College Football Prediction Model

Kenneth Woodard
University of Tennessee, Knoxville
Department of Computer Science
kwooda13@vols.utk.edu

***Abstract* - The datasets in question are from a website called *College Football Data*. Four datasets were chosen from the site to train a multilayer perceptron. The datasets chosen provided the following for each FBS college football team: roster talent rating, coach rating, and results for each game completed. All data is from the 2022 season prior to bowl season (including conference championship weekend). The goal of this prediction model is to gain an edge in gambling on college football point spreads. The resulting model does not necessarily choose one side of a point spread, but merely predicts the margin of victory or defeat of the home team. The resulting model can be deemed a success when compared to the current point spreads available on FanDuel Sportsbook. However, predicting actual results on the field has proven challenging. The model predicts the correct winner and margin of victory with around a five percent success rate.**

Introduction and Motivation

First, a quick note about betting point spreads for those who are unfamiliar: a point spread is just a line set for a game's final score. Here is an example; Clemson is favored to win by 6.5 against the Volunteers. If you think Tennessee will lose by 6 or less or win the game outright, then you take Tennessee to cover. The typical terminology for a gambler would be Clemson is -6.5 against Tennessee (or Tennessee is +6.5). Four different data sets from the *College Football Data* website (www.collegefootballdata.com) were chosen for the model. These are the chosen datasets available on the website: "Games and results", "FBS team list", "Team talent composite ranking", and "Coaching records and history." Statistics from seasons prior were only factored into the coaching rating. Data from the other datasets pertain only to the current (2022) season. The FBS team list data was used as a base for merging the following datasets into one for the model.

Team talent composite ranking is a simple dataset that quantifies the overall talent on a given roster. Transfer portal moves are included in the talent ratings. It is a composite ranking in nature because it takes each player's recruiting ranking from each website and averages them together. These recruiting ratings are pulled from *Rivals*, *ESPN*, and *247Sports* among others.

The Coaching records and history dataset provides a rating on each head coach based on the "simple rating system." This basically takes a coach's history (as a head coach) and puts a rating on them. This rating weighs recent results heavier than past results (i.e. Jimbo Fisher is one of the lowest rated coaches in the SEC despite having won a National Championship in 2014). The rating weighs all blowout victories as a 24 point victory and all wins of 7 points or less as a 7 point margin of victory. This is to avoid skewing the rating.

The games and results dataset provided three key pieces of data for each completed matchup: home team margin of victory/defeat, which week in the season the matchup occurred, and whether or not the game was a neutral site game. The margin of victory is important for measuring the strength of team in a given week, but teams change throughout the course of a season. Some players get injured while others improve their performance; therefore, teams can improve or deteriorate throughout the course of a season. A neutral site game is a game played anywhere other than a college campus; most are played at NFL venues. It is important for the model to factor this in when processing the game results, so it can properly weigh each result throughout the season. Home-field advantage is more prevalent in college football than other sport in the world. This is because kids that recently came from high school are more susceptible to crowd noise and other obstacles that come from playing on the road when compared to the National Football League. Oddsmakers in Las Vegas value home-field advantage in the NFL as about 3 point advantage, but some college venues are valued up to a touchdown advantage for the home team. This is also in-part due to the stadiums in college (specifically in the SEC) being much larger than NFL venues. For perspective, there are 14 college home venues that are larger, in terms of capacity, than the biggest NFL venue: MetLife Stadium.

The overarching goal of this model is to predict the outcome, specifically margin of victory, of college football games and assist bettors in picking point spreads. Fans of college football and gamblers alike let their eyes deceive them. Humans are susceptible to having knee jerk reactions to unexpected outcomes, but a machine learning model will factor it in appropriately. A typical fan's eyes and short-term memory deceive them while watching these games. In contrast, the model is unbiased in its approach in predicting the outcome of these matchups.

A multilayer perceptron was used because college football data is the furthest thing from linearly separable. It is one of the most volatile sports in the world, making it difficult to predict. The game is played by male students aged 18-22. It is safe to assume that social media, student life, and other distractions can play a massive role in a team's success in a given week. One off-the-field incident can turn a team's victory into defeat.

