🏀 Box Score to Bracket

Evaluating the Predictive Power of NBA Regular Season Performance on Playoff Outcomes

> *"Can watching a few games tell you who's making a deep playoff run?"*

📘 Overview

The NBA regular season packs 82 games per team—but how much of it truly predicts what matters most: the playoffs?

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.

We aim to answer:

- Can we predict if a team will **make the playoffs**?

- Can we estimate how far a team will go—**1st round exit, Finals appearance, or NBA Champions**?
- Are **later season games more predictive** than early ones?
- Does **game-by-game** vs **rolling averages** impact prediction quality?

📊 Data Collection

- **Source:** NBA.com (via `requests`, `pandas`, and `nba_api`)
- **Data:**
  - **Box scores (standard + advanced)** for every regular season game (2014-2024)
  - **Playoff outcomes**: number of playoff wins per team
- **Preprocessing:**
  - Aggregated team-level statistics over rolling 5–10 game chunks; engineered playoff outcome labels (missed playoffs, round exits, champions)
  - 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
- Our **visuals** were created using `matplotlib`
- Our **neural network** was created using the `keras` library in Python

🧠 Model Architecture

We use a **neural network classifier** trained to predict playoff success based on regular season data segments.

Core Experiments:

- **Sequential chunks** (mimicking watching games as the season progresses)
- **Random chunks** (simulating selective viewing like a casual fan)
- **Different window sizes** (5 vs 10 games)
- **Raw games vs seasonal averages**

Our goal isn't just classification—it's to explore **how predictive value evolves over time**.

💎 Key Insights

- 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

📚 Credits

- Developed by Andrew Scheiner and Sid Lamsal for **Dickinson College's** DATA400 (**Data Analytics Senior Seminar**)