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

Kenneth Woodard

Background & Motivation

- Average bettors lack access to the same data insights as oddsmakers, leading to bias in their bets. Personal biases and gut instincts often influence where bettors put their money, leading to losses and profits for sportsbooks.
- However, with the help of data science, sports bettors can level the playing field and make more informed decisions.
- Touchdown Forecast aims to provide a data-driven approach to sports betting, utilizing an artificial neural network to predict outcomes and inform betting decisions. By minimizing personal biases and using real data, our model can improve returns and reduce losses for College Football bettors.

Related Work

- ESPN Football Power Index: FPI is a widely used ranking system that was developed by ESPN. It provides weekly predictions for games based on win percentage.
- Bill Connelly's SP+ Ratings: SP+ takes into account factors such as returning production, recent recruiting, and coaching changes. Provides a detailed look at team performance as well as strengths and weaknesses.
- College Football Nerds Podcast/Youtube Channel: These guys have their own model that is similar to mine except they use more advanced metrics (from cfbstats.com). They also use more than just an ANN.
- KenPom Ratings: Ken Pomeroy developed his model for college basketball based on advanced metrics such as offensive and defensive efficiency, pace of play, and home court advantage.

Methodology

Gather data from CollegeFootballData.com and preprocess it. This data includes information about matchup history, roster talent, coaching rating, and the vegas betting lines for each game.

Train and test a multilayer perceptron regressor from scikit-learn using a multitude of different parameters as well as different weeks throughout the 2022 season. Use a heatmap to view each score and decide on the best parameter. Accuracy is determined in two ways: how close the prediction is to the actual score and whether the prediction is on the winning side of the point spread.

A function using the newly optimized neural network is ready to predict games. The predict_game_score function reads this input: week, home_school, away_school, spread, neutral_site. It returns a prediction such as "Alabama(-7.25) at Tennessee." In this example, Alabama was a 9 point favorite (-9) against the Vols, and the model was on the correct side of the spread because Tennessee won the game outright.

Summary

Here are some visuals that depict the data that is being used to train the model.
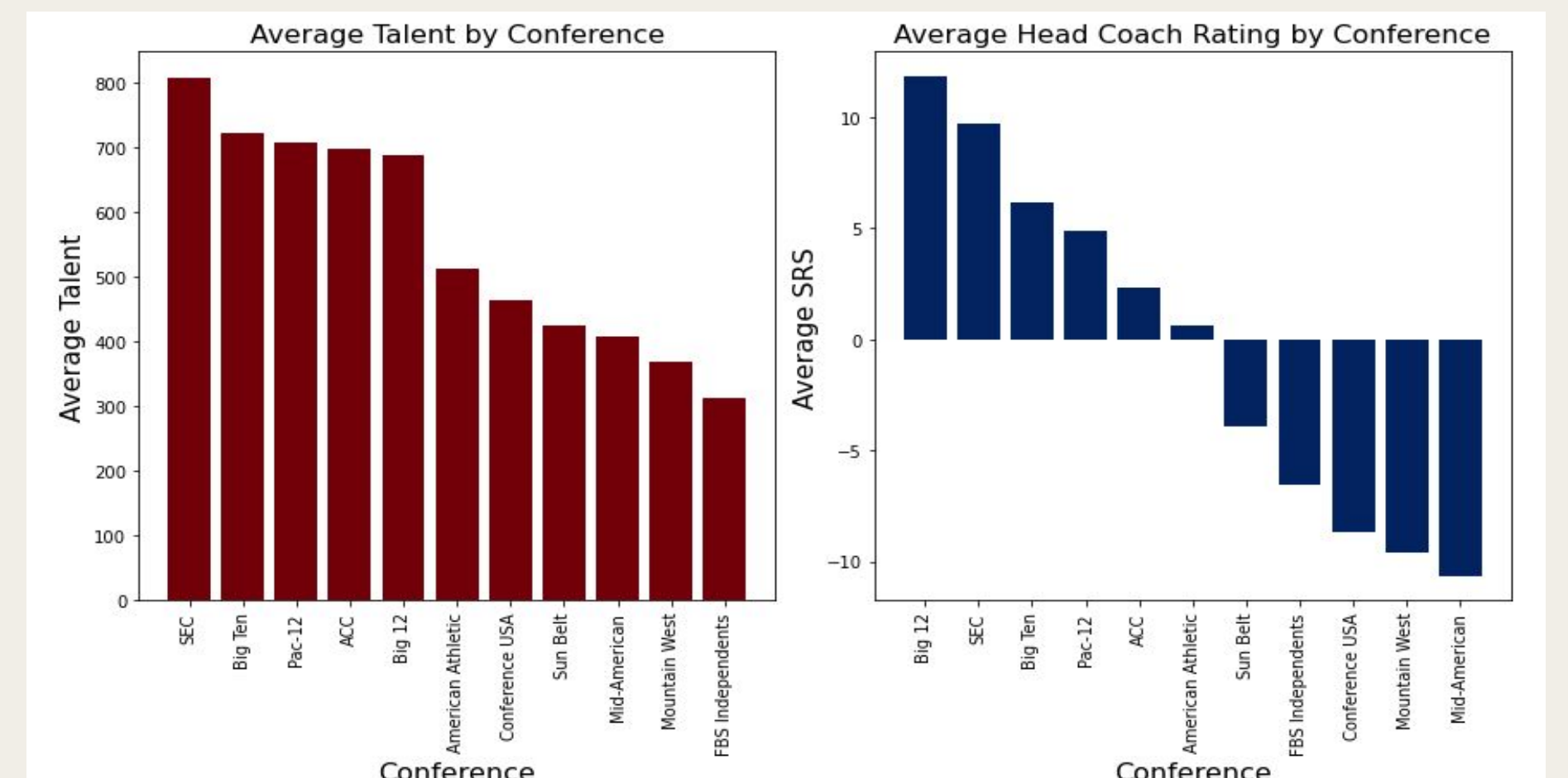

