Cricket Winner Prediction

S Surya – 20BCE1071
SCOPE
VIT Chennai
surya.s2020@vitstudent.ac.in

Nilavan I – 20BCE1080
SCOPE
VIT Chennai
nilavan.i2020@vitstudent.ac.in

ABSTRACT

Motivation: Predicting the outcome of a match has always been at the center of sports analytics. Indian Premier League (IPL), a professional Twenty20 (T20) cricket league in India, has established itself as one of the biggest tournaments in cricket history.

Objectives: In this paper, we propose a model to predict the winner at the end of each over in the second in-nings of an IPL cricket match.

Method: Our methodology not only incorporates the dynamically updating game context as the game progresses, but also includes the relative strength between the two teams playing the match. Estimating the relative strength between two teams involves modeling the individual participating players' potentials. To model a player, we use his career as well as recent performance statistics.

Results: Using the various dynamic features, we evaluate several supervised learning algorithms to predict the winner of the match. Finally, using seven ML algorithms, we achieved an accuracy of 81%-90%with an overall accuracy of 83.42%.

Keywords— ML Algorithms, Ensemble Learning, Accuracy, Winner Prediction, Sports Analytics.

ACKNOWLEDGMENT

We wish to express our sincere thanks and deep sense of gratitude to our project guide, Dr. Jothi R, Assistant Professor (Sr Grade 1), School of Computer Science and Engineering, for her consistent encouragement and valuable guidance offered to us in a pleasant manner throughout the course of the project work.

We are extremely grateful to Dr. Ganesan R, Dean, School of Computer Science and Engineering, VIT Chennai, for extending the facilities of the school towards our project and for her unstinting support.

We express our thanks to our Head of the Department Dr. P. Nithyanandam for their support throughout the course of this project.

We also take this opportunity to thank all the faculty of the school for their support and their wisdom imparted to us throughout the course.

We thank our parents, family, and friends for bearing with us throughout the course of our project and for the opportunity they provided us in undergoing this course in such a prestigious institution.

I. INTRODUCTION

The use of statistical analysis in sports has been growing rapidly since the past decade. It has not only changed the way game strategies are formed or the players are evaluated, but also has impacted the way sports is viewed by the audience. Cricket is one of the most followed team games in the world with billions of fans all across the globe. The complex rules governing the game along with many other player-dependent and natural parameters provide ample opportunities to model the game from various perspectives.

Cricket has evolved over time. Today, it is played in three major formats – Test Matches, One Day Internationals (ODIs) and the T20 cricket. T20 cricket is the latest and the most exciting format of the game. Ever since its inception in 2007, IPL has been a huge success and has generated a billion-dollar industry. It is played during April and May of every year by teams representing Indian cities. Therefore, in this paper, we focus our study on the IPL cricket matches. We propose a model to predict the winner at the end of each match. Apart from various game dependent features such asthe number of balls remaining, the number of runs to be scored remaining, and the number of wickets remaining, we have used the relative team strength between the competing teams as a distinctive feature in predicting the winner of the match. A team is composed of players, hence, estimating the relative team strength between two competing teams requires us to estimate the potential of the players. Therefore, using the recent and career performance statistics of a player we define novel methods to render his batting and bowling capabilities, the two major roles of a player in the game of cricket. Using these features, we have evaluated various machine learning algorithms to predict the winner of the match.

II. METHOLOGY

A. *Overall Batting Average*

Calculation of Overall Batting Average: Batting Average is defined as the average number of *runs* scored by the batsman before he gets *out*. Batting average for the career statistics is calculated in the following way

(1*total_bruns+1*total_lbruns+2*b_non_zero-1*b_dot-3*b_dis)/5

where,

total_bruns – total runs in byes
total_lbruns – total runs in leg-byes
b_non_zero – total runs from bat
b_dot – total dot balls
b_dis – batsman dismissed

