Project Report: Predictive Modeling of IPL First Innings Final Scores

Problem Statement

The Indian Premier League (IPL) is a dynamic and influential cricket league where the prediction of match outcomes can greatly enhance the strategic and analytical aspects of the game. This project endeavors to develop a predictive model that can estimate the final scores of the first innings in IPL matches based on comprehensive ball-by-ball match data, contributing to a deeper understanding and appreciation of the game's nuances for teams, analysts, and fans alike.

Approach

A robust dataset spanning from 2008 to 2022 was utilized, encompassing detailed ball-by-ball match data and match outcomes. The methodology involved rigorous data preprocessing, sophisticated feature engineering to extract meaningful insights from the data, and exploratory data analysis to uncover underlying patterns and relationships. Machine learning models, including Linear Regression, Random Forest, and XGBoost, were trained and evaluated using Mean Absolute Error (MAE) to measure prediction accuracy.

Summary and Conclusions:

The project culminated in the successful development of a predictive model with the ability to estimate the final first innings scores in IPL matches. After extensive data analysis and model evaluation, the Random Forest and XGBoost models demonstrated high predictive accuracy, outperforming the baseline model. The project findings could provide IPL teams and stakeholders with actionable insights, enabling more informed decision-making and strategy formulation. Ongoing collection and integration of new match data will continuously refine the model's predictive power, ensuring its relevance and utility in the ever-evolving cricket landscape.