**IPL Win Probability Predictor**


*Executive summary*


**ADTA 5340 Section(s) 501 and IPAC 4340 Section(s) 501**

**Discovery and Learning with Big Data**


*by*


**Group-3**

Venkata Sai Kumar Rachagundla

Rahul Varma Muppala

Likitha Reddy Mettu

Pavan Yerukula

*Submitted to*

Dr. Tony Fantasia

*May 5, 2024*

**Topic:** - -

Forecast the probability of winning in the IPL matches. This project uses machine learning models with the data from IPL history 2008-2019 to implement advanced analytics to offer real-time insights during live matches.

**Problem statement: -**

The dynamics and uncertainties of T20 cricket, especially the IPL, create a unique challenge that most conventional models fail. The accuracy and reliability of a prediction are refined by accounting for the influencing factors, such as player performance and the history of the team, the match's venue, and the weather conditions. The model does not only serve as a source of prediction for team strategists but also aids sports analyst and betting agencies to get sound predictions.

**Approach**

- **Data Collection:** We utilize two uncleaned detailed datasets "matches.xlsx" and "deliveries.xlsx" from Kaggle which contain the complete match results, player scores, and over-by-over data of each match.

- **Exploratory Data Analysis (EDA):** We performed a scientific method of analyzing data sets to understand patterns and trends of data sets.

  We use this EDA to analyze:

  - ➤ Performance differences between each year's IPL teams.
  - ➤ Differences toss choices make to match outcomes.
  - ➤ Ranking players by their wins for the match
  - ➤ Distribution statistics of the win margin.

- **Model Development:** We then run the collected data through several machine learning models to predict which team will win the match. We run multiple ML models and validate their score, accuracy, precision, and recall.

- **Models Used**

  - ➢ **Logistic Regression:** This model gives a baseline for performance metrics.

  - ➢ **Random Forest:** Random forest provides insights into the importance of different types of features and helps to improve its accuracy due to its ensemble approach.

- **Validation:** We divide our available IPL data into historical and current data to train the model specifically to not overly fit or underfit the model.

**Key Findings**

Important findings of the predictive model involve a preliminary accuracy rate near 75% as compared to traditional methods used to predict IPL match outcomes. Furthermore, the predictive model emphasizes the crucial contributions of venue, team matchups, player performance and team composition has to match outcomes and in-turn, a team's strategy. Our results clearly indicate the dynamic nature of the IPL match outcomes dependent not on fortuitousness, but, relatively, on a dynamic understanding of player form and performance on team outcomes in T20 matches.

**Conclusions and recommendations**

The IPL Win Probability Predictor proposed in the study can be used to benefit different stakeholders in the context of Indian professional cricket. The interested parties include: – Team management: The model findings should inform decisions made about players and their playing roles and strategies.

➢ **Sports Media and analysts:** The information from the predictions generated by the model can be adopted to enhance post and pre-game publicity and player analysis.

➢ **Betting and gambling markets interested in the tournament:** The study's IPL predictor outcomes can be used to modify the odds set on betting end markets creating additional interest.

Following our developments, in the future, it is paramount to include live data input to the predictor to remain  relevant to the complex sphere of professional cricketers.