International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 10 Issue: 06 | Jun 2023 www.irjet.net p-ISSN: 2395-0072

Prediction Of Right Bowlers For Death Overs In Cricket

Lakshmi Shaju[1], Rahul K Sajith[2], Reshma N P[3], Vinayak N[4], Dr. Paul P Mathai[5]

[1,2,3,4] *Student, Dept. of CSE, FISAT, Ernakulam, India*
[5] *Professor, Dept. of CSE, FISAT, Ernakulam, India*

Abstract -

Predicting the right bowlers for the death overs in cricket is crucial for the success of a team, as these overs can often determine the outcome of a match. Death overs refer to the final overs of an innings, which are usually considered to be the most crucial because they can determine the result of the match. The death overs are typically the last 5-6 overs of an innings, and are known for being high-pressure situations due to the need to score runs or take wickets. Batting team has always an advantage in winning the match, by selecting the right bowlers to bowl in death overs, it would create an edge for the bowling team to win the match. Selecting the right bowlers for death overs requires an in-depth understanding of the strengths and weaknesses of different bowlers and the ability to analyze various factors such as the pitch conditions, the opposition team, and the current game situation. Some of the key factors that can be used to predict the right bowlers for death overs include the bowler's past performance in similar situations, their ability to bowl under pressure, and their ability to take wickets, their ability to bowl yorkers and slower balls. In order to accurately predict the right bowlers for death overs, it is important to consider all of these factors and use statistical analysis and machine learning techniques to build models that can accurately predict their performance. These analysis can then be used to build models that can accurately predict which bowlers are likely to perform well in the death overs based on these factors.

1.INTRODUCTION

Cricket is a bat-and-ball sport played by two teams of eleven players on a field with a 22-yard (20-metre) pitch in the centre and a wicket at either end, each with two bails balanced on three stumps. The batting side scores runs by striking the ball bowled at one of the wickets with the bat and running between the wickets, while the bowling and fielding side attempts to prevent this (by preventing the ball from leaving the field and getting the ball to either wicket) and dismiss each batter (so they are "out").Being bowled, when the ball strikes the stumps and dislodges the bails, and the fielding side either collecting the ball after it is struck by the bat but before it touches the ground, or hitting a wicket with the ball before a batter can cross the crease in front of the wicket, are all methods of dismissal. The innings finishes when ten hitters are struck out, and the sides switch roles. In international matches, the game is judged by two umpires, who are assisted by a third umpire Twenty20 (T20) cricket is a shorter game format. In 2003, the England and Wales Cricket Board (ECB) introduced it at the professional level for inter-county play.[1]Each side has one innings in a Twenty20 competition, with a maximum of 20 overs.Twenty20 cricket, along with first-class and List A cricket, is one of three modern forms of cricket acknowledged by the International Cricket Council (ICC) as being at the top international or domestic level. A typical Twenty20 game lasts about two and a half hours, with each innings lasting about 70 minutes and a 10-minute formal pause in between. This game is significantly shorter than previous iterations, and it corresponds to the length of other famous team sports. It was intended to create a fast-paced game that would appeal to both on-the-ground fans and broadcast watchers. The pastime has spread throughout the cricket world. Most international tours include at least one Twenty20 encounter, and every Test-playing country has a local cup tournament. Following that, the Big Bash League, Bangladesh Premier League, Pakistan Super League, Caribbean Premier League, and Afghanistan Premier League all used identical formulas and were well-liked by supporters. The Women's Big Bash League was established by Cricket Australia in 2015, and the Kia Super League began in England and Wales in2016.The Mzansi Super League of South Africa began in 2018. Numerous T20 competitions follow the general pattern of a group stage followed by a Page playoff system among the top four teams in which: • The first- and second-placed team in the group stage encounter, with the winner progressing to the final. • The third- and fourth-place teams square off, with the loser ousted; and • the two teams that have notyet progressed to the final after the previous two matches are played fight for the second spot in the final. In the Big Bash League, an extra encounter is contested to decide whether the fourth or fifth-placed club will qualify for the top four. We acquire the necessary data from websites such as Cricinfo, Cricbuzz, and Cricmetric, which will provide us with our information. The data is then assembled in order to extract the required traits as effectively as possible. Bowling average, wickets gained, economy, and dot ball average are used to evaluate the bowler against a specific batter, whereas hitting strike rate and batting average are used to evaluate the striker. These acquired data are preprocessed so that the ml model can predict correctly. We suggest using Decision Tree with tagging, Random Forest, and SVM as ML models. As a consequence, for each batter, a similar bowler from the opposing side will be expected. This improves the hitting team's odds of winning.

