

NFL Super Bowl Predictor
Zak Beach
Level of Interest 10/10

Description

There are many different statistics in the NFL that have been used to compare teams to each other. Some old adages such as “defenses win championships” have been questioned over the past couple of years as high level offenses have seemed to dominate the sport. The importance of different positions have seemed to change as things in the NFL shift. The NFL is a copy-cat league. Most of these shifts happen first to the teams at the top and then trickle down below and most analysts do not necessarily notice these changes until almost every team in the league is doing it. My idea for this project is to use machine learning to better understand which features makes a team more likely to win a Super Bowl and to hopefully train a model to be able to predict which team will win or atleast make a deep playoff run. This will also indicate trends that will most likely trickle down to the rest of the NFL as the features that really stand out will most likely not be used by everyone.

Data

One of my ideas that I have is looking at the winning team’s passing yards per attempt in contrast to their rushing yards per attempt. Higher pass per attempt is more indicative of being able to threaten opponents with the deep ball on a more consistent basis. If it is shown over the years that the more successful teams have this ability then one can come to the conclusion that your team is more better suited investing in a deep-threat wide receiver who can make plays like this as opposed to one who has strengths in mid route concepts. Because there is so much data (much more than I could put in this paper) there is a lot of room to test which is better for training.

| Team | Passing Yards Per Attempt | Rushing Yards Per Attempt | Turnovers | Points allowed per game | Defensive Yards Allowed | Won SB |
|-----------|---------------------------------|---------------------------------|-----------|-------------------------------|-------------------------------|--------|
| KC Chiefs | 7.5 | 4.7 | 23 | 21.7 | 5579 | 1 |

Gathering The Data

NFL stats can be web scraped from sites like pro football reference; you can get data for each game of the season. The challenge will come from cleaning the data and organizing it. The data is in the form of game logs that show team stats for each game. What people who work on this project will need to figure out is if it is more advantageous to consider stats on a game by game basis or collectively as a season. For example a team may have 6 fumbles all season but in the first game of the season they happened to have 4 so after the first game they only had 2 fumbles. Differences like this could make a difference to the model so the team may have to adjust for things like this. The data is not ready for training at all and it will take some thought to get it ready to train