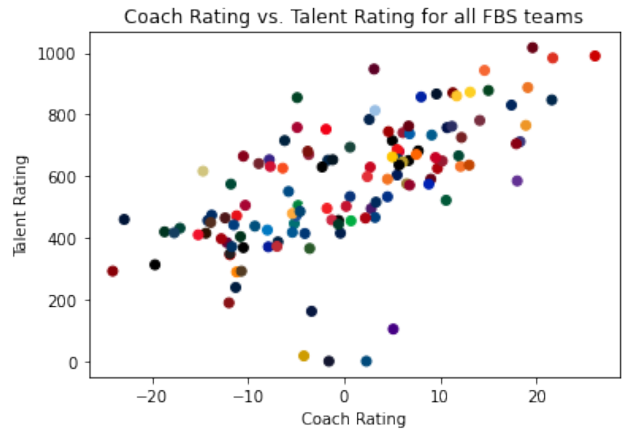
This was the key reason for choosing to use a multilayer perceptron. The success of the resulting model was measured by comparing the projected outcomes of the New Year's six bowl games to the current point spreads available on FanDuel Sportsbook.

Dataset

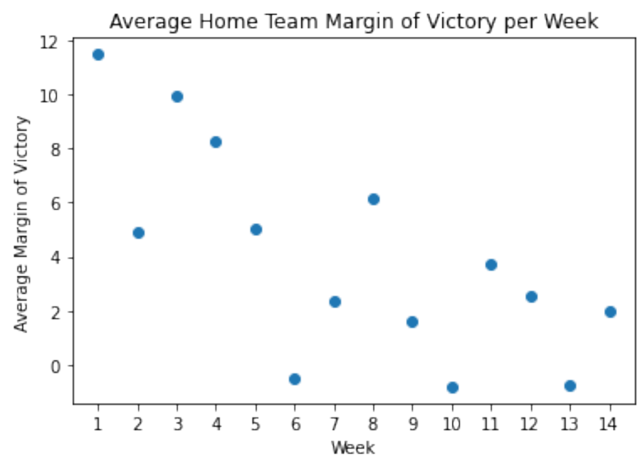
The datasets used for the training of this model are all of the game results from the 2022 season up to this date, head coach rating, and team talent rating. As mentioned above, the FBS team list dataset was used to help compile the more important data sets into one data frame. This made handling the model more efficient. Most of the columns from Game results data were removed because the only information to be concerned with was margin of victory, home team, away team, whether the game was played at a neutral site. Removing all "elo" ratings was logical because those take prior seasons into account. Given the new landscape of college football with the transfer portal and how rosters completely turn over every for years, it made sense to cut out this rating. The excitement rating of a game was also removed because it has nothing to do with what is actually on the field. The model obviously should not be concerned with the average fan's excitement level. Removing the scores from each quarter of a game from the data set was also important. A field goal is worth 3 points in the first quarter, much like it is worth 3 points in the fourth quarter. All matchups that included a non-FBS opponent were also removed from the dataset. This is because the talent gap between your average FBS and FCS team is so large that it is fair to say that they do not even play the same sport. The data from those matchups do not help the model.

The coaches offensive rating and defensive rating was removed from the coaching record and history dataset because there was no special teams rating. Special teams is an aspect of football that is overlooked in most places. There are three phases to football, not just offense and defense. Games are sometimes won and lost based on special teams performance. This is why only the head coach rating and the school of the coach were pulled from this dataset. All of the data from the team talent composite ranking was used because it was only a two column data set: school and talent rating. New columns were then added to the game results dataset: home talent, away talent, home coaching, away coaching. Iterating through the game results dataset allowed the team talent and coach rating to be put into the correct place for each matchup.

Here is a scatterplot on each teams talent rating and coaching rating, colored by the teams primary color:



Here is a scatterplot showing the home teams' performance week-to-week over the course of the 2022 season:



Machine Learning Approaches and Methodology

A multilayer perceptron was chosen because college football is an extremely volatile product, unlike its professional counterpart. Not every college football team responds to a past result in the same fashion. Some teams follow up a loss with more losses. This is referred to by many in the industry as "free-falling." Teams that achieve victory sometimes grab momentum and appear to improve week-to-week; this is called "ascension mode." Attempting to differentiate whether a team is free-falling or ascending as opposed to just having an outlier performance is one of the most difficult aspects of sports gambling. One of the primary goals for this model was to try and differentiate between the two. Regression back to the mean is concept that successful bettors use to their advantage. While the public mostly overreacts to a result, smart sports gamblers use this as a key reason for placing a bet (and most likely betting against the public). Roughly, the same players take the field for a team weekly, but a different team shows up on the field from week-to-week. Teams get "fired up" for certain games for miscellaneous reasons: rivalry, revenge for last season, trash talk, etc. Here's an interesting example, when a starting quarterback is ruled out for a game most casual fans would expect a regression in that teams performance. However, this can have a reverse effect on the team because the rest of the team realizes they need to step it up to perform well. This is why it is quite common to see a team without their starting quarterback over-perform relative to expectations.

