Box Score to Bracket: Evaluating the Predictive Power of NBA Regular Season Performance on Playoff Outcomes

**Andrew Scheiner '25, Sid Lamsal '25, Department of Data Analytics,** Dickinson College

Introduction

This project explores how accurately NBA regular season data can predict playoff outcomes using neural networks trained on game-by-game statistics. The goal is to identify patterns in predictive accuracy as the season progresses, uncovering whether certain segments of games better indicate postseason success. By addressing doubts about the regular season's value, this study offers potential insights to enhance sports analytics, inform league policies, and showcase the dynamic applications of machine learning in real-world scenarios.

Data Processing

- We first scraped and collected our data using the requests library in Python (see data sources section).
- Using the *pandas* library in Python, we were able to process scraped box score data for each NBA team (using global and team identifiers) and merge it together by season.
- The visuals were created using *matplotlib*.
- Our neural network was created using the *keras* library in Python.

Results
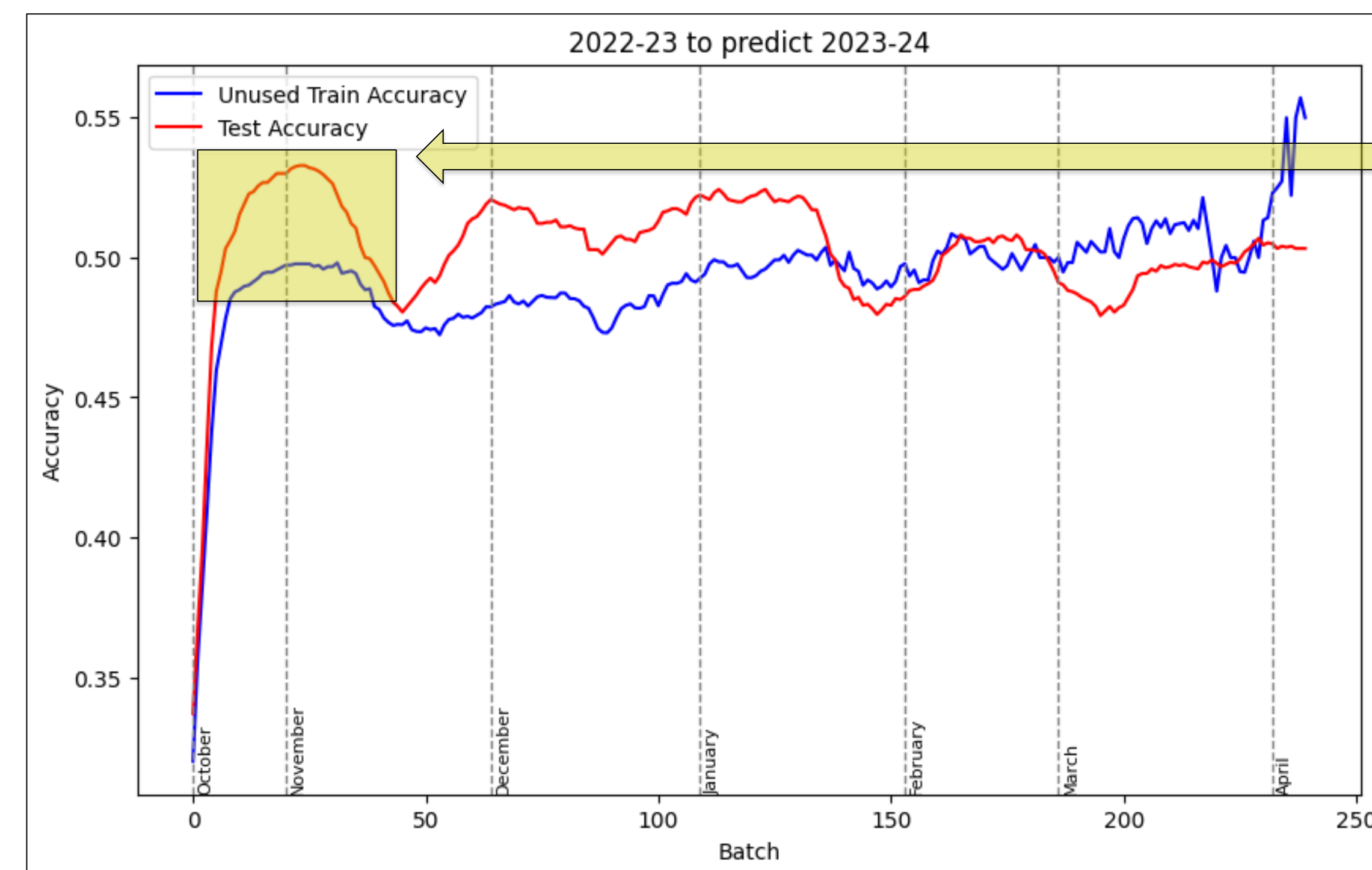
The graph on the left signifies how accurate our model was at predicting playoff performance for the 2023-24 season trained on single-game performance in the 2022-23 (previous) season. Our x-axis represents batch number, where one (1) batch is a group of five (5) NBA games. We find that around batch 20, our model's test accuracy, or ability to predict playoff result for the 2023-24 season, is capped out at around 53%. This means that after approximately 100 games played in the 2023-24 season, which is around the start of November, we can tell a team's correct playoff result at a bit better than a 50/50 rate.
