2025 chanakya university(CU)
Devanahalli, India. May 15-17, 2025

Analysis and Prediction for the Indian Premier League

Harshit Barot, Arya Kothari, Pramod Bide, Bhavya Ahir, Romit Kankaria
,Computer Engineering
Sardar Patel Institute of Technology
Mumbai, 400058
Email:imharshitbarot18@gmail.com, aryakothari26@gmail.com, pramod_bide@spit.ac.in, bhavya17ahir@gmail.com, romitvjain@gmail.com,

Authors: [Amrutha Girish Hegde], [Swapna K]
Affiliation: [Chanakya University]
Email: [hegdeamruta28@gmail.com,swapnak9019@gmail.com]

***Abstract*—Cricket is the most popular sport after football. Indian Premier League or the IPL is the most popular T20 domestic league in the world. Cricket involves lots of data and statistics. In the game of cricket, several parameters can be used to predict the outcome of the game. The factors affecting a cricket match can be combined with Machine Learning to predict the outcome of a match. This research has focused on analysing the features of the cricket matches in IPL. Moving further towards the analysis of the Indian Premier League, this paper has rated the Batsmen and Bowlers in a unique way based on their performance. A few crucial factors like team form and team strength in predicting the match outcome apart from the conventional features like the toss, venue of the games etc., have been added. Further, a novel analysis of Batting and Bowling has been proposed based on Batting Index and Bowling Index. Machine Learning algorithms like SVM, Logistic Regression, Random Tree, Random Forest and Naive Bayes have been applied, for match predictions. Lastly, the results, based on which algorithm gives the best accuracy, have been plotted. Decision Tree and Logistic Regression algorithms have given an accuracy over 87% and 95% respectively.**

***Index Terms*—Cricket, Sport, Machine Learning, Naive Bayes, Logistic Regression,Support Vector Machine, Random Forest, Decision Tree, Indian Premier League**

I. Introduction

Sports and statistics go hand in hand. Every professional sport has a scoreboard and stats of an individual or a team. Every team sport has a professional sport analyst who gives insights of the sports to the team coach or management. This not only helps to analyse the game, but also helps in improvement of skills for a particular player. Lately, there have been many sport companies that analyse the relevant sport statistics and numbers before broadcasting the matches.Taking a step further, based on the most crucial factors affecting a sport, several insightful predictions have also been made to improve sport performances. Lastly, analysing sports and showing it to the viewers in terms of simplified tables and graphs also has been popular and results in increasing revenue for the sports industry.

A. What is Cricket?

Cricket [15] is a game that originates from England. It was first played in the 16th century and spread globally after the expanse of the British empire. Cricket is played with a bat and a ball, on a cricket field, which is usually oval or circular. Two teams of eleven players each, compete against each other and the winner is decided by comparing the runs scored by each team. Cricket is globally recognised and viewed by more than 2.5 billion people. It's major following comes from South Asia, Australia and England.

There are three current major formats of Cricket, namely One-Day Internationals, Test Cricket and T-20 Cricket. T-20 Cricket is the latest of the three and it has gained popularity due to the entertainment it provides and the glamour it showcases.

B.The Indian Premier League

The Indian Premier League(IPL) is one the biggest T-20 Cricket leagues in the world. It was founded in 2008, and it takes place annually. Currently, 8 teams take part in the league and the top 4 progress to the play-offs, where the two teams who would complete for the trophy in the final are decided. The IPL has drawn attention from all over the world due its fast-paced action and roaring atmospheres.

The IPL has not only generated huge amounts of revenue every summer through media rights, ticket sales and sponsors, but also been a tournament where a young, aspiring cricketer can compete, perform, and prove himself against the world's best players. Young and successful talented cricketers have busted into international cricket after their impressive IPL performances.

C. Scope and Overview

Section II talks about the Literature Review of all the other papers referred. Further, in Section III the paper talks about our Implementation . This paper has tried to predict the outcomes of the IPL games by using Machine Learning algorithms and Analysis in this paper. For the best results, the records of the past 5 years - from 2015 to 2019 have been considered. The aim of this paper is to analyse all the insights from data, further moving to match predictions using these insights. Section IV speaks about Inference from the methodology. Lastly, Section V Concludes the paper along with the future scope of the research.

II. Literature Survey

Since the turn of this millennium, data science has found its way into every domain, including Cricket. Several researchers have contributed towards the analysis of Cricket.

Sasank et al. [1] found the batsman and bowler ratings and used the relative strength of the team to predict the outcome in the second innings of an IPL match dynamically. Chellapilla Deep Prakash, C. Patvardhan,C. Vasantha [3] presents multiple approaches to predict the winner of IPL season 9. Priyanka S, Vysali K, Dr K B PriyaIyer [4] used previous data from the previous editions of IPL and implemented data mining algorithms to predict the outcome of the 2020 edition of IPL.

Ananda Bandulasiri [2] analyzed the advantage a team has when playing on their home ground, and found the correlation between the winner of the coin toss and the outcome of the match. Ananda Bandulasiri [2] also analyzed the Duckworth Lewis Method and shows its effectiveness. Shilpi Agrawal, Suraj Pal Singh, Jayash Kumar Sharma [5] proposed a model to predict the winner of a match using three different machine learning algorithms and achieved a high accuracy with all three algorithms. The additional parameters which were considered in the paper were the batting strike rate and the bowling run rate, both considered separately even during the Power Play overs. Amal Chaminda Kaluarachchi, Aparna S. Varde [6] found that classification is the best approach to predict the winner of a match. The authors analysed the factors affecting the outcome of the match and brought forth a tool which could generate the chances of winning a match for a particular team based on some parameters. Madan Gopal Jhanwar, Vikram Pudi [7] used the past data of a player as well as the recent form to create a model about the player. The authors also used the k-Nearest Neighbor algorithm to determine the winner.