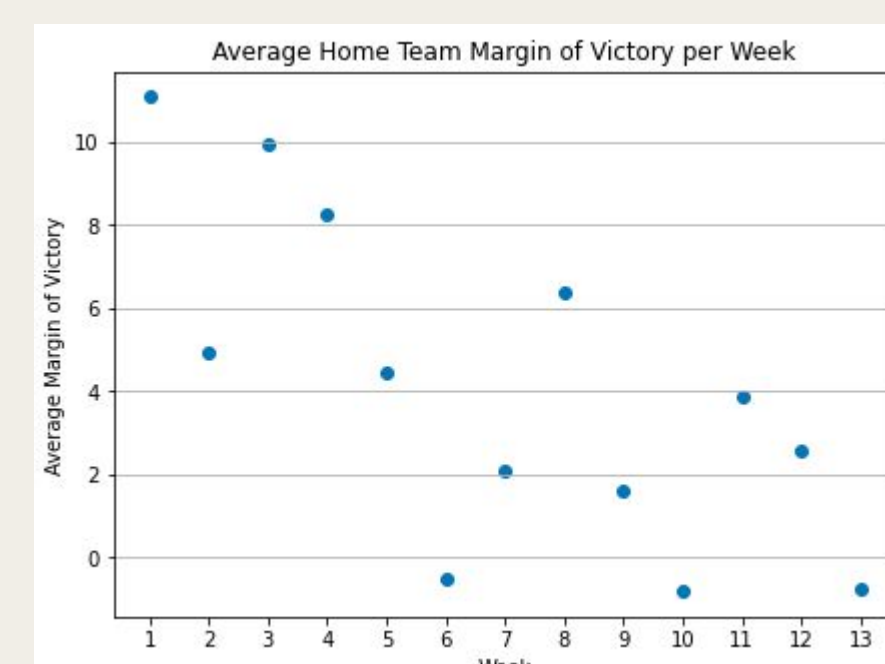Figure 1: Team talent and coaching rating by conference.

Figure 2: The improvement of performance for the away team during the 2022 regular season.
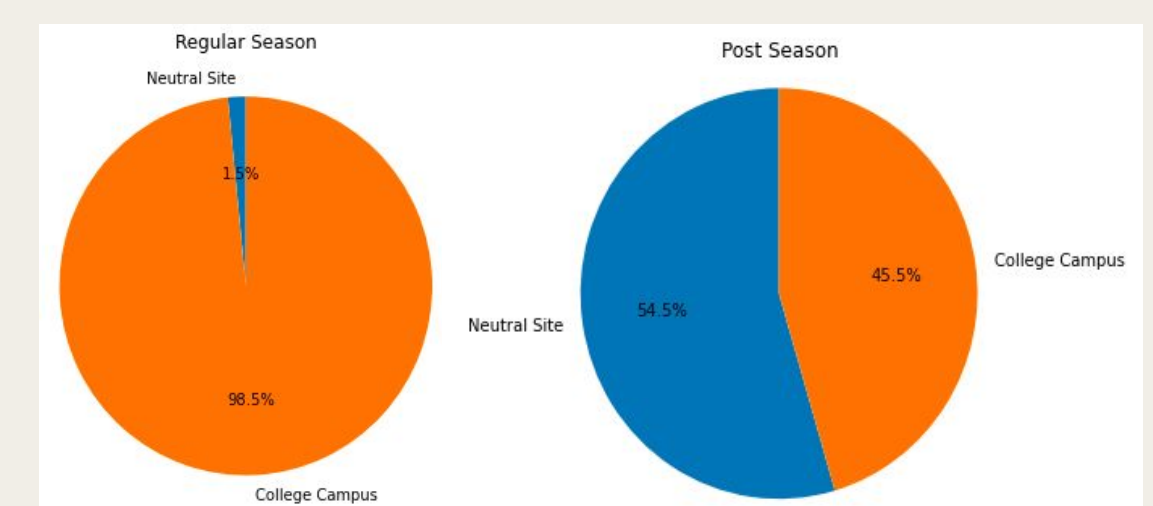

Figure 3: The percentage of games played at neutral sites during the regular season and the postseason. This depicts why this model was trained for the regular season: home field advantage.

