**Title:** Predicting NBA Success with Combined Psychological and Physical Statistics

**Introduction:** Physical performance attributes are undoubtedly important when it comes to professional basketball. While it is believed that the psychology of an athlete is important in determining future success it has historically been difficult to quantify. Research has been hampered by a lack of data, due in large part to the load that traditional psychological assessments impose on athletes. Access to athletes is typically very limited, further hindering advancements in this area. In addition, traditional psychological and psychometric assessments rely upon self-reporting, which is known to be problematic as candidates can be biased towards providing the answers they think assessors want to hear. Unlike traditional assessments, language psychology builds profiles based on the way people communicate. In particular it focuses on the unconscious use of function words, a group of words that are contextually uninformative but syntactically important. Profiles constructed using language are far less open to conscious manipulation and thus provide a more accurate picture of a candidate, while also negating the need to have them submit themselves for assessment. In this paper we explore how combining language psychology features with physical performance statistics can predict which aspiring college basketballers will be signed by NBA teams using machine learning techniques. We also explore applications of survival analysis techniques for forecasting player career durations.

**Methods:** Using the Linguistic Inquiry and Word Count (LIWC) language psychology framework, we analyzed post-game interview transcripts obtained from ASAP Sports of >1,500 NCAA basketball players prior to their NBA draft. We matched this language sample against college basketball game statistics for each athlete to obtain physical performance measures. Of this sample, approximately half were signed by NBA teams while the other half were unsuccessful. We built separate supervised classification models to predict the probability each player would be signed, and a combined meta model that leveraged the output of both the physical and psychological models. We also built Cox Proportional Hazard survival analysis models to predict career duration probabilities using these same features.

**Results:** Using 10-fold cross-validation we achieved a prediction accuracy of 72% for the psychological model, 83% for the physical model, and 90% for the combined model. We also found that combining physical and psychological features enhanced the predictive power of our survival analysis model with concordance values of > 0.8 achieved.

**Conclusion:** Our research confirms that while physical performance provides the most impact, psychology also plays an important role in determining success in high performance sport. Using language psychology we can quantify the mindset traits that are correlated with success and build models to predict the likelihood that an athlete will be signed by an NBA team. We also demonstrate the combination of physical and psycholgical traits that most strongly influence the career durations of athletes. This work has implications for talent scouting and player development, potentially revolutionizing how teams identify emerging talent. While this research focussed entirely on basketball, it can also be applied to any other high-performance sport.

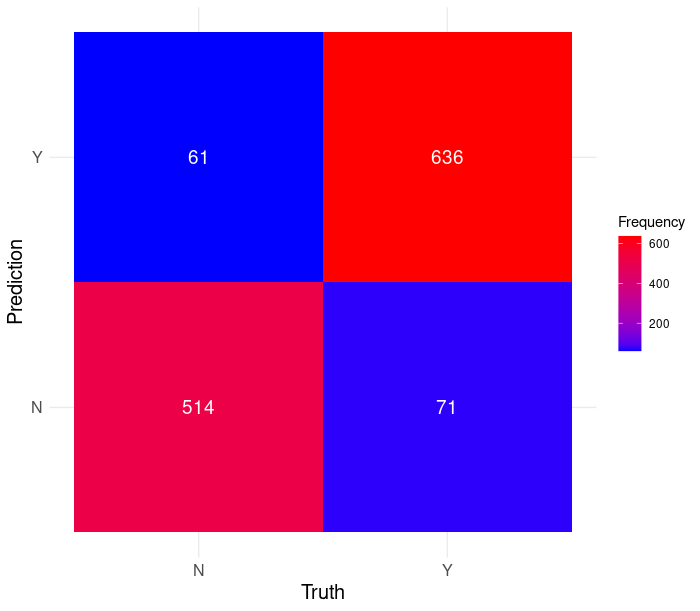


Figure 1 - Confusion matrix for the logistic regression model combining the outputs of the psychological and physical models.

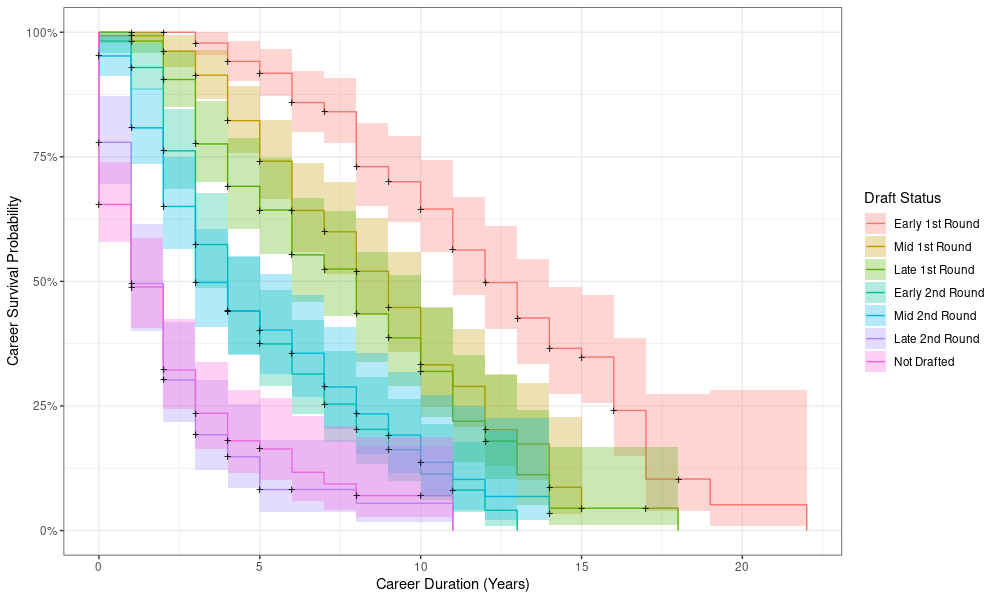


Figure 2 - Survival curves for NBA player career duration probabilities by how they were recruited.