III. Implemented Methodology

A. Dataset

The dataset used for analysis and prediction was collected from www.kaggle.com [9] and has been scraped from www.stats.espncricinfo.com [10], where the data of the previous editions of IPL was available. Two datasets have been used. The first gives us the ball-to-ball information of every match ever played in the IPL, like the batsman, bowler, runs, wicket, and more, on each ball of the match. The second dataset gives us the summary of each match, which includes the teams playing, the winner, the winner of the toss, and more, for every match played in the IPL.

B. What affects the cricket game

A cricket match is affected by numerous factors [8]. These can be directly related to the game or external factors.
The list of factors affecting the game are as follows:

1) Toss: The toss is very crucial in determining whether a team bats first or fields first. A toss winning team thus always has this advantage.

2) Weather: Cricket is a game played outdoors. Thus weather is a crucial factor in determining the match winner. A Sunny weather is always good for batting, however when the weather becomes overcast the ball starts to swing thereby favouring the bowling team.Thus, weather plays a crucial role in a match outcome. However, most IPL games are played in the month of April and May in India, thus it's always a favourable weather.

3) Pitch: In cricket, pitch is a very important factor. A hard pitch is good for batting, a dusty pitch has the ball spinning more and a green pitch has the ball seaming more off it.

4) Bat First or Field First: After assessing the conditions and winning the toss, batting first or fielding first is a very important decision to make.

5) Venue: Every IPL team has a home ground advantage since it has more experience in playing on that ground. Thus the place where the game is played is also an important factor in the match outcome.

6) DLS Applied: If the weather is rainy and it rains during an ongoing game, the DLS rule is applied. In this case, the game length is generally shortened and targets are revised. Thus, DLS rule plays a role in a match outcome.

7) Team Strength: Cricket is a team sport. Thus by the collective efforts of all the players, a team is successful. Team Strength generally consists of batting strength(the collective performance of all the batsmen) and bowling strength(the collective performance of all the bowlers).

8) Team Form: The team's recent form often indicates its performance. If the team has been winning games lately, the momentum is towards their side. Thus, their performance is better.

C. Pre-processing and Feature Extraction

Attributes	
Season	DL_Applied
Toss_Winner	Home_Team
Team_A_Batting_Average	Team_B_Batting_Average
Team_A_Bowling_Average	Team_B_Bowling_Average
Team_A_Total_Runs	Team_B_Total_Runs
Team_A_Overall	Team_B_Overall
Team_A_Form	Team_B_Form

The data was pre-processed and cleaned to extract the necessary analysis, and also for the prediction module. The dataset prepared has considered the records of the last 5 years, that is, seasons 2015, 2016, 2017, 2018 and 2019. A total of 298 matches are taken into consideration over this period.

D. Analysis

1) The Toss Factor: According to the analysis, 55% of the toss winning teams end up winning the match as shown in Fig. 1. Winning the toss gives you the option to bat first or field first, which, when analysed beforehand, can be very advantageous to the team that wins the toss. Thus, the winner of the toss is worth analysing.

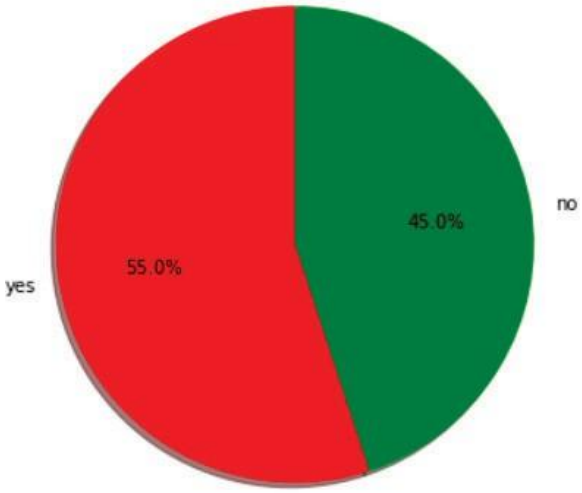


Fig. 1. Toss Winners are Match Winners

2) Bat and Win or Chase and Win: From the given pie-chart, we infer that teams chasing targets have won 57% of the times as shown in Fig. 2. Thus, chasing targets is seen more preferable rather than setting targets.

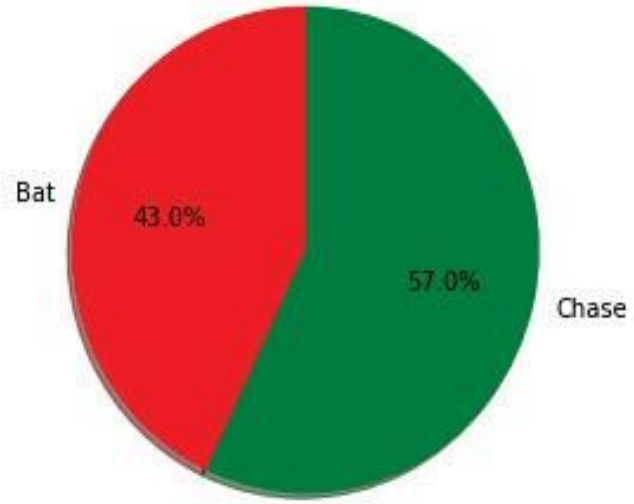


Fig. 2. Bat first or bat second and Win

3) Targets chased: As mentioned above, chasing a target is preferred in the IPL matches. However, when the target to be chased is greater than 200 runs, the inference is as follows.

In the figure given, the percentage of a target of 200 or more runs in the second innings in IPL is shown. It is seen that on only 15.6% of the times, is a target of 200 or more than 200 runs chased by the team batting second in an IPL match as shown in Fig. 3.