Results

It is impossible to determine right now whether the success of this model will be replicable in future seasons. However, the model as of right now can be deemed a success based on its performance during the 2022 season. The model performed well after week 5, once it had enough data from the current football season. Here is the model's accuracy charted during each week of the 2022 regular season.
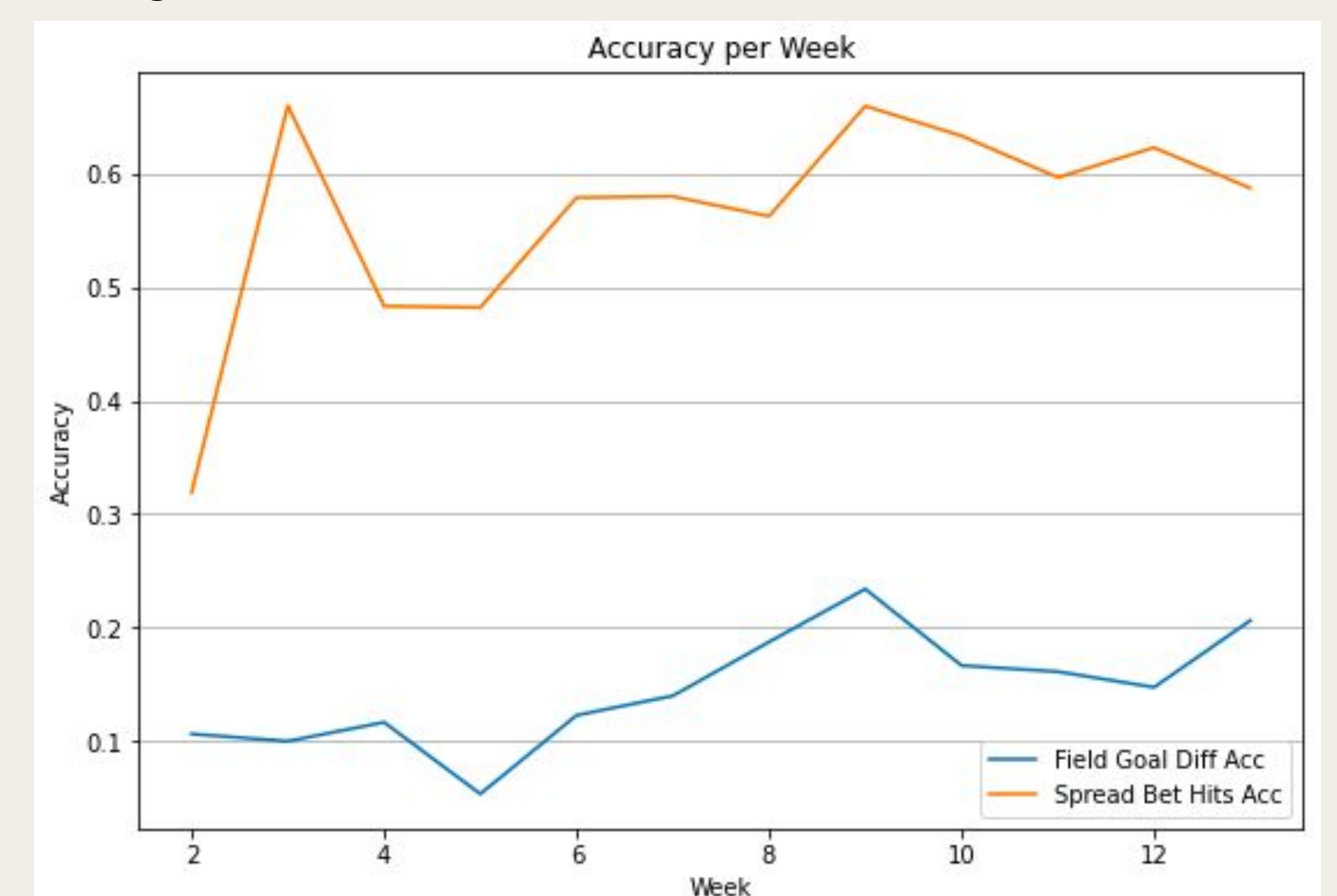
Conclusion

The results of this project are quite encouraging. Even with the limited available data, this model can be deemed a success. This project has proven that with more advanced statistics, the model can be improved enough to gamble on. The cut that the sportsbooks take (10%) is too much to justify using this model in practice as of right now. More updates will be implemented, more seasons will be tested (to ensure replicability), and the upcoming Touchdown Forecast 2.0 should be finished before the 2023 season starts.

THE UNIVERSITY OF TENNESSEE KNOXVILLE