B. *Overall Bowling Average*

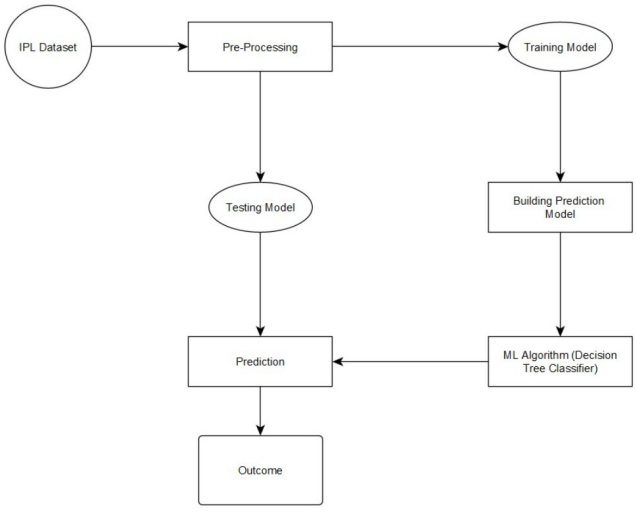
Calculation of Overall Bowling Average: Bowling Economy is defined as the average number of *runs* conceded by the bowler per *over* he bowls. Bowler average for the career statistics is calculated in the following way

(-1*total_wr-1*total_nb-2*total_pr-1*total_er-2*b_non_zero+5*b_dot+10*b_dis)/7

where,

total_wr – total wide bowled
total_nb – total no-balls bowled
total_pr – total penalty runs conceded
total_er – total extra runs conceded
b_non_zero – bowled a non-zero run ball
b_dot - bowled a dot ball
b_dis – dismissed a batsman

BLOCK DIAGRAM



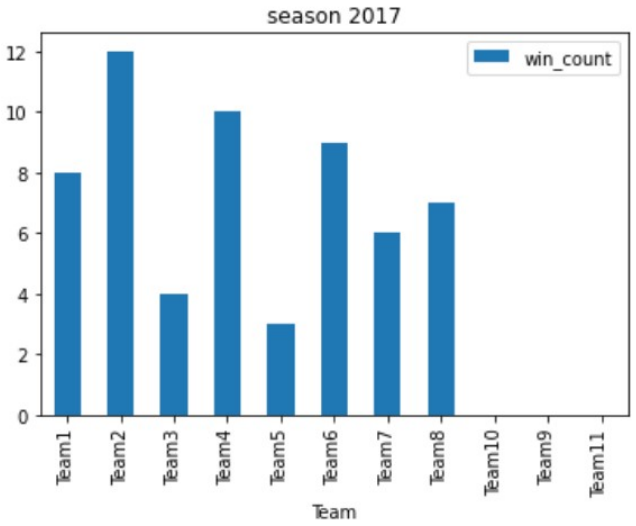
III. EXPERIMENTS AND RESULTS

A. *Dataset*

The dataset can be broadly divided into two categories – historical data: pertaining to the career statistics of players, and ball by ball data: pertaining to various states of a match. The dataset for career statistics has been scraped from the cricinfo website for all the matches played in the seasons 1-10 of IPL. The ball by ball data for each match in seasons 1-10 of IPL has been provided by the cricsheet website. The dataset constitutes the match statistics recorded after each ball, including runs scored, wickets lost, current batsmen, current bowler, winner of the match, date of fixture, etc. We combined data from both these sources to build our prediction model. IPL seasons 1-7 have been used for training our model, season 8 has been used for validating the parameters, and seasons 9 and 10 have been used as the test data.

