

The Bank Is Open: AI in Sports Gambling

Students: Gaurav A Gopalakrishnan (SRN: PES1PG25CS027), Mani Deepak A (SRN: PES1PG25CS037)

Course: Stochastic Models and Machine Learning

Problem Statement

The objective of this project is to design a machine learning system capable of predicting the total number of points scored in an NBA game using historical team-level performance data. This prediction directly relates to the Over/Under betting market, which is a widely used indicator in sports gambling. Accurate estimation of total points can assist in understanding how data-driven models compare with sportsbook predictions.

Approach

The problem is formulated as a supervised regression task. Initially, a Linear Regression model was implemented to serve as a baseline and establish a reference performance. Subsequently, a Random Forest Regressor was employed to better capture non-linear relationships and interactions between multiple team performance metrics. Model selection focused on balancing simplicity, interpretability, and predictive accuracy.

Implementation

The dataset consists of historical NBA game statistics at the team level, including rebounds, assists, and turnovers for both home and away teams. The target variable was defined as the total points scored in a game. Data preprocessing included feature selection and splitting the dataset into training and testing sets using an 80-20 ratio to evaluate generalization. All models were implemented using Python with standard machine learning libraries.

Results

Model performance was evaluated using Mean Squared Error (MSE). The Linear Regression model achieved an MSE of 841.26, indicating limited ability to model complex scoring dynamics. In contrast, the Random Forest Regressor significantly improved predictive performance with an MSE of 86.92, demonstrating the effectiveness of ensemble-based non-linear models for sports analytics and betting-related prediction tasks.

Conclusion and Challenges

This project demonstrates the practical application of machine learning techniques to sports gambling analytics. While the results are promising, the primary challenge was the limited size of the dataset, which can affect generalization. Future work could incorporate player-level statistics, recent form indicators, and betting odds data to further improve predictive accuracy and robustness.