Touchdown Forecast: Using a Multilayer Perceptron to Project NCAA Football Games

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

BACKGROUND, INTRODUCTION, AND MOTIVATION

First, a note on the model. This project is actually a followup to previous work done by the author. References to the previous project will be used throughout this abstract to explain decisions made regarding the new, over-hauled version of the model. The first college football model was built on flawed logic in numerous ways. After testing the old model against last postseason, its performance revealed that it was completely broken and beyond repair. This failure to prove that neural networks could project these matchups was the inspiration and motivation for this project. If the originality of this project is in doubt, refer to the previous paper, *College Football Prediction Model,* which is included in the GitHub repository. The only element of the broken model that was retained was the original datasets. However, an additional dataset that included betting information was added to this version.

A version of Touchdown Forecast beyond 1.0 is currently in the works, but this model is not going to be included or discussed in this report due to the value of the findings made during its development. Additional columns of data, accuracy metrics, and other features are included in the new and improved version. However, the author plans to monetize Touchdown Forecast's predictions in the future. Therefore, due to the extremely promising performance of Touchdown Forecast 1.5, the details will be withheld from this submission entirely to maintain the author's intellectual property. The previous version of the model will be referred to as "Touchdown Forecast 0.0" because of its broken nature and inability to forecast these matchups (its record is also in the repository). While Touchdown Forecast 1.0 is not as good as its updated counterpart, the performance of the model depicted here is enough to prove that neural networks can indeed be used to project college football games.

This study focuses on developing a multilayer perceptron prediction model trained on a combination of datasets obtained from the free *CollegeFootballData.com* website. The selected datasets include roster talent ratings, coach ratings, betting data, and game results for FBS college football teams during the 2022 regular season. The model aims to provide an advantage in gambling on college football point spreads by predicting the margin of victory or defeat for the home team (if unfamiliar, see the introduction of the author's previous report, *College Football Prediction Model,* for information on how betting lines work).

The gap between the general public and oddsmakers in accessing data insights for sports betting is significant. Oddsmakers utilize math and data science to set accurate lines, giving them an advantage over the average bettor who lacks such information. This project aims to bridge that gap by empowering bettors with similar tools to make informed decisions. The goal is to enhance returns and minimize losses for the average bettor, who is often exploited by sportsbooks. While there may be instances where personal intuition proves correct over the model's predictions, it is essential to recognize that Las Vegas preys on gamblers who rely on their instincts rather than data-driven analysis.

The choice of a multilayer perceptron stems from the non-linear nature of college football data and its volatility (Woodard, College Football Prediction Model 2022). The success of the model is evaluated by comparing proximity of the result to the prediction and whether or not the prediction was on the correct side of the spread. The model has hovered right below 60 percent during the course of last season.

DATASET

The dataset used for training the model consisted of game results and betting lines from the 2022 regular season, head coach ratings, and team talent ratings. The game results dataset was utilized to consolidate the essential data into a single data frame. Many irrelevant or unnecessary columns were removed, and all matchups that included a non-FBS opponent were also taken out of the dataset. This is because the gap between your average FBS and FCS team is so large that it is reasonable to say that it is like comparing apples to bowling balls. Therefore, the data from those matchups do not help the model (Woodard, College Football Prediction Model 2022). Spread data was added to this version so that the model would pick up on patterns regarding teams' past performance in comparison to the expectation of oddsmakers and the betting public. Adding this data has shown very positive results.

Touchdown Forecast 0.0 was attempting to predict only bowl games. Its lackluster performance shifted the goal of this model significantly. Its spread picks for bowl postseason were correct less than 50 percent of the time. After reviewing the failures of the previous model, the focus on the regular season was chosen due to the sizable home-field advantage in college football. Unlike its professional counterpart, there is a greater emphasis on which team is hosting the game, meaning that this model is significantly

predicated on where the game is being played. This model was not tested on conference championship weekend or bowl season for this reason; and therefore, postseason data was excluded. The final dataset included every matchup's final score, betting point spread, week number, and game location data for each game in the 2022 regular season. The coaching data and roster data were included as well to describe and identify each team for the model.

Here are 2 pie charts depicting the difference in the frequency of neutral sight games between the regular season and the postseason. This shows why the decisions regarding the postseason were made in the new model:
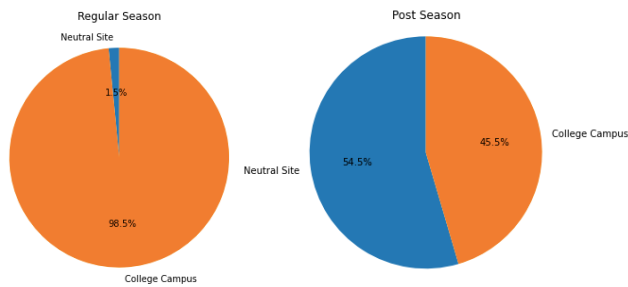
Here is a bar graph depicting the average coaching rating and roster talent for each conference. This shows the importance of these ratings in projecting non-conference regular season matchups between FBS teams. This is because teams that are not in the same conference rarely share common opponents:
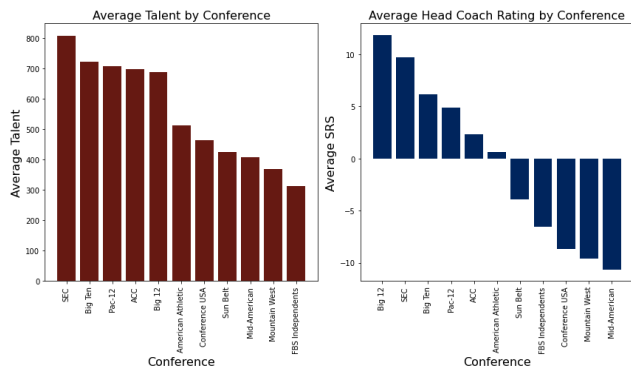
METHODOLOGY

