Capstone Two - Project Proposal

How can I use NBA advanced statistics to predict, with 75% accuracy, which players will make the All-Star team each season?

Each year, 24 players from the NBA are selected as All-Stars. The way All-Stars are selected has changed slightly over the years, but today's selections are made by a combination of votes from fans (50%), players (25%) and sports medai (25%). Many people consider individual player stats, team success, and an "eye test" when judging All-Star selections. Can individual player data, as well as team wins and home game attendance allow us to accurately predict whether a player will make the All-Star Team with 75% accuracy?

Accurately predicting All-Star Team selection is valuable for NBA players, sponsors, and teams. All of those parties are motivate to know which statistics are most valuable for All-Star selections, because once a player is crowned an All-Star their name recognition and ability to sell products and tickets will greatly increase. In addition, sports betting companies and fans will benefit from this information.

This project will use data from Sumitro Datta's Kaggle data set which original came from Basketball-Reference. I will use three tables from his data set to get the following 55 features. A more detailed description of each stat can be found in this spreadsheet.

- Season ID
- Season
- Player ID
- Player Name
- Position
- Age
- Experience
- Team
- Games
- Minutes Played
- Player Efficiency Rating
- True Shooting %
- 3-point attempt rate
- Free throw rate
- Offensive Rebound %
- Defensive Rebound %
- Total Rebound %
- Assist %
- Steal %
- Block %
- Turnover %
- Usage %
- Offensive Win Share
- Defensive Win Share
- Win Share
- Win Share per 48 mins
- Offensive box +/-
- Defensive box +/-
- Box +/-
- Value over replacement player
- Field Goals
- Field Goal Attempts
- Field Goal %
- 3 pointers
- 3 point attempts
- 3 point %
- 2 pointers
- 2 point attempts
- 2 point %
- Effective Field Goal %
- Free Throws
- Free Throw Attempts
- Free Throw %
- Offensive Rebounds
- Defensive Rebounds
- Total Rebounds
- Assists
- Steals
- Blocks
- Turnovers
- Personal Fouls
- Points
- All-Star (T or F)
- Home Game Attendance
- Team Win %
- Season's Average Attendance

I will use data from 1979 and onwards, because, before that time, there were multiple professional basketball leagues in the U.S and there was no three point line. The majority of these features are found in the "Advanced.csv" and "PlayerTotals.csv" files of the data set. However, I will need to use a left-outer join to add All-Star team selection information from the "End of Season Teams.csv" file. I will also need to create a Team Win % column from the "Team Summaries.csv" and use another left-outer join to add Team Win % and attendance to my table.

For this project, I will deliver my code, a written summary, and a slide deck.