez, “Classification based on decision tree algorithm for machine learning,” *Journal of Applied Science and Technology Trends*, vol. 2, no. 01, pp. 20–28, 2021.

[20] Priyanka and D. Kumar, “Decision tree classifier: a detailed survey,” *International Journal of Information and Decision Sciences*, vol. 12, no. 3, pp. 246–269, 2020.

[21] D. J. Hand, “Principles of data mining,” *Drug safety*, vol. 30, no. 7, pp. 621–622, 2007.

[22] F.-J. Yang, “An implementation of naive bayes classifier,” in *2018 International Conference on Computational Science and Computational Intelligence (CSCI)*, pp. 301–306, IEEE, 2018.

[23] Y. Y. Liu, M. Yang, M. Ramsay, X. S. Li, and J. W. Coid, “A comparison of logistic regression, classification and regression tree, and neural networks models in predicting violent re-offending,” *Journal of Quantitative Criminology*, vol. 27, no. 4, pp. 547–573, 2011.

[24] P. Kazienko, E. Lughofer, and B. Trawiński, “Hybrid and ensemble methods in machine learning j. ucs special issue,” *J Univers Comput Sci*, vol. 19, no. 4, pp. 457–461, 2013.

[25] T. G. Dietterich, “Ensemble methods in machine learning,” in *International workshop on multiple classifier systems*, pp. 1–15, Springer, 2000.

[26] A. B. Shaik and S. Srinivasan, “A brief survey on random forest ensembles in classification model,” in *International Conference on Innovative Computing and Communications*, pp. 253–260, Springer, 2019.

[27] L. Bottou, “Stochastic gradient descent tricks,” in *Neural networks: Tricks of the trade*, pp. 421–436, Springer, 2012.

[28] N. Dhieb, H. Ghazzai, H. Besbes, and Y. Massoud, “Extreme gradient boosting machine learning algorithm for safe auto insurance operations,” in *2019 IEEE International Conference on Vehicular Electronics and Safety (ICVES)*, pp. 1–5, IEEE, 2019.

[29] N. K. Sinha, M. Khulal, M. Gurung, and A. Lal, “Developing a web based system for breast cancer prediction using xgboost classifier,” *International Journal of Engineering Research Technology (IJERT)*, vol. 9, 2020.