4) Runs Scored by teams per over on an average: Fig. 4 shows the total runs scored by a team in a particular over on an average. Here, it can be seen that all teams have a spike in their runs scored between the 1st and 6th over, the overs of the Powerplay, which is the period of overs where not more

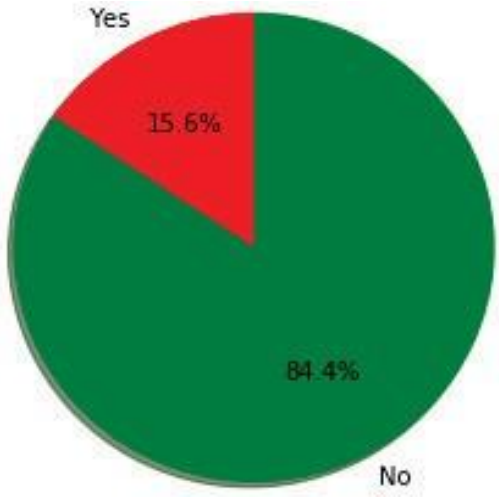


Fig. 3. Targets over 200 chased sucessfully

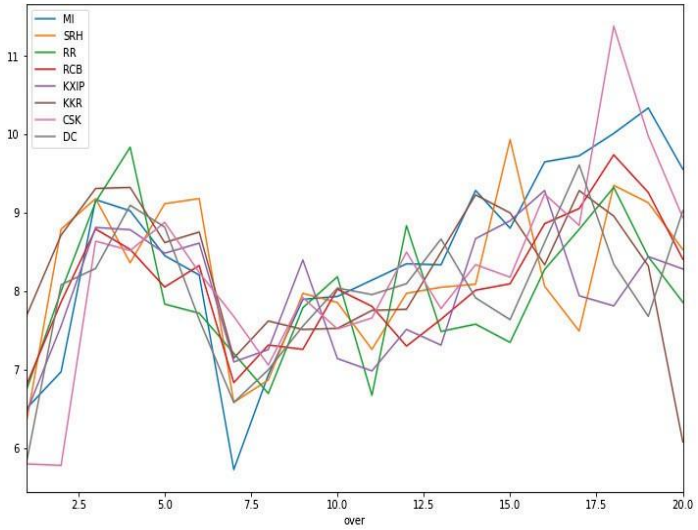


Fig. 4. Runs scored on Average every Over

than 2 fielders can stand outside the 30-yard circle. After this over there's a dip. Another spike in runs is seen at the end of the innings. This is where majority of the power hitters come and bat.

5) Batting Index: -

Batting Strike Rate- The strike rate of a player who is batting is defined as 100 x (the ratio of number of runs he scores to the number of balls he faces). Since IPL is a T20 league, a good batter needs to have a high strike rate.

Batting Average - The Batting average of a player is defined as the number of runs scored divided by the number of innings of a batter. The higher the average of a batsmen, the better he is. Batting Average marks the consistency of a batsmen.

Batting Index - This takes all the factors like runs scored, average,strike rate , number of matches played into consider-

ation. Mathematically ,

$$BattingIndex = \frac{(BattingAverage) * (BattingSR)}{100}$$

Higher the batting Index, better the batsman performance, as shown in Fig. 5. Thus the Batting Index is an apt parameter for judging a batsman.

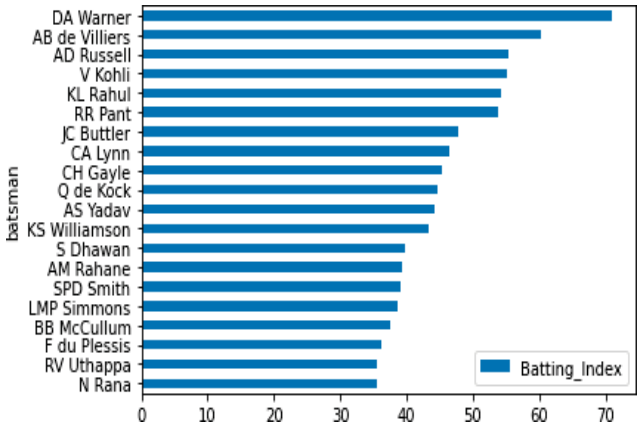


Fig. 5. Batting Index: The higher, the better the batsman

6) Bowling Index: -

Bowling Strike Rate - The strike rate of a bowler is defined as 100 x (the ratio of number of balls bowled per wicket taken). Lower the strike rate for a bowler, the better the bowler. A good bowler needs to have a low strike rate.

Bowling Average - The Bowling average of a player is defined as the number of runs conceded by the bowler divided by the number of wickets he picks up. The lower the average of a bowler, the better he is. Bowling Average marks the consistency of a bowler.

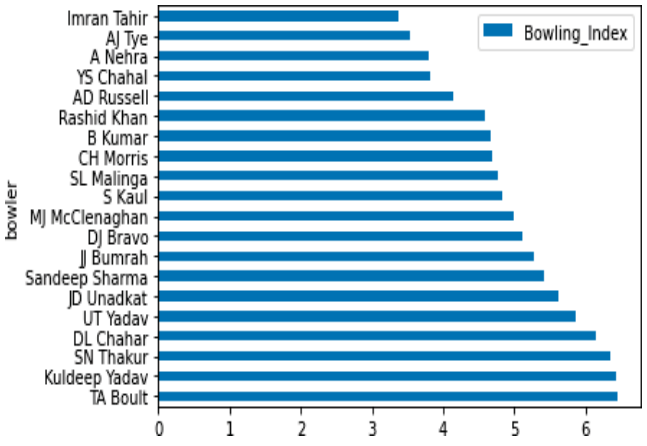


Fig. 6. Bowling Index: The lower, the better the bowler

Bowling Index - This takes all the factors like runs given, bowling average, strike rate of the bowler , number of matches played and number of wickets taken. Mathematically ,

$$BowlingIndex = \frac{(BowlingAverage) * (BowlingSR)}{100}$$

Bowling Index should be as low as possible for a good bowling performance, as shown in Fig. 6. Thus, the Bowling Index is an apt parameter for judging a bowler.

Performance of Bowler - In the scatter plot given below, we analyse the wickets taken and the economy rate of the top bowlers in IPL 2019, as shown in Fig. 7. It can be deduced that the bowlers featuring in the top left corner of the plot are much superior bowlers since their wicket-count is high and their economy rate is low-the desired combination. Whereas, the bowlers in the bottom right corner have an undesirable combination of low number of wickets and a greater economy rate.