2. Literature Survey

2.1 Performance Prediction and Squad Selection

The goal is to forecast a player's performance in a forthcoming competition so that the team management may make preparations. The performance of the participants (bowlers and batsmen) in a forthcoming tournament is anticipated by taking into account the results of the event's previous iterations. For forecasting the players' next tournament performance, many machine learning techniques are used, including Random Forest, Decision Trees, K-nearest Neighbors, Support Vector Machine, and Naive Bayes. After the players are forecasted, the best squads are determined using two well-known optimization algorithms, NSGA-II (Non-dominated sorting genetic algorithm II) and SPEA2 (Strength Pareto Evolutionary Algorithm II). Using machine learning and multi-objective optimization techniques to predict a cricketer's tournament wise performance and select a squad for a cricket team is a complex task, as there are many factors that can influence a player's performance in a given tournament. These factors could include the player's form, the quality of the opposition, the pitch conditions, and the player's role within the team, among others. To select a squad for a cricket team, you could use a multi-objective optimization technique such as genetic algorithms. This would involve defining a set of objectives (such as maximizing the team's overall batting and bowling strength) and using the optimization algorithm to search for a combination of players that best meets these objectives.

2.2 Extraction of strong and weak region of cricket batsman

In this paper, extraction of strong and weak regions of cricket batters through text-commentary analysis is done. The paper is prepared considered the latest format of game i.e. T20 format. The ground is divided into different regions like slip, gully, point , cover , mid-off etc .They remove the unwanted words and pun marks using NLTK libraries in Python. Stemming and wordNet lemmatizer are used to make text consistent and organized then the data is converted into csv format. This data is then matched with regular expressions and results are taken out and thus strong and weak regions of batsmen are determined. In this proposed methodology the overall functionality is divided into five steps. Regular Expression Algorithm is used in this paper to predict the strong and weak region of the batsman. Various frameworks are used to retrieve text comments while avoiding the impact of page scrolling. This cricket stroke written description includes every move that occurred during the real-time occurrence. After gathering the necessary data, pre-processing is performed to minimize the complexity of the data. Reduce the quantity of irrelevant data to increase efficiency. The trivial debate between the pundits is dropped.

2.3 Cricket Team selection using Evolutionary Optimization

They suggested a multi-objective strategy based on the NSGA-II algorithm to maximise the team's overall batting and bowling strength and locate team members within it. A player's batting average is calculated by dividing the total number of runs scored by the number of times he has been out. Only players who have scored at least 300 runs in international T-20 format are classified as batsmen. A bowler's bowling average is calculated by dividing the total number of runs conceded by the bowler by the number of wickets taken. An examination of the resulting trade-off solution revealed a favoured team with higher batting and bowling averages than the winning team of the previous IPL competition. The basic goal of the article is to create a squad of 11 players with optimal bowling, batting, and fielding performance while staying within budget and rule limits. After issue formulation, they use the elitist non-dominated sorting genetic algorithm to perform multi-objective genetic optimization on the team (NSGAII). A batsman's batting average is regarded as a good indicator of an individual player's ability. A bowler's bowling average is calculated by dividing the total number of runs conceded by the bowler by the number of wickets taken. The team's bowling performance, as measured by overall bowling average, is lowered. As a result, a fictitious penalty bowling average for all non-bowlers must be introduced to the objective function to lower the likelihood of the optimization method setting the team's net bowling average to zero. The proportion of matches won out of the total number of matches played in the role of captain, which is another quality of a player, is used to measure captaincy performance. It is also a key decision-making criterion. They would rather choose the team represented by a knee point.

2.4 Predicting the performance of batsmen in test cricket

The study used a three-level hierarchical linear model(HLM) to predict the performance of cricket players, taking into account anthropometric characteristics such as height and handedness. The analysis considered both inter-individual and intra-individual characteristics that could affect a player's performance. The collected data had a hierarchical structure of three levels and a longitudinal nature. At level one, individual performance was a function of match-related characteristics and random error. At level two, the level one intercept was a function of measurable individual characteristics and random error. At level three, the level two intercept was a function of higher-level characteristics and random error. The data revealed that several left-handed batsmen consistently scored runs during the study period. The study emphasized the need to establish general developmental principles that apply to all individuals in cricket, focusing on inter-individual variation.

