



Predictive Modeling of IPL First Innings Final Scores

Rohit Bhamidipati



Context

The Indian Premier League (IPL) is a premier cricket league that attracts global attention. Predicting the final scores in the first innings is crucial for strategic decision-making and enhances the viewing experience. This project aims to leverage historical data to forecast these scores accurately



Problem Statement

Accurately predicting the final score of the first innings in IPL matches presents a significant analytical challenge. The ability to make such predictions would be highly beneficial for team strategies, betting markets, and fan engagement.



Results Overview

The predictive models were evaluated, and the results indicate that:

The most correlated features with the final score were: average runs per ball in the last 3 matches, average runs per wicket in the last 3 matches, current number of wickets fallen.

The Random Forest model significantly outperformed other models with an MAE of 0.71, suggesting its robustness in predicting the final scores.



Criteria for Success

The project's success will be evaluated based on the Mean Absolute Error (MAE) of the predictive model, with the goal to outperform the baseline model's MAE and provide actionable insights for stakeholders.

The MAE in this project will have the interpretation of the difference between the predicted final score and the actual final score.



Data Sources

Comprehensive ball-by-ball match data from 2008 to 2022 was used, including detailed statistics and match outcomes. This rich dataset provides the foundation for our predictive modeling efforts.

[IPL 2008 to 2022 All Match Data](#)



Data Wrangling and Feature Engineering

Data cleaning and preparation were performed to ensure quality inputs for the model. Features were engineered to encapsulate important aspects of the game such as:

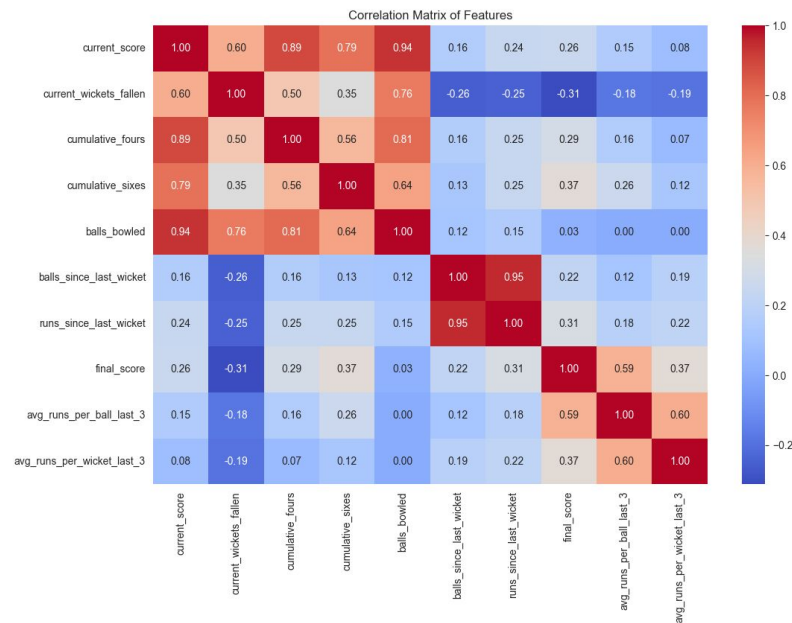
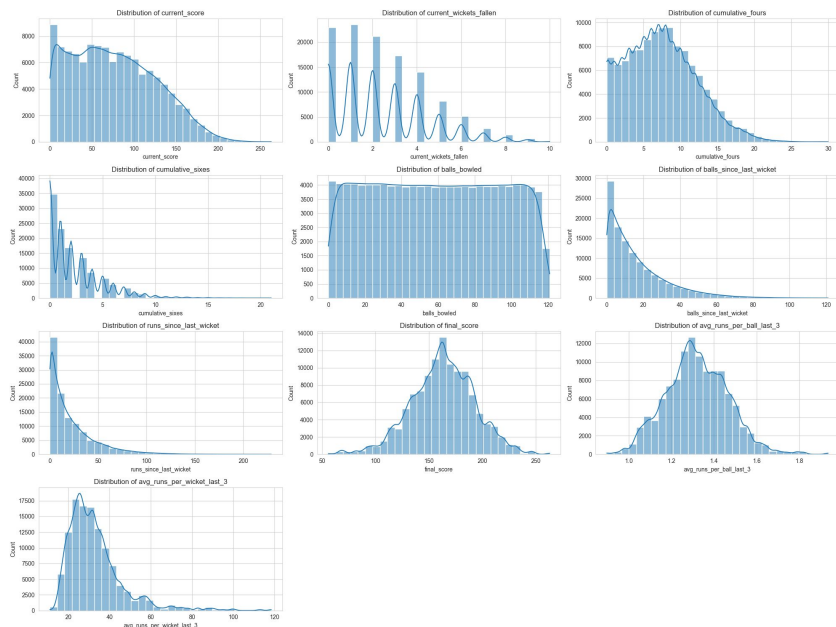
- Match aggregate data: Current score, current no. of wickets, etc.
- Match informational data: Match venue
- Team form data: Runs per ball in the last 3 matches, runs per wicket in the last 3 matches



Exploratory Data Analysis (EDA)

An initial exploration of the data was conducted to understand the distributions and relationships between different variables. This analysis included examining the correlation between current scores, wickets fallen, and the final scores of the innings.