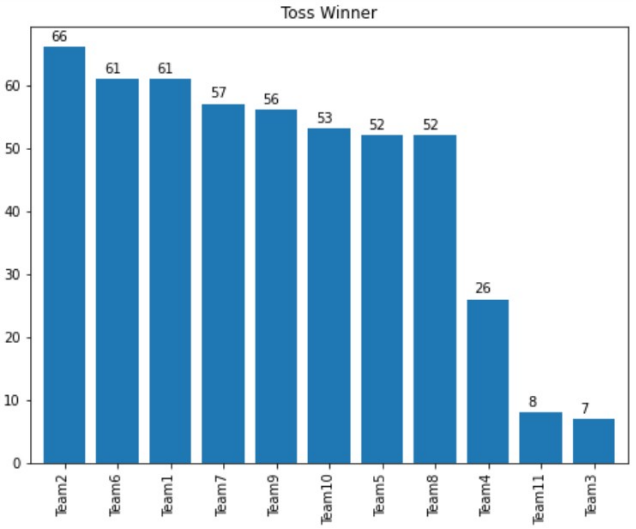
B. *Visualization*

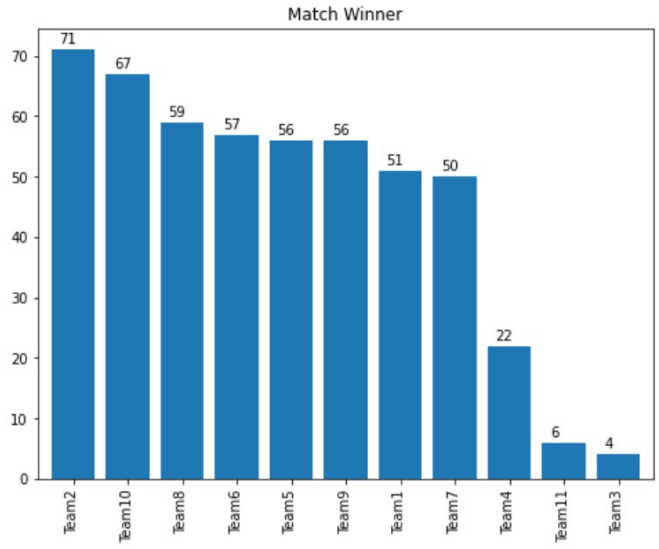
Graph 1: Count of Matches won by each team in 2017



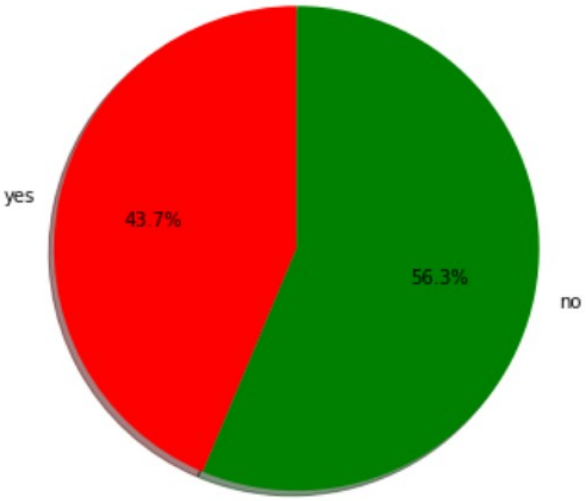
This graph shows number of matches won by each IPL team in the 2017 season.