2.5 Optimal batting orders in cricket

The difficulty of scoring runs and being dismissed is heavily influenced by the pitch quality, which differs significantly between matches and usually deteriorates during a traditional play. For cricketers, the most significant figure was their average, which is a measure of how many runs a batsman gets per game, or the total number of runs split by the number of dismissals. A continuous possibility of dismissal per run scored implies a geometric distribution for batsmen's scores. Folklore holds that batsmen are more prone to dismissals early in their innings, may become nervous or cautious when the score hits 90, and they strike out later in the game. They computed the difference directly in this article, using simplified game models and solving them with dynamic programming For a large number of overs (say, 80 or more, as in test cricket),the standard batting order in decreasing order of average is optimal, with the team expected to score about 245 runs and nothing to gain from a variable batting order. Similarly, with a limited number of overs (in this case 40 or fewer),it is advantageous to start with the faster-scoring batter, and flexible batting order yields minimal benefit. The models presented here are simplified, and the captain is arguably the models may be used to highlight the benefits of varied batting orders and persuade captains that such a strategy is favourable and that he can simply pick the best squad for the match using this model. The captain who enters a match with a fixed attitude and does not contemplate changing the batting order based on the conditions is not increasing his team's chances of winning. Selectors and captains have learned that rapid scoring is just as crucial as good averages in limited-overs cricket, and naturally fast scorers have been elevated to one-day sides as well as further up the batting order.

2.6 Decision Tree Algorithm

A decision tree algorithm is a machine learning algorithm used for both regression and classification tasks. The algorithm creates a tree-like model of decisions and their possible consequences. In a decision tree, each internal node represents a decision based on a particular feature, and each leaf node represents a class label or a numerical value. The goal ist o split the data into subsets that are as pure as possible, meaning that all of the examples in each subset belong to the same class. The algorithm works by selecting the best feature to split the data based on a particular criterion, such as information gain or Gini index. It then recursively splits the data into subsets based on the selected feature, creating child nodes and splitting those further until the stopping criteria is met. Decision trees are easy to interpret and can be used for both classification and regression tasks. However, they tend to overfit the data if the tree is too deep or if there are too many features. To avoid overfitting, techniques such as pruning or ensembling can be used.

3. COMPARISON

For cricket team prediction, we can compare the performance of Support Vector Machines (SVM), Decision Trees, Random Forests, and Naive Bayes algorithms based on their strengths and weaknesses for the task. SVM is a powerful classification algorithm that works well with both linearly and non-linearly separable data. It can handle high-dimensional data and can find complex decision boundaries. In the case of cricket team prediction, SVM can work well if there are many features that need to be considered in the selection of the team. However, SVM can be sensitive to the choice of hyperparameters, and training the model can be computationally expensive.

METHOD	ACCURACY
Random Forest	96%
Decision Tree	95%
SVM	92%
Naïve Bayes	88%
NSGA-II	86%

Decision Trees are easy to interpret and can work well for small to medium-sized datasets. They can handle both categorical and numerical data and can handle missing values well. In the case of cricket team prediction, decision trees can be useful for understanding which features are most important in the selection of the team. However, decision trees can be prone to overfitting, and their performance can degrade with increased complexity.

Random Forests are an extension of decision trees that can reduce overfitting and improve performance by combining multiple decision trees. They can handle high-dimensional data and can be effective in dealing with noisy or missing data. In the case of cricket team prediction, Random Forests can work well if there are many features that need to be considered, and if there is a risk of overfitting with decision trees. However, Random Forests can be computationally expensive and can be difficult to interpret.

Naive Bayes is a simple yet effective classification algorithm that works well with small to medium-sized datasets. It assumes that the features are independent of each other, which can be unrealistic in some cases. In the case of cricket team prediction, Naive Bayes can work well if there are relatively few features that need to be considered in the selection of the team. However, Naive Bayes can be sensitive to the quality of the input data, and may not work well if the assumptions of independence are not met.

In summary, the choice of algorithm for cricket team prediction depends on the specific requirements of the task. If there are many features that need to be considered, SVM or Random Forests may work well. If interpretability is important, decision trees may be a good choice. If the dataset is small and simple, Naive Bayes may be sufficient. Ultimately, it may be beneficial to try multiple algorithms and compare their performance on a validation set. In this project we propose to use decision tree(with bagging) and SVM for training and prediction.

4. METHODOLOGY

Fig. 1. Work Flow

4.1 DATASET

The proposed system includes data from websites like cricbuzz, cricinfo and cricmetric which will be converted to a csv form including all the details required. Then all the necessary data like dot ball percentage, bowling average, economy, wickets taken etc will be sorted. All the available data are present in raw form in websites like cricmetric, cricinfo and cricbuzz. The data is then either taken manually or using scraping tools in python and converted to a usable digital CSV format which will act as the primary data that can be used throughout the prediction process.

