**University of Toronto**

**Faculty of Applied Science and Engineering**

**MIE368 Analytics in Action**

**Project Proposal**

| **Student Name** | **Student ID** |
| --- | --- |
| Michael Furlano | 1008072899 |
| Mikhael Saikaly | 1008811127 |
| Hryhorii Ovcharenko | 1008776371 |

NBA teams invest millions of dollars each year in draft selections in an attempt to secure their next franchise cornerstone. Despite extensive scouting and evaluation of players worldwide, outcomes remain extremely variable. Many second-round picks or even undrafted players go on to carve out roles as reliable starters or, in rare cases, develop into All-Stars. While on the other hand, highly ranked lottery selections often fail to meet expectations, with an overwhelming 35–40% of lottery picks being out of the league within their first four seasons. This unpredictability in scouting and drafting provides an understanding of the difficulty of projecting long-term success in a player. By introducing objective, data-driven evaluation, analytics can strengthen traditional scouting, improving the assessment of player potential while reducing the risk of poor draft selections and missed opportunities.

A key challenge for NBA teams is predicting which draft prospects will translate their college performance and physical characteristics, such as height, weight, hand size, and overall athleticism, into sustained professional success. To address this issue, we propose to classify prospects into qualitative categories such as Star, Starter, Role Player, or Bust. Success will be defined over the first eight NBA seasons, capturing not only foundational metrics such as minutes per game, Win Shares, and Player Efficiency Rating, but also milestones including All-Star selections and individual awards. Beyond classification, our objective is to compare traditional predictive models with modern machine learning methods to assess whether machine learning models are capable of providing greater accuracy in forecasting draft success.

To address this question, we will integrate data from several different sources. College statistics will be obtained from Sports-Reference’s NCAA basketball database, which provides both traditional box score measures and advanced statistical and performance metrics. Draft history and full career outcomes will be collected from Basketball-Reference, including indicators such as Box Plus-Minus, Win Shares, and Value Over Replacement Player. Finally, Stathead will be utilized, as it enables advanced queries and bulk downloads, allowing us to link players' NCAA careers with their NBA performances.

Our analysis will begin by constructing a labeled dataset from NBA players with at least eight seasons of experience, categorizing them as Superstar, Star, Starter, Role Player, or Bust based on performance thresholds (e.g., minutes played, Win Shares, Player Efficiency Rating), career longevity, and milestones such as All-Star selections and awards. This dataset will allow the model to learn how college-level statistics and performance overall translate into long-term outcomes and will then be applied to current players with fewer than eight seasons to project their likely trajectories. We will begin with exploratory data analysis to examine distributions, correlations, and baseline trends by draft round and position before implementing predictive models. We will implement standard baseline models (e.g., random forests) for interpretability and compare them against a neural network that’s trained on standardized features to capture nonlinear relationships. Model performance will be evaluated using accuracy, precision, recall, and F1 scores, adapted particularly for multi-class classification. Confusion matrices will further detail class-specific accuracy, focusing on distinguishing high-impact players from those unlikely to translate to the NBA. Finally, projections will be evaluated against actual historical outcomes and pre-draft consensus rankings to assess how data-driven methods compare to scouting evaluations.

This project will generate insights into which college performance metrics and player characteristics best predict long-term NBA success. By comparing traditional approaches with machine learning models, we aim to demonstrate whether data-driven approaches trump current draft forecasting. This project provides a practical framework for NBA teams to supplement scouting decisions, reduce draft risks, and identify undervalued prospects through objective analysis.