Predictive Analysis of Athlete Performance in Basketball

Elia Harb, Mason Jeffers, Tyler Zastrow, Maximus Simon

Professor Ben Labaschin

MGSC 310: Statistical Models in Business

November 14, 2023

1. Introduction & Business Case

This project aims to use historical NBA player data from 1996 to 2022 to predict future player performance. Our objective is to provide actionable insights for team management and player development, utilizing data analysis to inform strategic decisions. The core challenge is to accurately forecast individual player performance in upcoming seasons, thereby enhancing team composition and game strategy. Addressing this challenge presents significant business value, potentially leading to improved game outcomes and financial benefits from optimized player investments.

2. Dataset Overview

Dataset Name: NBA Players

Source: Justinas Cirtautas, Kaggle

Description: Comprehensive dataset encompassing biometric, biographic, and basic box score stats of NBA players from the 1996 to 2022 seasons.

Key Variables: Player biometrics (height, weight), biographical details (age, experience), and performance stats (points per game, rebounds, assists, etc.).

Link: https://www.kaggle.com/datasets/justinas/nba-players-data

3. Predictive Models

Elia Harb: Random Forest Model:

- **Question:** Using a Random Forest model, how accurately can we predict an NBA player's points per game in the upcoming season based on their previous season's points per game, height, weight, and years of experience in the NBA?

Mason Jeffers: Linear Regression

- **Question:** How to predict a players Plus/Minus, using Linear Regression, to create the most efficient Starting 5, offensively and defensively.
- **How to Answer:** I will create a Linear Regression model with the data and have Plus/Minus be the output. I will then train the model to predict the players plus minus. I will then filter the players by position to create the most efficient starting 5.

Tyler Zastrow: Decision Tree Model

- **Question**: Can we predict the likelihood of an NBA player being selected for the All-Star game based on their current season's statistics like points per game, rebounds, assists, and shooting accuracy?
- **How to Answer:** I will create a Decision Tree model using current season statistics as input features. The model will classify players based on the likelihood of being selected for the All-Star game. After training the model with historical data, it can be used to predict current players' All-Star selection chances. The decision tree's structure will help in understanding which factors most strongly influence All-Star selection.

Maximus Simon: Time Series Analysis

- **Question:** Using Time Series Analysis, can we identify and explain seasonal trends in plus/minus points scored per game for players throughout a regular NBA season.

4. Analysis Focus

Target Variable: Future season performance metrics (e.g., points per game, efficiency rating).

Independent Variables: Historical performance data, age, height, weight, years of experience, etc.

5. Expected Outcomes

Anticipated insights into factors influencing player performance.

Recommendations for player development and team strategy.

6. Conclusion

Emphasizing the potential impact of this analysis on strategic decision-making in NBA team management.