

Final Project Topic Selection:

Predicting the 2024 NBA Champion With Machine Learning

Zareb Islam | Aadhil Mubarak Syed | Jaynor Singson | Dylan Tran
ECS 171: Machine Learning | Professor Setareh Rafatirad
Department of Computer Science | University of California, Davis

Project Objective

The game of basketball is dynamic and unpredictable, with continuous evolution in team strategies, player performance, and game statistics. It is indubitable that players as a whole possess a lot more talent today compared to older eras. For instance, in the recent 2024 season, the Golden State Warriors held the 10th seed in the Western Conference despite having 45 wins. To put this into perspective, 45 wins would have awarded them the 4th seed in the previous year's standings. With such talent in the league, traditional methods of predicting season outcomes, based on subjective analysis or simple statistical approaches, often fall short in accuracy. A prime example of this can be found during last year's postseason when the Miami Heat, an 8th seed in the Eastern Conference, defeated many teams that were title favorites on their path to the NBA finals, taking the entire NBA fanbase by surprise. Given such stochasticity, it is tempting to conclude that statistics do not tell the entire story of a team's outcome... or do they? The recent surge in Machine Learning (ML) offers a sophisticated way to handle vast amounts of data and uncover patterns that may not be immediately apparent, providing a more robust method for prediction. In this project, we aim to use Machine Learning to predict which team will be crowned the 2024 NBA Champions. By employing various ML techniques and algorithms, we aim to analyze various features of overall and advanced team performance statistics from historical data to make informed predictions for the outcomes of the current season.

Background Research

The datasets for this project will be sourced from Basketball-Reference (<https://www.basketball-reference.com/>), a website containing "statistics, scores, and history for the NBA, ABA, WNBA, and top European competition." We will make use of the per-game team statistics and the advanced team statistics for every season. Since the game has evolved drastically since its inception in the mid-1900s, we decided that using data from this time would likely make our model less accurate. Therefore, we will be using datasets spanning specifically the last 30 years (from 1993 to 2023). We contemplated various machine learning techniques including multiple linear regression, probability networks, and deep neural networks (DNN). DNNs, capable of handling complex patterns in large datasets, are suitable for the nuanced analysis of sports statistics. On the other hand, regression models, while simpler, can provide quick insights into relationships between variables and outcomes. The final choice between these models will be made based on initial exploratory data analysis and feature relevance to the championship outcomes.

Project Deliverables

The class project deliverables include a comprehensive research paper detailing the problem, literature review, dataset analysis, methodology, model development, and evaluation, concluding with a discussion of the results. All source code will be documented and uploaded to GitHub. Additionally, a basic HTML interface will be developed to demonstrate the model's functionality, accompanied by a recorded video presentation of the demo. These components are designed to showcase the application of machine learning in predicting NBA champions and to facilitate further academic and practical exploration.

Project Timeline

- | | |
|--|---|
| <input type="checkbox"/> April 22: Submit 1-Page Project Topic Selection | <input type="checkbox"/> May 19: Model Training, Evaluation, Deployment |
| <input type="checkbox"/> April 28: Data Extraction & Cleaning | <input type="checkbox"/> June 2: Project Report Rough Draft |
| <input type="checkbox"/> May 3: Exploratory Data Analysis | <input type="checkbox"/> June 6: Project Report Final Draft |
| <input type="checkbox"/> May 5: Mid-Quarter Progress Report | <input type="checkbox"/> June 8: Submit Project Report, Source Code, Demo |
-