4.2 DATA PREPROCESSING

Data preprocessing is reassembling original data into appropriate data sets, because robots cannot use data that they cannot comprehend. Primary data is frequently insufficient and formatted incongruously. The success of every endeavour that needs data analysis or prediction is linked to the sufficiency or inadequacy of data preparation. Data preprocessing includes both data validation and imputation. The goal of validation is to determine if the data is both complete and exact. The goal of data imputation is to correct mistakes and fill in missing values before splitting the dataset into two independent datasets, one for analysis and the other for prediction. It is required to turn raw data into clean data for machine learning algorithms to function. To make machine learning algorithms operate, raw data must be converted into a clean data set, which implies the data set must be converted to numeric data. We do this by converting all category labels into column vectors with binary values. The presence of missing values, or NaNs (not a number), in the data collection is a concern. The missing rows are removed.

4.3 FEATURE SELECTION

- The dot ball percentage of the bowler against the batter.
- The bowling average of the bowler against the batter.
- The economy of the bowler.
- The number of times the bowler has taken the wicket of the batter.
- The batters strike rate against the bowler.
- The batters average against the bowler.

Data is compiled so as to extract required features effectively. Against a particular batsman, bowling average, wickets taken, economy, dot ball average are the features to evaluate the bowler whereas batting strike rate and batting average are the criteria used to evaluate batsman. These acquired data is pre-processed for Machine Learning model to predict accurately. We try to reduce the bias towards the batters by predicting the best bowlers for the batter during the death overs. The Machine Learning models intended to use are Decision Tree with bagging and SVM. The output expected is that for a particular batsman, corresponding bowler will be predicted among the 5 bowlers in the opposite team. This will increase chances of winning for the batting team.

5. RESULT

The results obtained from the experiments are presented in this section. It was observed that Random forest provided a greater accuracy, precision, recall and f1 score when compared and had an accuracy of 96%. A comparative analysis of the Random Forest and Decision Tree models is performed based on their performance metrics. Any significant findings are analysed along with their implications for bowler prediction.

6. CONCLUSIONS

Machine learning is now used in the majority of real-world scenarios. Therefore, by introducing machine learning in the game of cricket it would be possible to find out the batsman's and bowler's performance in the forthcoming match. We will have an idea about the player's success during a competition as a result of this procedure, and the team administration will be able to select the best player. This work is used to predict the best possible team members that could give the best possible result in a match. Not all criteria are included

in the strategy mentioned in the papers chosen. In the death overs, the captain of the bowling team can send their best bowlers against the respective batsman and win the match according to the way how the match is being played rather than having 50:50 chance of winning. Also the batting team has more advantage when compared to the bowling team. This work aims to eliminate such advantage in the game.

REFERENCES

[1] Cricket, http://en.wikipedia.org/wiki/Cricket.

[2] Tirtho, Devopriya, Shafin Rahman, and Md Shahriar Mahbub."Crick-eter's tournament-wise performance prediction and squad selection using machine learning and multi-objective optimization." Applied Soft Computing 129 (2022): 109526.

[3] Rauf, Muhammad Arslan, et al. "Extraction of Strong and Weak Regions of Cricket Batsman through Text-Commentary Analysis." 2020 IEEE 23rd International Multitopic Conference (INMIC). IEEE, 2020.

[4] Wickramasinghe, Indika Pradeep. "Predicting the performance of batsmen in test cricket." Journal of Human Sport and Exercise 9.4 (2014): 744-751.

[5] Ahmed, Faez, Abhilash Jindal, and Kalyanmoy Deb."Cricketteam selection using evolutionary multi-objective optimization." Swarm, Evolutionary, and Memetic Computing: Second International Conference, SEMCCO 2011, Visakhapatnam, Andhra Pradesh, India, December 19- 21, 2011, Proceedings, Part II 2. Springer Berlin Heidelberg, 2011.

[6] Norman, John M., and Stephen R. Clarke. "Optimal batting orders in cricket." Journal of the Operational Research Society 61.6 (2010): 980-986.

[7] Mahbub, Md Kawsher, et al. "Best Eleven Forecast for Bangladesh Cricket Team with Machine Learning Techniques." 2021 5th International Conference on Electrical Engineering and Information Communication Technology (ICEEICT). IEEE, 2021.

[8] Anik, Aminul Islam, et al. "Player's performance prediction in ODI cricket using machine learning algorithms." 2018 4th international conference on electrical engineering and information communication technology (iCEEiCT). ieee, 2018.

[9] Ahmad, Haseeb, et al. "Prediction of rising stars in the game of cricket." IEEE Access 5 (2017): 4104-4124.

[10] Somaskandhan, Pranavan, et al. "Identifying the optimal set of attributes that impose high impact on the end results of a cricket match using machine learning." 2017 IEEE International Conference on Industrial and Information Systems (ICIIS). IEEE,