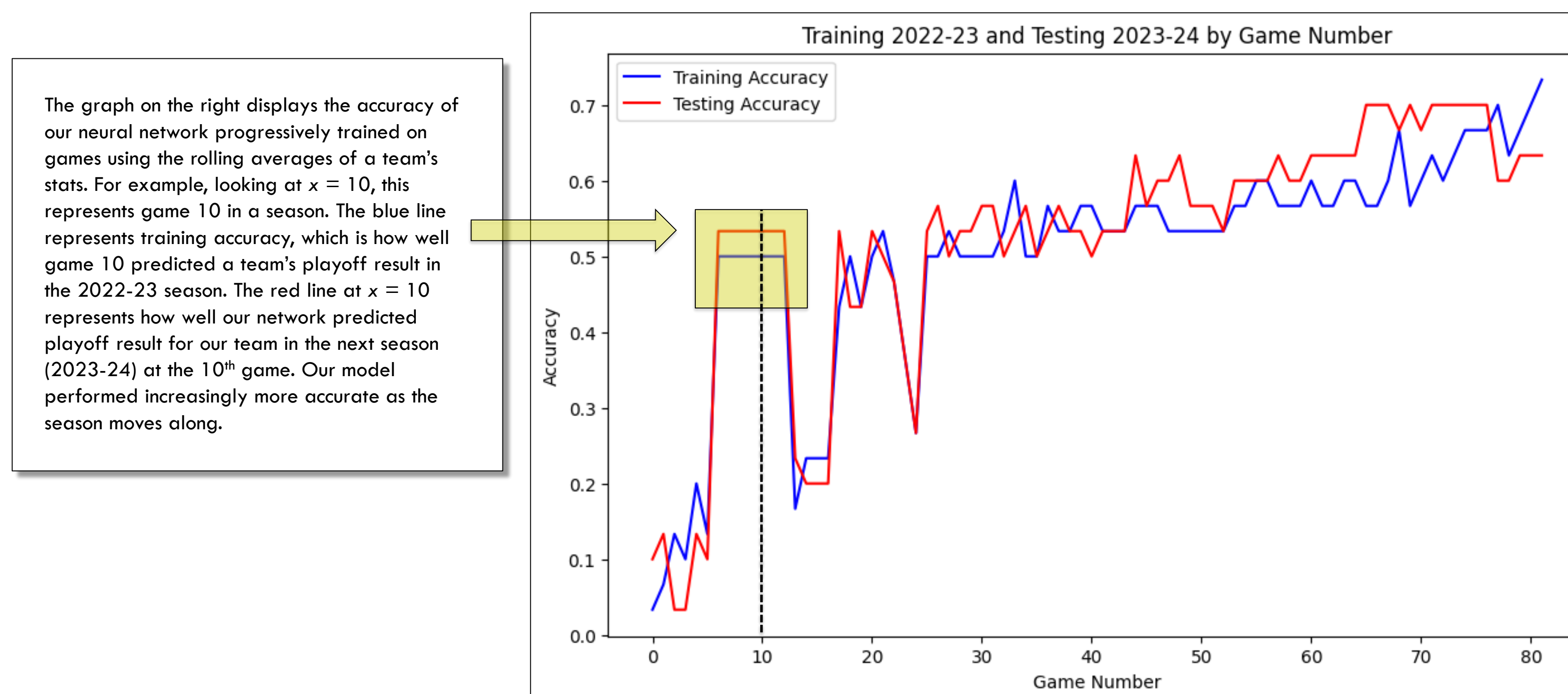
The graph on the right displays the accuracy of our neural network progressively trained on games using the rolling averages of a team's stats. For example, looking at x = 10, this represents game 10 in a season. The blue line represents training accuracy, which is how well game 10 predicted a team's playoff result in the 2022-23 season. The red line at x = 10 represents how well our network predicted playoff result for our team in the next season (2023-24) at the 10th game. Our model performed increasingly more accurate as the season moves along.

Findings

- Our neural network predicted NBA playoff outcomes with an average accuracy of 50%, peaking at 55%.
- Game-by-game training enhances predictive accuracy, following a semi-logarithmic trend.
- Predictive performance improves sharply at first, then plateaus, with accuracy declining slowly after the initial surge.

Implications

- Our model could tell early on what level of performance correlated to a particular playoff finish, but it was not entirely accurate.
- Our predictive margin of error was around +/- 3 games of playoff result – which is the difference between losing or advancing to the next round.
- Rolling average performance over stretches of games will affect prediction ability.

Future Work

- Creating specialized models for each team.
- Experimenting with different model types other than neural networks, hoping to find more accurate models for predicting.

Data Source

- NBA.com's official stats website.
- Used the stats.nba.com API to scrape traditional and advanced box score data.