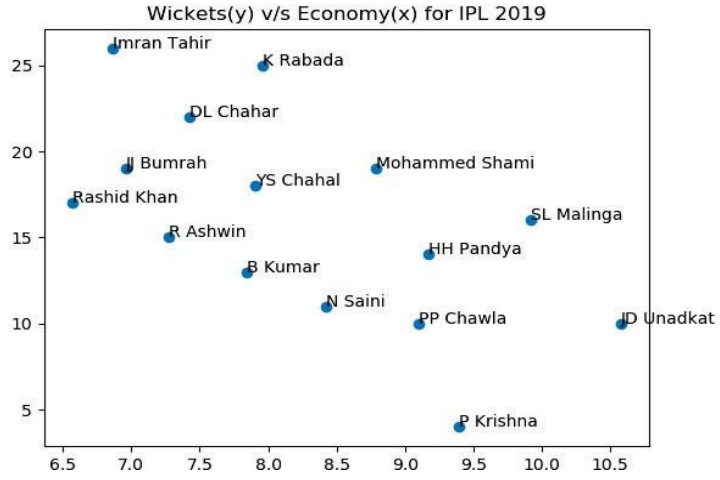


Fig. 7. Wickets taken VS Economy Rate

Dot Balls in the final over - The chart below gives the list of the bowlers who have bowled the most number of dot balls in the final over of an inning, as shown in Fig. 7. This can prove

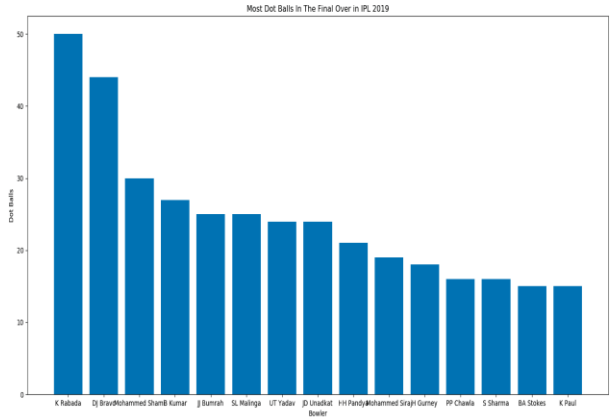


Fig. 8. Dot Balls Bowled in the final Over

to be extremely instrumental to a team's chances of victory. The chart also proves the ability of a bowler to bowl under pressure.

7) Form: Form - The form of a team was extracted from the data available. Here, form is considered as how well the team has fared in the previous 5 matches of the same season of IPL. The last 5 matches have been noted. A value of 2 has been added to the form for each victory, a value of 1 for each tied-match and a value of 0 for every loss. Therefore form for both the teams playing a particular match has been stored as a value out of 10.

E. Prediction

1) Naive Bayes Classifier: -

Naive Bayes [13] is a fast classification algorithm that is

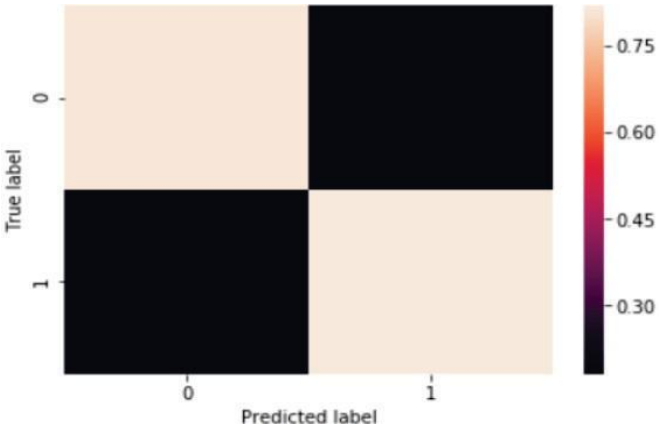


Fig. 9. The accuracy using Naive Bayes is 81.63%

suitable for a big dataset. It takes a specific feature and analyses its effect, independent of the other features. The Bayes' Theorem is as follows:

$$P(H/C) = \frac{P(C/H)P(H)}{P(C)},$$

where P(H) is the probability of the hypothesis H being true, P(C) is the probability of the characteristic taken into consideration to be true irrespective of the hypothesis, P(C/H) is the probability of the characteristic being true given that the hypothesis is true, and P(H/C) is the probability of the hypothesis being true, given the data of the characteristic. Fig. 9 shows the normalised confusion matrix for Naive Bayes Classifier.

2) Support-Vector Machine Classifier: -

A Support Vector Machine or SVM is a supervised machine learning algorithm which is used for both regression and classification problems. For classification, it involves a hyperplane which separates the 2 closest points from different groups in an N-dimensional plane. To obtain good results the margins of the hyperplane should be of maximum width. An SVM has multiple kernels like RBF, linear etc. which can be used according to the required need of the problem. Fig. 10 shows the normalised confusion matrix for SVM.

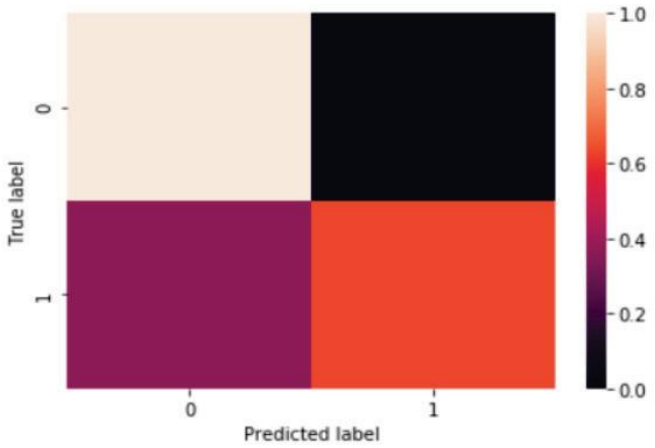


Fig. 10. The accuracy using SVC Classifier is 83.67%

3) Decision Tree Classifier: -

The Decision Tree Classifier [11] will analyse the different attributes in the data which are presented to it, and create subsets by splitting the data using these attributes. The Classifier will keep trying to split the data further into subsets till it finds the subset which contains all equal values for the target attribute. When the Classifier reads the test data, it will look at the subset in which the test data falls into and chooses the most-suited class, and thus predicts the outcome. Fig. 11 shows the normalised confusion matrix for Decision Tree Classifier.