It was quite tricky tuning the hyper-parameters for the model. Sometimes it would overfit to the data in one direction or another. At one point in this process, the model would only project 3 point margins of victory for a given matchup. This is due to the fact that a field goal victory is the most common margin of victory. Since 2005, around 9 percent of matchups have been determined by 3 points. The goal was for the neural network to cut straight through this data and decide which outcomes are outliers and which outcomes are indicative of future performance. The success of the resulting neural network was determined by comparing projected margin of victory from the model to current point spreads. However, the real success of the model will be determined while watching bowl games actually played on the field over the next month.

RESULTS

The results of the model have been quite pleasing. There are no graphs or plots to show for it, but I do have a chart that compares the model's expected outcomes compared to the current FanDuel Sportsbook point spreads. Here are the model's projections for the New Year's six bowl games side by side with the Vegas point spread

Matchup	Current FanDuel Line	Projected Outcome from Model
Peach Bowl	UGA over OSU by 6.5	UGA over OSU by 3
Fiesta Bowl	MICH over TCU by 7.5	MICH over TCU by 7
Orange Bowl	CLEM over TENN by 6.5	CLEM over TENN by 7
Rose Bowl	UTAH over PSU by 2.5	PSU over UTAH by 7
Sugar Bowl	BAMA over KSU by 3.5	BAMA over KSU by 3
Cotton Bowl	USC over TULANE by 1.5	USC over TULANE by 3
Natl. Championship	UGA over MICH by 8.5	UGA over MICH by 7

The model is nearly spot on in its projections for 5 of the 6 bowl games. The only mishap is with Penn State being projected to beat Utah by a touchdown. A 10 point difference between the point spread and the model's projection is definitely significant. However, if the Penn State wins by a touchdown or more, the model will prove correct given the result on the field. The model clearly does not punish Penn State much for their losses to Michigan (ranked number 2) and Ohio State (4) as much as it punishes Utah for their losses to Florida (unranked), UCLA (18), and Oregon (15). The model is probably also expecting regression back to the mean for Utah after they splattered USC in the Pac-12 Championship Game. Simultaneously, Penn State has been one of the most consistent teams all season. It is safe to assume that these are the reasons behind the model's projection for the Granddaddy of Them All (a.k.a. The Rose Bowl Game). In practice, this would be an example of the model telling the user to bet on Penn State to cover the point spread or the Nittany Lions to win outright. The model also slightly leans toward Ohio State to cover (but Georgia to win) in the Peach Bowl. For the rest of the bowl games listed, the model has no lean either way. Assuming the model is correct and both Georgia and Michigan win their bowl games, the model projects the Bulldogs to defeat the Wolverines by a touchdown and win their second national championship in a row.

DISCUSSION, CONCLUSION AND FUTURE WORK

This model is just getting started, and there is a lot that can be improved about this prediction model. To begin, it would be nice to be able to get final score projections, so that the model can assist with betting point totals as well (this means betting over or under a certain total of points in a game). Using key in-game statistics would be one of the most important additions to this model. Here are a few key stats that are on-deck to be baked into the model: yards per pass attempt, 3rd done conversion rate, red zone efficiency, and many other efficiency statistics. Efficiency statistics (i.e. passing yards per attempt) are more indicative of who will win a football game than generic stats (i.e. total passing yards).

It is obvious that the model picked up on games being decided by field goals and touchdowns more often. But, incorporating the ability to spit fractions of a point out when projecting margins of victory would be a useful tool. The model seems to produce winning by 7 instead of wining by 6 just because winning by 7 is more likely overall. For example, if the model is somewhere between 3 and 6 points, the current model would spit out 3 or 7. It would be a useful if it could produce a 4.9 projection instead. This indicates a more precise measure of confidence for the model.

The ability to move players in and out of a given team's contributing roster would be key when it comes to handling roster changes, for both newly injured and returning players. One key player can change the outcome of a game, so being able to manipulate the model's view of a position group (defensive line, running backs, etc) would increase the performance of the model. During the season a lot changes about a team, and incorporating injuries into the model would without a doubt would increase performance. The backup quarterback dynamic that is mentioned above could also be exploited with the model keeping track of which injuries to key players. There is a whole lot more that could and will be incorporated in this model in the future.

AUTHOR INFORMATION

Kenneth Woodard, Student, Department of Computer Science, University of Tennessee.