NFL Game Prediction Model: CS 445 Final Project

Objective—The goal of this project is to collect and process data about the NFL for the purpose of building a model to predict the outcomes of games. This knowledge will then be used to locate specific betting lines that likely would be profitable.

I. MOTIVATION

With the advent of legal sports betting in many states across the country there has become increased demand for accurate predictions for the outcomes of many sporting events. In the US the NFL is the most popular sports league and as such has the most demand for advanced analytics. The motivation for this project is to attempt to use a wide range of data about previous NFL games to accurately predict the outcomes of future games. These outcomes would be compared to the available betting lines for different games in order to find markets that will likely be profitable. The end goal of this project is to create a model accurate enough to consistently make money from sports betting.

II. DATA COLLECTION

Sports fans have for a long time been interested in predicting sporting events so there is no shortage of data available. However the difficulty lies in collecting the correct data and analyzing it properly. Putting emphasis on the right variables and filtering out less meaningful factors. For this project we will focus on two different types of data. Generic data about NFL games from prior years, and more recent data from games the last two seasons.

A. Historical Data for Game Outcomes

While the composition of specific teams can change significantly from year to year there are many general trends that existed in prior years that can shed light on what may happen in the present. The main data we would be interested in from prior years would be win margins, total points scores, date time and location of games, and trends between consecutive games. It should not be too difficult to locate/create a database with this information.

1. Win Margins

The first piece of data we are interested in collecting is the win margins of different games. A popular bet in NFL sports betting is the spread bet. This involves betting on a team to win by at least a certain amount of points, or to lose by no more than a set amount of points. The number of points a team is predicted to win by is often called the line of the game. This number is generated by sportsbooks based on their own research and the amount of money placed on both sides of the line. For example if Team A was favored to win by 3 points, and 80% of the money was on Team A to cover the spread, sportsbooks will often move the line to encourage more people to bet the side with fewer bets, guaranteeing them more money. This could result in the line moving to -4.5, or Team A needing to win by 5 to cover the spread. Many sportsbooks also allow bettors to move the line either up or down for either raised or lowered odds.

By performing analysis on common win margins we will be able to determine which spread bets are reasonable and where it would be advantageous to modify the spread or bet the other side. For example the most common ways of scoring points in football are field goals, worth 3 points, and touchdowns, usually worth 7 points. As a result a team winning by either 3 or 7 points is fairly likely, compared to other margins. If you think a team is going to win a game, it is not unreasonable to bet on them to win the game by 7 points and this will often add a fairly large odds boost if that game is a toss up. Performing an analysis on win margins would allow us to maximize our payouts by betting on teams to win by the more common win margins and avoiding uncommon ones.

2. Game Information

Another interesting data set we plan to look at is the date, time, and location of games to determine if these variables have an effect on the outcome or how the game is played.

i. Home / Away

In the NFL each game has a home team and an away team. It is generally understood that the team playing at home will likely do better than expected and the team playing away will do worse. Our goal is to quantify this “home field advantage” in order to determine how strong an effect it has on predicting games.

ii. Day of the week / Time of day

Additionally during the regular season there are games played on Thursday night, Sunday early and late afternoon, Sunday Night, and Monday night. These different days and times of

day can influence the play of the teams involved, and we would like to quantify this influence and take it into account.

iii. Momentum, Streaks, and Rest Time

Another interesting piece of data we would like to evaluate is how previous games affect future games. Examples of scenarios we could examine would include: teams coming off of a big win, teams coming off of a tough loss, teams returning home after a long stretch of away games, teams playing after a bye week, or after a long time without one, teams playing on a short week, (Sunday - Thursday night), (Monday night - Sunday), or teams playing off of extended rest, (Thursday night - the following Sunday). It is likely these factors have an impact and our goal is to quantify it.

iv. Divisional and Rivalry games

The final piece of data we wish to collect from historical games concerns divisional and rivalry games. The NFL is divided into 8 divisions with the teams in these divisions playing each other twice a year. These games may behave differently than others because the players are often familiar with each other and are often more motivated to win. The same would go for rivals outside of the same division. Understanding the impact of these games could provide valuable insight into predicting the outcome.

B. Recent Data for Player Prop Prediction

In addition to general data from past seasons we will also add in more detailed information from the current season and the past season. This data would include offensive and defensive performance, specific player performance, and tendencies of each team.

1. Offensive/ Defensive Performance

Different teams have different strengths and weaknesses that are important to consider when determining the outcome and how players will perform. For example if a team has a really good run defense and a poor pass defense it is reasonable to assume a team will try to pass it against them more often to avoid their biggest strength. Additionally if a team is really good at passing the ball it could be assumed the other team might focus more attention on defending the pass, and would as a result allow more rushing yards. Analyzing this data could be useful for predicting how teams will stack up against each other and for determining the likely game script.

2. Specific Player performance

Another common market for sports betting is player prop bets. For these, people can bet on whether a specific player will gain over or under an amount of yards, or if they will score a touchdown, etc. By analyzing how specific players do in certain matchups a lot of useful information could be gathered and applied to many connected prop bets. For example, if a defensive player is known to be very good at guarding the best receiver on a team it would be reasonable to bet for that receiver to gain under his expected total, and potentially for the second best receiver to gain over his expected total. Betting on both of these connected lines together in a parlay could lead to some large payouts. For this project we would like to look into matchup specific data to attempt to find these scenarios.

III. TEAM MEMBER RESPONSIBILITIES

A. Buddy Swan - Model Development

B. Ryan Seamons - Data Source Acquiring, Model Development

C. An Phan - Model Development

D. Seth Johnson - Data collection and game predictions

IV. MILESTONES

October 16: Finish performing analysis on win margin data.

October 23: Finish performing analysis on Home Field advantage

October 30: Finish performing analysis on day of the week / time of day and momentum / streaks and rest time.

November 6: Finish performing analysis on information from past games and look ahead to current season stats.

November 13: Examine trends in Offensive and defensive performance for the current season.

November 20: Collect data on specific players and matchups for the current season.

December: Produce a finished model to predict games.

V. EXPECTED OUTCOME

The end goal of this project is to create a model, using as much data as reasonably possible, to accurately predict the outcomes of NFL games. Additionally we would like our model to be able to predict more about the game than just the winner, including the expected margin of victory, total number of points scored, and individual player statistics. These results will then be compared to betting lines to find the bets with the most expected value.