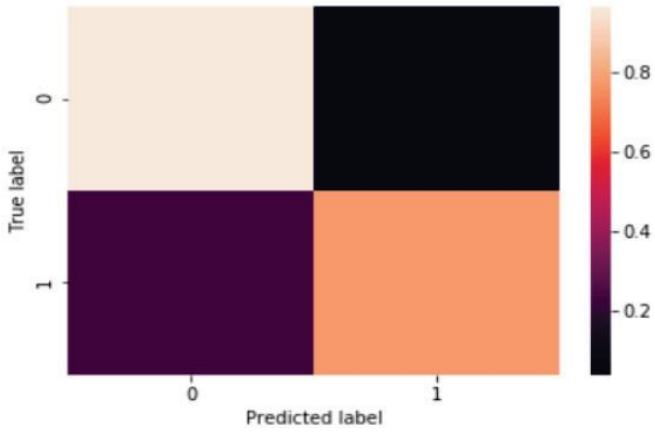


Fig. 11. The accuracy using Decision Tree Algorithm is 87.75%

4) Random Forest Classifier: -

The Random Forest Classifier consists of multiple decision trees, each with a random subset of attributes from the overall set of attributes. Each individual decision tree gives out a class prediction, and whichever class has the highest number of votes is then selected as the model's prediction. Fig. 12 shows the normalised confusion matrix for Random Forest Classifier.

5) Logistic Regression: -

In a Logistic Regression [14] model, only a binary outcome is possible. There is a threshold value set. Anything under the

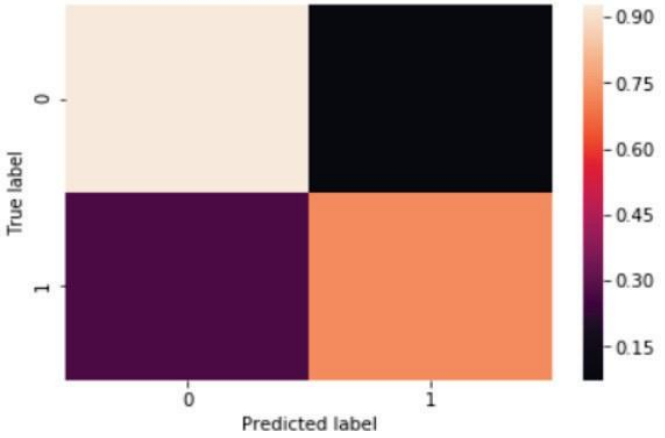


Fig. 12. The accuracy using Random Forest Algorithm is 83.67%

threshold value gives an output of 0 and anything above it gives an output of 1. It consists of a sigmoid function which gives an output between 0 and 1. It is a supervised machine learning algorithm. Fig. 13 shows the normalised confusion matrix for Logistic Regression

$$sigmoid(Z) = \frac{1}{1 + e^{-Z}},$$

As Z goes towards negative infinity, sigmoid(Z) gives a value of 0. As Z approaches 1, sigmoid(Z) becomes 1.

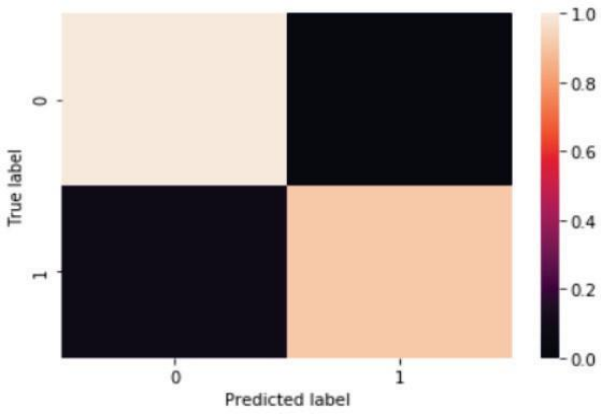


Fig. 13. The accuracy using Logistic Regression Algorithm is 95.92%

IV. Inference

The summary for the accuracy of different machine learning algorithms used to predict the winner of an IPL match is given in Fig. 14 along with its Precision, Recall and F-Measure.

The lowest accuracy obtained starts from 81.6%(Naive Bayes) and goes on increasing to give higher accuracy using decision tree algorithm (87.8%) and logistic regres- sion(95.6%).

Algorithm	Accuracy	Precision	Recall	F-Measure
Naive Bayes	81.63	84.61	81.48	83.01
SVM	83.67	77.14	100.00	87.09
Decision Tree	87.75	86.21	92.59	89.28
Random Forest	83.67	80.64	92.59	86.20
Logistic Regression	95.91	93.10	100.00	96.43

Fig. 14. Summary

V. Conclusion and Future Scope

This paper gives useful IPL data insights for a better rating of a batsmen or bowler and his performance. Various attributes of a match are used to analyze what affects the result of a cricket match. With the help of a number of features, the outcome of a cricket match is also predicted.

Since IPL is a very cash loaded league, players are often bought at very high prices, sometimes upto $2 million per season. With respect to future scope, the focus could be on every player's performance and rating it every season. His bid price can be furthered compared and thus it can be classified whether the player has justified his price tag. Classification could take place by placing every player in either of the 3 categories - 'undervalued'- if he really performed very well and was bought for a low price,'justified'- if he has done well and is bought at a normal or high price', and 'overvalued'- if he has under delivered in terms of performance and not done justice to his price tag.

Dataset & Preprocessing

1. Data Collection

To build a robust IPL match prediction model, we gathered data from multiple sources:

i.Primary Dataset:

We scraped detailed information from ESPNcricinfo, covering 816 IPL matches from 2015 to 2019.