The previous, broken version of the model was built on faulty logic in numerous ways, leading to the methodology used in Touchdown Forecast 1.0. As mentioned above, the previous model was designed to project games on neutral sites but was trained using regular season games that were played on college campuses 98.5 percent of the time. This proved very problematic once bowl season was completed at the beginning of this year.

The previous model also employed a random test-train-split on the data. Upon surface-level inspection, some would think this would make sense when testing parameter combinations for a multilayer perceptron. However, when the random split happened, it did not take into account which weeks were being trained and which week would be tested. This led to Touchdown Forecast 0.0 producing accuracy numbers for certain hyperparameter combinations that were not very insightful or realistic. If the model knows what a team is going to do in week 10, but is testing what it will do in week 9; the model would have an unfair advantage when testing for accuracy. Once the final version of Touchdown Forecast 0.0 was trained using only past data and tested on the next round of matchups (bowl season), its performance suffered due to poor optimization. At the time, it was reasonable to think that the model would perform decently because the predictions were close to the point spreads. That is not how it turned out. That is why the 1.0 version of the model was only trained on weeks before the week being tested, and tested on each week during hyperparameter testing.

Touchdown Forecast 0.0 used a classifier to predict scores and was tested for accuracy based on whether or not the model correctly projected each score. This also caused problems for the old model because it would overfit to certain margins of victory like 3 or 7. This impacted the previous model's ability to project the correct side of the spread bet. Conversely, the new model uses a regressor to predict scores, and each score is rounded to the hundredth of a point. Because the new model can't project the exact score anymore, a new custom metric called "field goal accuracy" was used to measure the accuracy of the model. If the projection was accurate for this metric it meant that the projection was within 3 points of the result. Along with "field goal accuracy," the betting spread was also used for a new implementation that tested the accuracy. The optimal parameter configuration was determined by finding the score that performed best in terms of both types of accuracy.

RESULTS

The performance of the model throughout the 2022 season has been highly satisfactory. It demonstrated significant success, particularly after week 5 when it had accumulated sufficient data from the ongoing football season. Below is a chart that showcases the accuracy of the model for each week of the regular season in 2022:
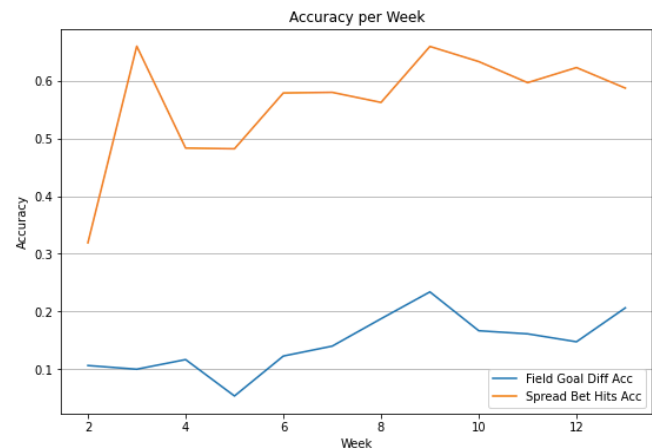
CONCLUSION AND FUTURE WORK

To conclude, the results of this project are highly promising, demonstrating the success of the model even with the limited publicly available data. However, the current cut taken by sportsbooks (10%) makes it challenging to justify using Touchdown Forecast 1.0 in actual betting scenarios. Thus, there are numerous areas for improvement. Incorporating in-game team and player data could be a game changer for its performance because it would allow the model to account for injured and returning players (Woodard, College Football Prediction Model 2022). Sadly, the accurate and verified data that is required for this is behind a $300 paywall (it also happens to be the official dataset of the *College Football Playoff*). While this project has shown promising results, there is a vast potential for incorporating additional features into this model in the future. With continued updates already in the works, this model could become a legitimate betting tool to have in the arsenal. Touchdown Forecast 2.0 should be ready before the start of the 2023 season.

AUTHOR INFORMATION

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

REFERENCES

"College Football Data," *CollegeFootballData.com*. [Online]. Available: https://collegefootballdata.com/. [Accessed: 08-May-2023].

Woodard, Kenneth. "College Football Prediction Model." Dec-2022. *Intro to Machine Learning*, University of Tennessee, Report