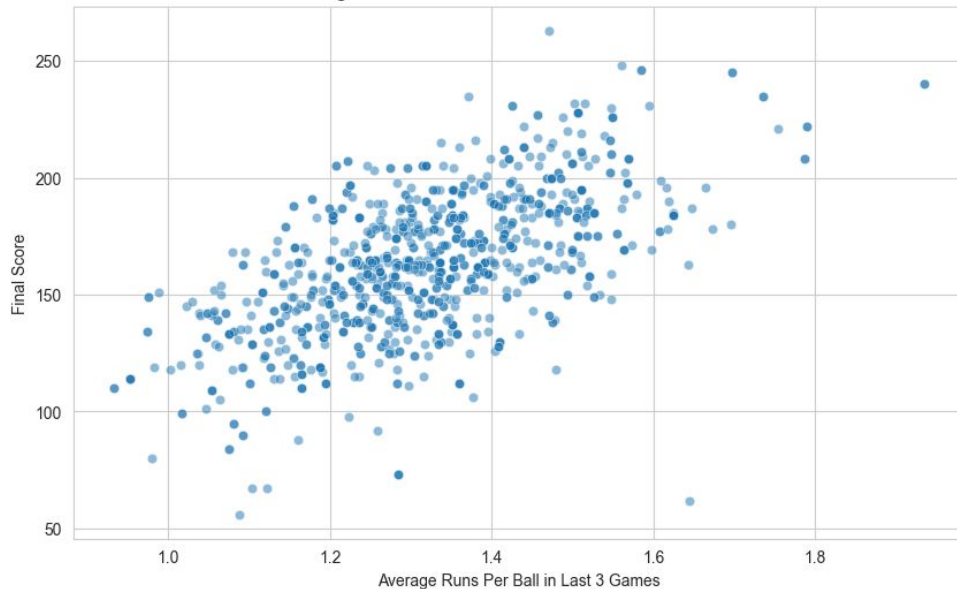
Exploratory Data Analysis (EDA)



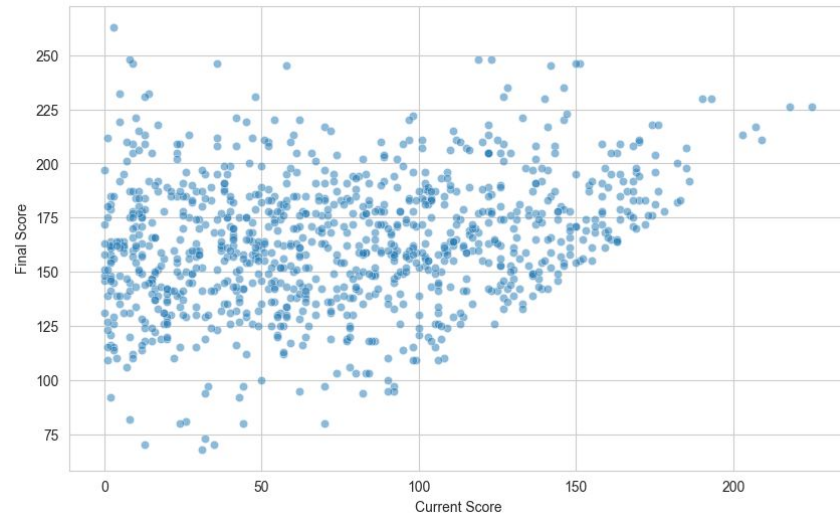


Exploratory Data Analysis (EDA)

Average Runs Per Ball in Last 3 Games vs Final Score



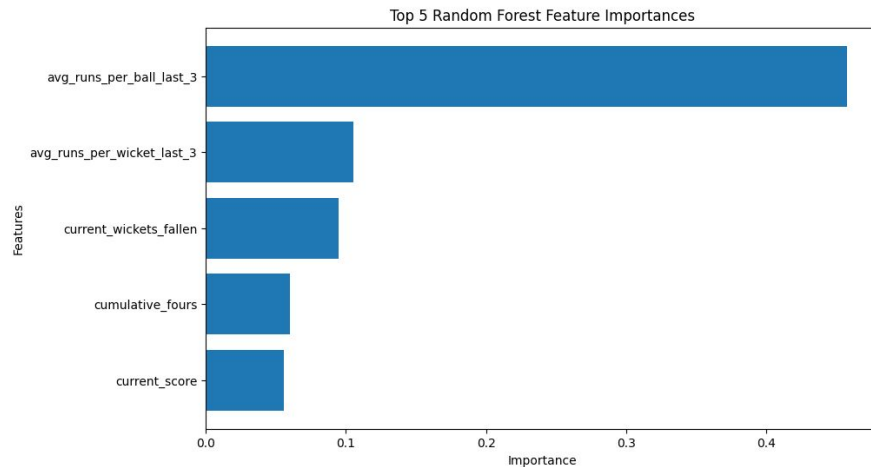
Current Score vs Final Score





Feature Importances

The Random Forest model identified several key features that impact the final score prediction. These include the current score, wickets fallen, and historical performance metrics.





Future Work

Future improvements will focus on integrating real-time data and exploring advanced modeling techniques.

Further refinement of the feature set and model parameters is also planned to enhance prediction accuracy.



Conclusion

The project has successfully developed a model that predicts the final score of the first innings in IPL matches with high accuracy. These predictions can provide valuable insights for teams and other stakeholders involved in the IPL.