CFB Project Summary

Steve Czarnecki

2024-12-20

## The Model

I wanted to build a point differential model for CFB Games. The intention was to 1) predict games for fun and 2) try to win money vs sports books on spread and/or moneyline bets. The second goal is hard but this is my attempt.

I used cfbfastr to aggregate a bunch of data from previous games, using rolling means from both play-by-play and game-by-game data. In the end, I threw 109 predictors into an XGBoosted model and tuned.

The v1 results were fine. RMSE was at 17.4, which is deceivingly low due to the difficulty of predicting game results. After converting to binary predicted wins/losses, the accuracy was 72.6%, short of sportsbooks’ 74.6%. Finally, our accuracy when using our model to bet the spread was 52.03%, .5% short of the 52.5% mark I projected was needed to break even vs sportsbooks.

From here, I wanted to test the data more to see what our actual returns would be, using actual historical odds and spreads. I also wanted to make it easier to predict future games.

## The App

The goal of the app was to 1) create a dynamic dashboard-like page to evaluate the model’s betting performance and 2) create an environment where I can easily predict future games.

Starting out, I built the betting analysis page. This allows the user to see how the model performs when betting on either the spread or moneyline on our test data (everything that wasn’t used in training). The return on the spread is actually a lot better than expected at -.28% while moneyline is not great at -7%. Compared to the “average bettor” baseline of -4.53%, the model performs very well, almost breaking even. The user can also adjust the filters to see how well it performs in certain scenarios, which will be interesting to see (e.g. where can I improve model, where should I be betting specifically).

Next, I built the game predictor. The user can input any combination of two teams into the page and get an output of projected point differential in the matchup. Here, I can easily adjust if I want to see my model’s projections for the playoff games, for example, or if I just want to see a random score.

The final two pages will be coming out soon. The first will have an adjustable playoff feature, where you can input any 12 teams into the bracket and it will spit out who wins each game and eventually the entire playoff. The next will be a current week page where all of the games for the current week are predicted.

## Going Forward

After I finish the app’s final pages, I will be adjusting the model to make it better. There are a lot of predictors that I did not have enough time to include in v1. The biggest is sportsbook data, which was hard to find. It will be in the next version so I can build off the sportsbooks’ spreads. After I get these predictors in and re-tune, I will spit out the new model, adjust the app accordingly, and see how much better v2 performs.

In the end, I don’t expect to actually be able to bet with this. Sportsbooks are very good at setting lines, and even if I’m confident of the external validity of my test data, my returns will likely only be 1 or 2% at most. Since I don’t have thousands of dollars to throw at CFB games, this model won’t be very practical betting-wise. It will be cool just to have a model that outperforms sportsbook spreads and can give me good estimates on games in the future.