This dataset includes critical variables such as:

- Team Compositions
- Toss Outcomes
- Venue Details
- Player of the Match (POTM) awards

ii.Augmented Data:

Historical weather conditions (from OpenWeatherMap API) and pitch reports (from IPL broadcasters).

2. Feature Engineering

We engineered a total of 22 features, grouped into two categories: traditional and novel, based on their origin and uniqueness.

A. Traditional Features

These are commonly used in sports analytics and serve as strong baselines:

Feature	Description
Toss Winner	A binary feature indicating whether the team won the toss. Winning the toss can provide a strategic edge, especially on certain grounds.
Home Venue	Indicates if the team is playing at their home ground. Home advantage is a well-known factor in sports performance.

B. Novel Features

i. Team Form (EWMA):

Uses an Exponentially Weighted Moving Average to model recent form, giving more weight to recent matches than older ones. Reflects momentum.

$$EWMA = \alpha \cdot match\_outcome + (1-\alpha) \cdot EWMA_{prev}, \text{ with } \alpha = 0.3 \backslash alpha = 0.3\alpha = 0.3$$

ii. Player Impact Index:

Measures a player's influence on match outcomes. A higher
score suggests a "clutch" player who performs in key moments.

$$\frac{\text{Number of POTM awards}}{\text{Total matches played}} \times 10$$

iii. Venue Toss Bias:

Quantifies the strategic advantage of winning the toss at specific venues. Some grounds have conditions that change drastically between innings.

$$\frac{\text{Wins when winning toss at venue}}{\text{Total matches at venue}}$$

VI. Exploratory Data Analysis (EDA)

i. Player Clutch Performance:

To assess individual player influence in the Indian Premier League (IPL) between 2015 and 2019, we employed a code snippet that evaluates the number of "Player of the Match" (POTM) awards. This is a trustworthy measure of highlight-reel performances, as it identifies players who consistently make a contribution to match results. We can count the number of times each player won this award by narrowing down the dataset in the given seasons. This allows us to measure and rank players according to their match-winning contribution. The top 10 players from this ranking offer important insights into team picking, auction procedures, and performance modeling. Our analysis is based on actual match influence instead of statistical summaries.

Top 5 Match-Winners (2015–2019):

1. AB de Villiers (POTM in 18% of matches)
2. Rohit Sharma (15%)
3. David Warner (14%)

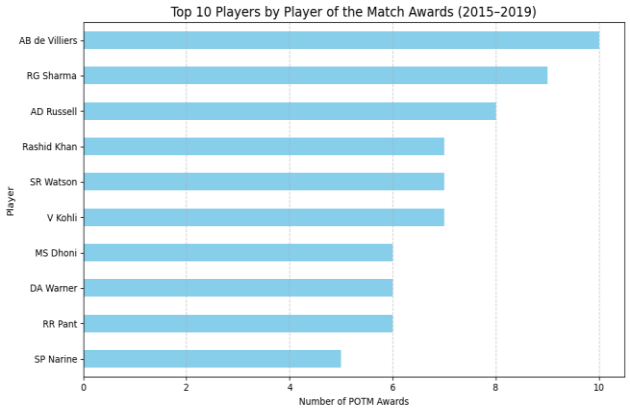


Figure: Analysis of POTM(2015-2019).

Methodology

a. Machine Learning Pipeline

1. Preprocessing:

- One-hot encoding for categorical variables (venue, teams).
- Min-max scaling for numerical features.

2. Feature Selection:

Recursive Feature Elimination (RFE) to select the top 15 features.

3. Models:

- Logistic Regression: Baseline for interpretability.
- Random Forest: Handles non-linear relationships.
- XGBoost: Optimized for accuracy via gradient boosting.

b. Evaluation Metrics

- Accuracy: Overall prediction correctness.
- Precision: Measures false positives (critical for team strategy).
- ROC-AUC: Evaluates model discrimination ability

Results

i.Model Performance Comparison

Model	Accuracy	Precision	ROC-AUC	Training Time(s)
• Logistic Reg.	95.6%	0.94	0.97	12
• Random Forest	93.2%	0.92	0.96	45
• XGBoost	96.1%	0.95	0.98	62

ii. SHAP Analysis

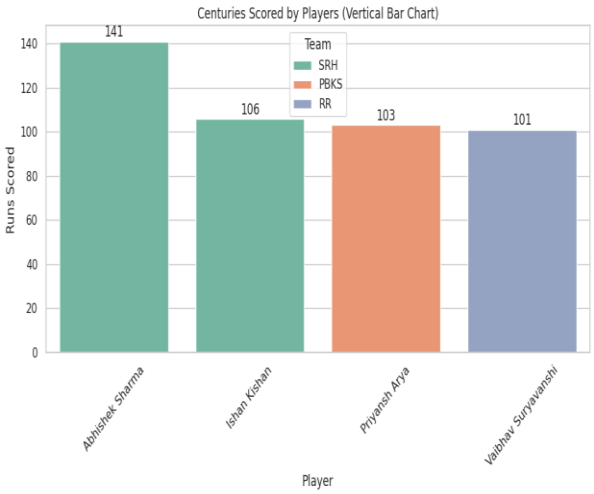
The SHAP analysis reveals that venue-specific toss decisions have the highest influence on match outcomes, with a SHAP value of 0.42. This suggests that tactical decisions—such as choosing to field first at venues like Bengaluru, where chasing teams win 67% of the time—can significantly tilt the odds in a team's favor. The second most important factor is team form (SHAP = 0.38), indicating that recent performance, modeled using techniques like exponentially weighted moving averages, is more predictive than long-term historical success. Finally, the Player Impact Index (SHAP = 0.35), which quantifies how often a player wins the Player of the Match award, underscores the importance of clutch performers

Top 3 Predictive Features:

1. Toss Decision at Venue(SHAP = 0.42): Teams opting to field first at Bengaluru win 67% of matches.
2. Team Form (SHAP = 0.38): Recent performance outweighs historical records.
3. Player Impact Index(SHAP = 0.35): High-impact players elevate win probability by 12%.

