**NBA Rookie Scouting Report Sentiment Analysis**

Group Leader: Tenzin Nargee

Group Members: Kian Putnam, Michael Hijduk

Course: DS4002

Date: 1/31/25

**Motivation:** The evaluation of NBA rookies during the pre-draft phase plays a crucial role in shaping team strategies and player careers. Scouting reports provide qualitative insights into players’ strengths, weaknesses, and potential, which are all often subject to the personal interpretations of scouts and analysts. Exploring and understanding the sentiment expressed in these reports can reveal underlying biases and trends that influence draft decisions and also reveal the accuracy of these reports.

Although existing research has explored the relationship between scouting reports and player performance, limited studies have applied sentiment analysis to these reports to systematically assess their impact [1]. By leveraging web scraping techniques to collect detailed scouting reports from reputable basketball websites, our project aims to quantify sentiment and correlate it with measurable performance metrics. This approach allows a data-driven perspective on how effective scouting reports are in predicting rookie success in the NBA.

**Goal Statement:** Investigate whether the sentiment expressed in 2023 NBA rookie scouting reports is a significant predictor of players’ performance metrics during their rookie season.

**Research Question:** Do the sentiment scores of the 2023 NBA rookie pre-draft scouting reports correlate with the players’ performance metrics in their rookie season?

**Modelling Approach:** We will begin by collecting pre-draft scouting reports for all 2023 NBA rookies through web scraping from various scouting report sources online. The text data will then undergo preprocessing steps including tokenization, stop-word removal, and lemmatization using Python’s NLTK library [2]. For sentiment analysis, we plan to use VADER because it has previously been used for sports related texts and works well [3]. We will also use a combination of correlation tests to define the relationship between the sentiment of the scouting reports and the rookie’s success in the NBA [4]. Specifically, we will use point biserial correlation to assess the relationship between sentiment scores and binary outcomes such as winning rookie awards, Pearson correlation coefficient to evaluate the relationship between sentiment scores and continuous variables like minutes played (assuming normal distribution), and Spearman's rank correlation for sentiment scores and ordinal variables such as voting for awards, which may not follow a normal distribution. Success in the NBA will be defined by the awards that the rookies received from their rookie year (rookie team awards) and from minutes played in the season and win shares for rookies.

**References:**

[1] C. Lawlor, J. Rookwood, and C. Wright, "Player scouting and recruitment in English men’s professional football: opportunities for research," *J. Qual. Res. Sports Stud.*, vol. 15, no. 1, pp. 57–76, 2021. [Online]. Available:<https://www.academia.edu/60947249/>. [Accessed: Jan. 30, 2025].

[2] NLTK, “Natural Language Toolkit — NLTK 3.4.4 documentation,” \*Nltk.org\*, 2009. [Online]. Available: https://www.nltk.org/. [Accessed: Jan. 29, 2025].

[3] C. J. Hutto and E. Gilbert, “VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text,” in \*Proc. 8th Int. Conf. Weblogs Social Media (ICWSM-14)\*, 2014. [Online]. Available: https://www.researchgate.net/publication/275828927\_VADER\_A\_Parsimonious\_Rule-based\_Model\_for\_Sentiment\_Analysis\_of\_Social\_Media\_Text. [Accessed: Jan. 29, 2025].

[4] R. P. Schumaker, A. T. Jarmoszko, and C. S. Labedz Jr., “Predicting wins and spread in the Premier League using a sentiment analysis of Twitter,” \*Decision Support Syst.\*, vol. 88, pp. 76–84, 2016. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0167923616300835. [Accessed: Jan. 29, 2025].