Graph 2: Toss Factor



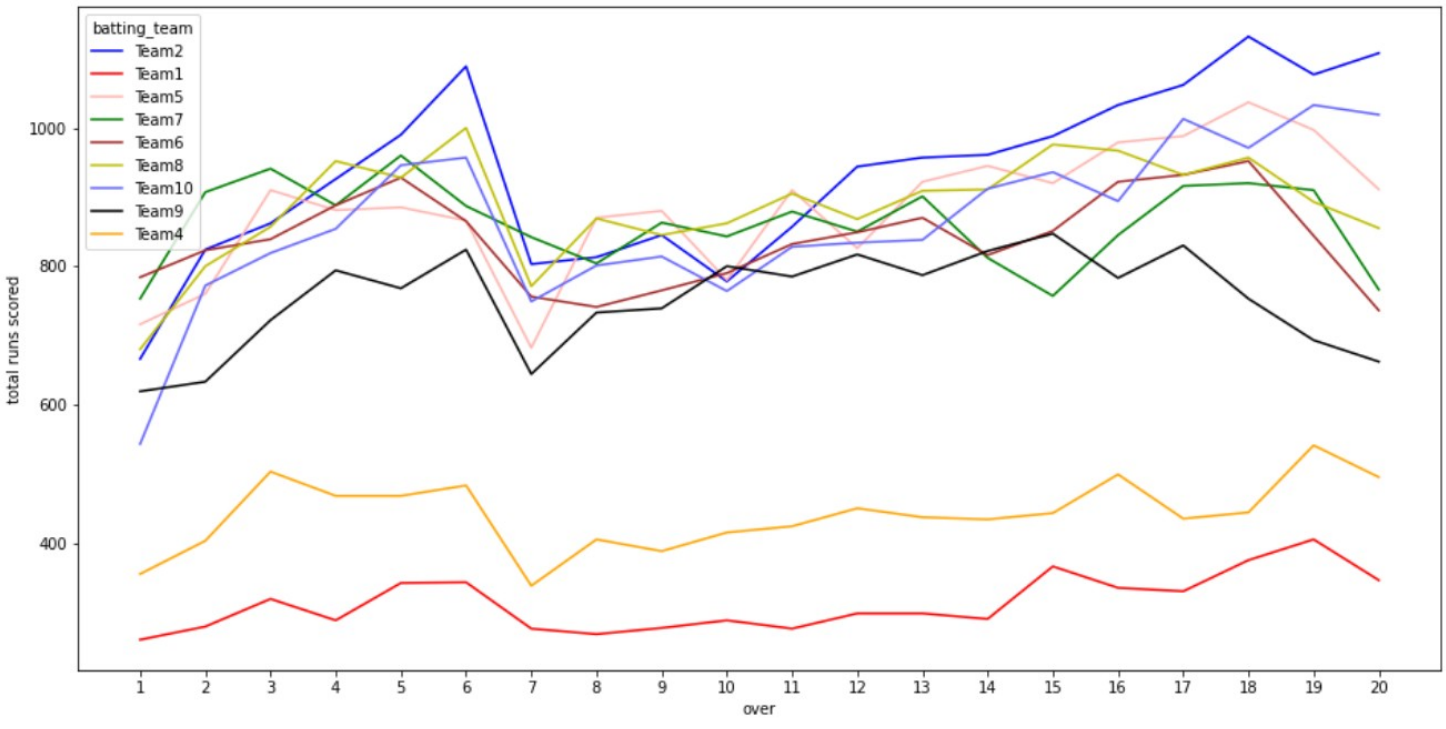


The two bar graphs shows the number of tosses won and matches won by each team respectively in the seasons 2008-2014.



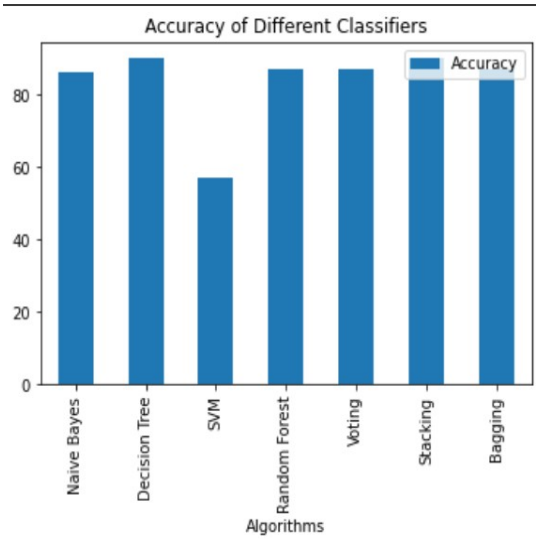
This pie chart indicates that teams that wins tosses wins 43.7% of the matches. Therefore, winning the loss might not be a better result for the team.

Graph 3: Number of runs scored by each team in particular over of a match



C. *Results*

	Algorithms	Accuracy
0	Naive Bayes	86
1	Decision Tree	90
2	SVM	57
3	Random Forest	87
4	Voting	87
5	Stacking	90
6	Bagging	87



Clearly, among the supervised machine learning models Decision tree is giving us the best accuracy. Also, among Ensemble techniques, Stacking is giving the highest accuracy. Whereas SVM model is giving us the least accuracy indicating that it is the least appropriate for this dataset.

IV. CONCLUSION AND FUTURE WORK

The problem of winner prediction in a Twenty20 cricket match has been successfully addressed in this paper. A combination of features (Toss, Stadium, Duckworth-Lewis method, Player of the match) which capture the state of the match have furnished promising results. Overall batting average and overall bowling average has been shown as an important feature that is successful in quantifying and comparing the strengths of the playing teams.

In order to expand this work, we can apply the same predictor for ODI cricket by adding some new features such as Home/Away, Timing of the game (Dew factor).

REFERENCES

1. D. Vistro, F. Rasheed, and L. David, "The Cricket Winner Prediction With Application Of Machine Learning And Data Analytics." [Online].Available: https://www.kclas.ac.in/wp-content/uploads/2021/01/The-Cricket-Winner-Prediction-With-Application-Of-Machine-Learning-And-Data-Analytics-.pdf

2. Omkar Mozar, Soham More, Shubham Nagare, and Prof. Nileema Pathak, "Cricket Score and Winning Prediction". [Online]. Available: https://www.irjet.net/archives/V9/i4/IRJET-V9I4181.pdf

3. S. Kumar and C. Nandini, "DATA SCIENCE APPROACH TO PREDICT THE WINNING FANTASY CRICKET TEAM-DREAM 11 FANTASY SPORTS." Accessed: Nov. 19, 2022. [Online]. Available: https://arxiv.org/pdf/2209.06999.pdf

4. S. Viswanadha, K. Sivalenka, M. Gopal Jhawar, and V. Pudi, "Dynamic Winner Prediction in Twenty20 Cricket: Based on Relative Team Strengths." Accessed: Nov. 19, 2022. [Online]. Available: https://ceur-ws.org/Vol-1971/paper-06.pdf