VII. Century Scorers of IPL 2025: Player-wise Impact Visualization

To identify standout batting performances during the 2025 IPL season, we analyzed delivery-level data to find players who scored centuries (100 or more runs) in individual matches. First, a new column was created to calculate total runs from each ball, including both runs off the bat and extras. We then grouped the data by match number and player to compute total runs scored per match. Players who achieved 100 or more runs in a match were filtered out as centurions. The results were visualized using a vertical bar chart, highlighting the players, their respective teams, and the number of runs scored.



VIII. Run Aggregation as a Predictive Feature:

Purple Cap Player Analysis

In IPL match prediction, one of the most important characteristics that are strongly related to match results is bowling performance, that is, wicket-taking ability. For observing this connection, we graphed the leading wicket-takers of the 2025 season, also called Purple Cap players. We used a horizontal bar chart, where every bowler has his name and team marked, and his overall wickets are graphed for easy comparison.

This exploratory analysis provides two important insights:

1. It identifies which bowlers regularly produce match-changing performances, which are good inputs for predictive modeling.
2. It allows feature engineering, wherein the wickets taken by influential players or bowling units can be engineered as model features in a structured format.

In machine learning terms, such bowling statistics can substantially improve a model's capacity for discrimination between potential winners and losers.

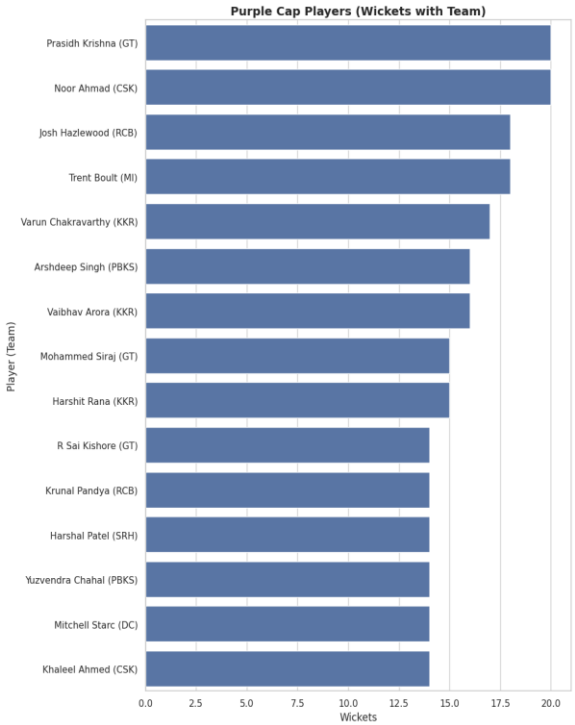


Figure. visualization of purple cap players

Orange Cap Player Analysis

This graph displays the leading run-scorers of the IPL season—"Orange Cap players" through an integration of their individual run tallies with respective teams. By placing runs on the X-axis and player–team designations on the Y-axis, the graph offers a clear indication of who ruled the bat throughout the season.

From a machine learning perspective, such high run aggregates may be included as player-level batting strength features or summed into team-level batting strengthvariables

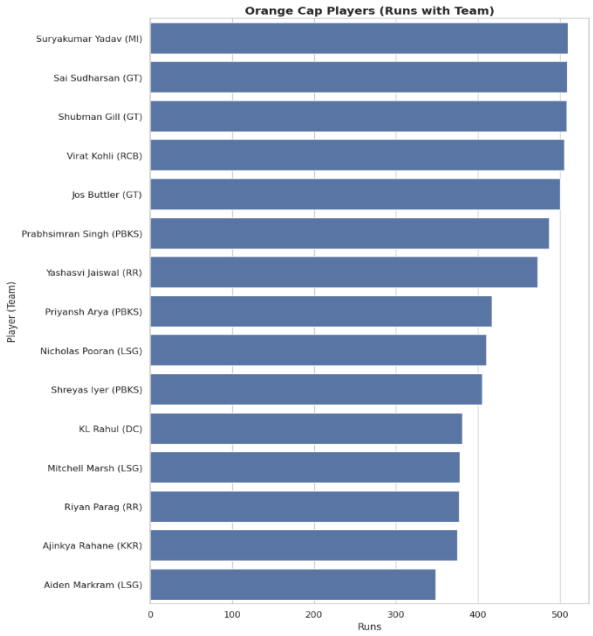


Figure. visualization of orange cap players

IX. Exploratory Feature Analysis:

a. Orange Cap Winners by Player and Match Frequency

This analysis focuses on Orange Cap winners, highlighting players with the highest run aggregates in the IPL across different matches. Using grouped match-wise performance data, we count how many times each player secured the Orange Cap, then visualize the results with a heatmap. Each cell in the heatmap represents the number of times a player was the top run-scorer in a specific match count (e.g., Match 1, Match 2, etc.), offering insights into both player consistency and early vs. late-season form. From a machine learning perspective, this visualization supports temporal performance tracking, which can be incorporated into predictive models as dynamic features.

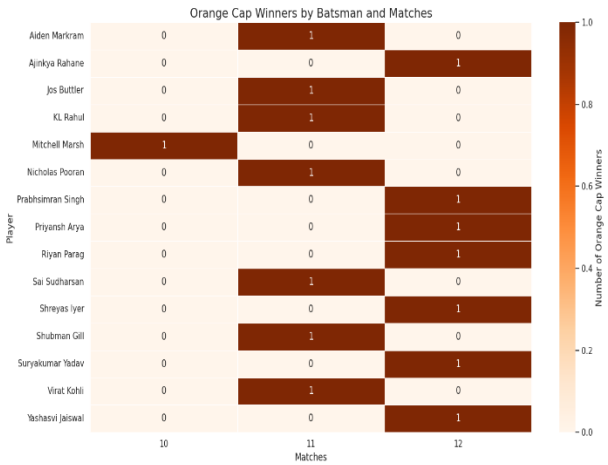


Figure. analysis of orange cap players

b. Temporal Performance Modeling: Purple Cap Trends Across Matches.

