

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 media (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 motivated 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 originally came from Basketball-Reference. I will use three tables from his data set to get the following 34 features. A more detailed description of each stat can be found in this [spreadsheet](#).

Season ID	Team	Offensive Rebound %	Usage %	Box plus/minus
Season	Games	Defensive Rebound %	Offensive win share	Value over replacement player
Player ID	Minutes Played	Total Rebound %	Defensive win share	All-Star (True/ False)
Player Name	Player Efficiency Rating	Assist %	Win share	All-Star Team # (1-3)
Position	True Shooting %	Steal %	Win share per 48 mins	Attendance per home game
Age	3-point attempt rate	Block %	Offensive box +/-	Team win %
Experience	Free throw rate	Turnover %	Defensive box +/-	

I will use data from 1977 and onwards, because, before that time, there were multiple professional basketball leagues in the U.S. The majority of these features are found in the "Advanced.csv" file 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-per-game to my table.

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