May 31, 2024

Carl Riemann Torres

Capstone Two Project Proposal: NBA Score Predictor Model

The NBA offers a rich dataset for predictive modeling. From my dataset searches I found a large set containing the Player Stats of all NBA players for the 2023-2024 season. With this information it is possible to predict future performances which can help businesses such as sport betting sites and fantasy leagues in making data driven decisions.

The primary goal of this project is to create a prediction model with good performance metrics that predicts points per game (PTS) and field goal percentage (FG%) based on data available from the 2023-2024 NBA regular season. Given the project goals a model like Linear Regression will be established first and then a more complex model Random Forest Regression will be used to capture non-linear interactions between the variables. For this goal I will use methods like Principal Component Analysis and ANOVA to assess the relationships between the variables, focusing on the variables that most affect PTS, FG%, Match Result. Mean Absolute Error, Mean Squared Error, and R-squared will be used to evaluate model performance.

Challenges expected during this project include applying techniques to handle duplicates, and outside variables such as player injuries, which can impact the model's performance.

Applications for this project are to provide predictive values of player performances, helping fantasy league players or sport bettors in their decisions. Another application is enhancing NBA fans engagement by creating players/teams performance discussions. Gaining hands-on personal experience on predictive modelling for games.