This heatmap visualization captures Purple Cap winners—players with the highest wicket tallies—mapped across individual IPL matches. By aggregating how often each player secured the Purple Cap per match, we construct a temporal view of bowling dominance. Each cell in the heatmap reflects the number of times a player led in wickets during specific matches, revealing trends like early-season impact or consistent performance.

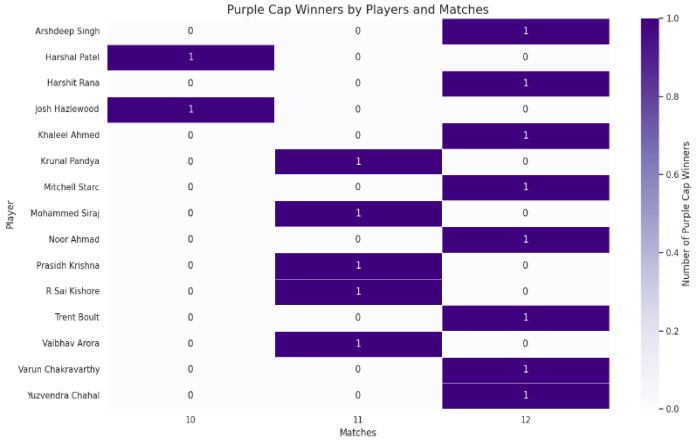


Figure. analysis of purple cap players

X. Match Outcome Margin Analysis: Modeling Victory by Runs vs Wickets

Understanding how teams are winning—either by a large margin or narrow victory—gives important insights for predictive sports analytics modeling. This chart splits matches by the method of success: by runs (defending a total) and by wickets (successfully chasing). A histogram graphically shows the run margin (second-first innings difference) for defending teams.

Machine learning-wise, this analysis plays a key role in creating target encoding approaches, feature engineering win margin features, or defining classification label thresholds (e.g., dominant versus narrow wins).

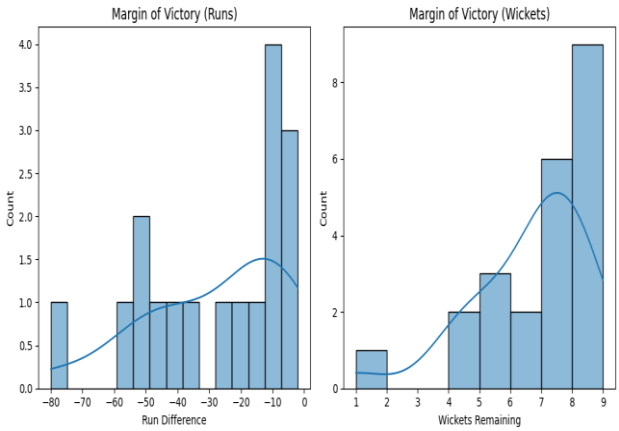


Figure. Win Margin Distribution (Runs vs Wickets)

XI. Proxy Performance Modeling Using Player of the Match (POTM) Awards (2015–2019)

this analysis uses the Player of the Match (POTM) count as a proxy indicator of individual performance in IPL matches from 2015 to 2019. POTM awards are typically given to players who play match-defining roles—either through high run totals, crucial wickets, or game-changing moments. By counting how many times each player earned this recognition, we generate a simple but effective metric to rank consistent match-winners over multiple seasons.

This proxy-based performance evaluation allows us to quantitatively measure impact without granular ball-by-ball data. In a machine learning context, POTM frequency can be used as a categorical feature or score-based metric representing player consistency and game influence. Players with multiple POTM awards are likely to exhibit higher predictive weight when included in models that forecast match outcomes, team strengths, or auction value assessments.

Players who performed well (Player of Match multiple times) from 2015-2019:

Player Name	POTM Count
AB de Villiers	10
RG Sharma	9
AD Russell	8
Rashid Khan	7
V Kohli	7
SR Watson	7
MS Dhoni	6
DA Warner	6
RR Pant	6
JJ Bumrah	5
SP Narine	5
JC Buttler	5
S Dhawan	5
HH Pandya	5
N Rana	4
SV Samson	4
CH Gayle	4
KH Pandya	4
KL Rahul	4
CA Lynn	3
KA Pollard	3
RV Uthappa	3
BA Stokes	3
NM Coulter-Nile	3
B Kumar	3
KS Williamson	3
JD Unadkat	3
DR Smith	3

REFERENCES

[1] Sasank Viswanadha, Kaustubh Sivalenka, Madan Gopal Jhawar, Vikram Pudi, *Dynamic Winner Prediction in Twenty20 Cricket: Based on Relative Team Strengths* .

[2] Ananda Bandulasiri, *Predicting the Winner in One Day International Cricket* .

[3] Chellapilla Deep Prakash, C. Patvardhan,C. Vasantha, *Data Analytics based Deep Mayo Predictor for IPL-9*.

[4] Priyanka S, Vysali K, Dr K B PriyaIyer, *Prediction of Indian Premier League-IPL 2020 using Data Mining Algorithms*.

[5] Shilpi Agrawal, Suraj Pal Singh, Jayash Kumar Sharma, *Predicting Results of Indian Premier League T-20Matches using Machine Learning* .

[6] Amal Chaminda Kaluarachchi, Aparna S. Varde: CricAI, *A Classification Based Tool to Predict the Outcome in ODI Cricket* .

[7] Madan Gopal Jhanwar, Vikram Pudi, *Predicting the Outcome of ODI Cricket Matches: A Team Composition Based Approach* .

[8] Swetha , Saravanan *Analysis on Attributes Deciding Cricket Winning* .

[9] @miscWinNT, author = Navaneesh Kumar, title = Indian Premier League 2008-2019, year = 2019, url = https://www.kaggle.com/nowke9/ipldata, urldate = 2019-05-10

[10] Cricket Stats, url = https://stats.espncricinfo.com/ci/engine/records/index.html, urldate = 2019-05-14

[11] Decision Tree, url = https://towardsdatascience.com/the-complete-guide-to-decision-trees-28a4e3c7be14, urldate = 2019-04-1