Project Executive Summary

Reggie Bain, Reid Harris, and Tung Nguyen

Recent legislation has changed the landscape of college sports, a multi-billion dollar enterprise with deep roots in American sports culture. With the recent legalization of sports betting in many states and the SCOTUS O'Bannon ruling that allows athletes to be paid through so-called "Name-Image-Likeness (NIL)" deals, evaluating talent and projecting results in college sports is an increasingly interesting problem. By considering both talent accumulation and recent on-field results, our models aim to predict relevant results for sports betting/team construction. In this iteration of the project, our targets are regular season win percentage (using a season level model that we'll call Model 1) and individual game results (with a game by game model we'll call Model 2) in the regular season.

Our dataset comes from a variety of sources including On3, ESPN, 24/7 Sports, The College Football Database, and SportsReference.com. For game by game results, we were able to acquire data back to 2002 reliably. For some of our features such as recruiting rankings and returning usage statistics, reliable data is only available from roughly 2010-2014. Since rules/policies in college football change often, we felt that a 10-year window from 2014 to the present was adequate for making predictions for the future in our season level model. We considered only FBS teams in our study, which are the top-division teams in the NCAA.

Stakeholders for this project include: university athletic departments (for allocating NIL funds), college coaching staffs (for assembling rosters), and sports gamblers. We determined several key performance indicators including: identifying features that determine on-field outcomes, predicting season win percentages accurately, and developing a highly explainable model. We explored a wide array of features, some of which we engineered ourselves. These included more traditional metrics such as passing/rushing yards and touchdowns scored, to more advanced metrics such as ELO rating and the so-called "blue chip ratio". We also engineered features to help describe team talent level and the recent success of a team's coach. Using various techniques, we found that ELO was the most important feature in both our season level and game by game models. Returning talent, coaching success, strength of schedule, and recent winning percentages were also important, particularly in the season level model.

In both approaches, we used a variety of different techniques to build effective models. For Model 1, our season-level approach, we set a baseline model of the "naive forecast" where the season winning percentage was set to the previous season's value. For Model 2, our game-by-game approach, our baseline was simply picking the team with the higher pregame ELO rating as the winner. Although we experimented with a variety of advanced models such as random forests, gradient boosted trees, and long short-term memory neural networks, the best performing models were Linear Regression (Model 1) and Logistic Regression (Model 2). In Model 1, Linear Regression performed ~35.5% better than baseline and in Model 2, Logistic Regression performed nearly ~38% better than the baseline. Cross-validation was performed for both models to help avoid overfitting. We found a dip in performance for both models in 2020, which may be explained by the COVID-19 pandemic causing abbreviated schedules, missing players due to opt out or illness, and altered rules, all of which likely made the season somewhat of an outlier.

A few key insights can be gained from our model. ELO rating was the most important factor in predicting wins. It played a larger role than any metrics related to talent or specific on-field statistics. Although more study is needed, this suggests that sports betters should focus on recent results over perception of talent when making bets. There are obviously still football-specific matchup considerations (player injuries, etc.) to consider. Unsurprisingly, we also found that coaching is important. The lifetime winning percentage of coaches was an important factor in the model. We showed that a given school can have very different winning % based on the coach even in a short time period.

Because the talent metrics (particularly the blue-chip ratio) played a smaller role, it brings into question how enthusiastically teams should invest money into acquiring highly rated high school recruits. Further study is needed to see how predictive recruiting ratings are of college success. Having "talented" players is, obviously, essential to having a successful team. However, the accuracy of the underlying recruiting ratings in predicting success needs further scrutiny. This study suggests that teams should perhaps focus resources more on evaluating players, especially those with college experience from the transfer portal, than in bidding for highly rated recruits. It is unclear if recruiting rankings provide a useful independent metric for evaluating players.