Comparing Models for Forecasting NBA Game Result

Yiming Zhang, Mingqi Zhang, Ruohan You

Abstract:

The objective of this project is to develop a predictive model for forecasting the outcomes of NBA games using data mining and machine learning techniques. The project will focus on analyzing historical NBA game data, player statistics, and team performance to create an accurate model capable of classifying game outcomes. By exploring various algorithms, such as logistic regression, decision trees, and KNN, this study aims to determine which approach provides the highest accuracy in predicting NBA game results.

Keywords:

KNN, Decision Tree, Predictions, NBA, Basketball

Introduction:

The National Basketball Association (NBA), composed of 30 teams, stands as one of the most popular professional basketball leagues worldwide, having millions of fans from everywhere. Fans are either watching live broadcasts or attending the games live to support their favorite teams. Many people want to know which team is more likely to win the matchup before the game. In this project, we aim to predict the outcomes of NBA games by comparing the results of two models: K-Nearest Neighbors (KNN) and Decision Tree. By examining the most recent matchups between teams, we will compare the predictive accuracy of each model, providing insights into their effectiveness for forecasting game results. We will also use graphs to visualize the differences in terms of results of different models.

Previous Work:

https://digitalcommons.bryant.edu/cgi/viewcontent.cgi?article=1000&context=honors_data_science

https://www.kaggle.com/datasets/wyattowalsh/basketball

https://github.com/luke-lite/NBA-Prediction-Modeling/blob/main/model_builder.py

https://www.kaggle.com/code/bbusath5/nba-shot-prediction-knn-model

Methods:

We will be using KNN and decision trees for this project. We compare the results we get from each algorithm to see which one predicts a better result. For the KNN approach, we will be using the dataset from Kaggle to get the most recent 5 matchups in the NBA, which helps us to predict the upcoming matchup. We will also get each team's stats, for example, free throw percentage, 3-pointer percentage, field goal percentage, etc., and the most important part of this is to predict the score each team gets. We will use normalization to get all these stats normalized to compare the differences between teams. We will also consider if the venue (home vs. away) affects the score of each team. In terms of the decision tree approach, we will use the same dataset and divide all of the data into smaller subsets and eventually get the win or loss for each team. By analyzing free throw percentage, 3-pointer percentage, field goal percentage, etc., we can get a deeper look into each team's performance. And eventually we compare the results from 2 models to see how these 2 models generate predictions differently.

Contribution:

Yiming: final essay and data collection

Mingqi: Introduction and use of data screening and modeling .

RuohanYou will finish the code part. He will use ChatGPT to help understand the techniques for analyzing the data.