**Capstone Topic Approval Form**

**Student Name:** Paul Fox

**Student ID:** 001859906

**Capstone Project Name:** Data Analytics Capstone (BHN1)

**Project Topic:** NBA All-Star Player Assessment

**Research Question:** Do an NBA player’s season statistics (points per game, assists, rebounds, shooting efficiency, and team success) predict whether that player will be named to the All-Star team?

**Hypothesis:** Players with high performance metrics in key statistical categories (points per game, assists, rebounds, shooting efficiency, team success) are more likely to be selected for the NBA All-Star team than players with average or below-average stats.

**Context:** NBA All-Star selections can be based on individual performance, team success, and player popularity. While fan voting, media, and coaches all play a role in the selection process, there's ongoing debate about whether the best players are always selected. Built-in bias and subjectivity, coupled with vague and ill-defined selection criteria make this topic a prime target for data analytics. By analyzing historical player statistics and All-Star selections, we can identify traits commonly shared by All-Stars. This can help teams, agents, and players better understand what metrics matter most in earning recognition. Predictive analytics can bring transparency to the selection process and help validate or challenge subjective decisions using concrete data. In the era of advanced stats and machine learning, data-driven insights like this can enhance decision-making and public understanding of the game.

**Data:** NBA player statistics (points, assists, rebounds, shooting percentages, steals, blocks, efficiency rating, win shares, team win percentage, All-Star selection status).

I plan to work primarily from a subset of the [NBA Stats (1947-present) dataset on Kaggle](https://www.kaggle.com/datasets/sumitrodatta/nba-aba-baa-stats), made available by [Sumitro Datta](https://www.kaggle.com/sumitrodatta).

The data exists in the public domain and is freely available for use.

**Data Gathering:** Iwill download the data from Kaggle in CSV format. I will use pandas and Excel to ensure the data I use is clean and normalized as appropriate.

**Data Analytics Tools and Techniques:** I intend to use Random Forest Classification to create a predictive model. I will be using Python and appropriate packages such as pandas and scikit-learn.

**Justification of Tools/Techniques:** Random Forest is well-suited for tabular data with mixed feature types. It handles non-linear relationships, manages feature interactions effectively, and provides built-in feature importance. It’s also robust to overfitting, especially useful for a dataset with many correlated basketball metrics.

**Application Type, if applicable (select one):** n/a

mobile

web

stand-alone

**Programming/Development Language(s), if applicable:** Python, pandas, scikit-learn, jupyter

**Operating System(s)/Platform(s), if applicable:** n/a

**Database Management System, if applicable:** n/a

**Project Outcomes:**

* A trained Random Forest model capable of predicting whether a player should be selected for the NBA All-Star team.
* Model performance metrics showing model efficacy
* Documentation and visualization as applicable

**Projected Project End Date:** 5/20/2025

**Sources:**

* <https://www.kaggle.com/datasets>
* <https://www.nba.com/>

**Human Subjects or Proprietary Information**

Does your project involve the potential use of human subjects? (Y/N): N

Does your project involve the potential use of proprietary company information? (Y/N): N

**STUDENT SIGNATURE**



**­­­­­­­­­­­­­­­­­­­­­­­­­­­­­\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**



**By signing and submitting this form, you acknowledge** that any cost associated with the development and execution of your data analytics solution will be your (the student) responsibility.

**TO BE COMPLETED BY AN INSTRUCTOR**

**The capstone topic is approved by an instructor.**

**INSTRUCTOR’S NAME AND SIGNATURE: **

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**INSTRUCTOR APPROVAL DATE: 5/5/2025**

**Project Compliance with IRB (Y/N): Y**