Rhoads MacGuire

Gambling Line Accuracy and Weather in the NFL

Proposal:

Football is the only major american sport in which weather is a factor. Of the four major American sports, basketball and hockey are played indoors so weather has no effect. Baseball games are cancelled in the presence of precipitation so weather does not truly have an opportunity to affect this game in a meaningful way. Football is the only sport where we regularly see weather impact the result of the game. Furthermore, football is also the most widely gambled on american sport, with casinos in Nevada handling over two billion dollars in wagers in 2018 (https://www.legalsportsreport.com/28482/nevada-sports-betting-revenue-december-2018/). Additionally, with the change in federal legislation, sports gambling will become more widely available and socially accepted in the United States. This will increase the size of the United States gambling market of which football is the most popular sport to wager on by a large margin. With this in mind, understanding the effects of weather on how the result of the game compares to the spread has large and growing monetary implications for sports books.

Data Collection and Wrangling:

The data I am using for this project comes from Kaggle and is comprised of NFL game data, including points scored by each team, the point spread and over/under lines, and weather details on the game. The data set did not require a lot of cleaning, but it did require some additions and alterations to better suit my intentions. Those intentions of course are to analyse the effects of weather on the accuracy of both the point and over/under spread.

The first cleaning step I took was to delete columns that would not improve my analysis. Most of these columns contained superficial details. Others contained data that could have been useful, but were so unpopulated it threatened their validity. For example the column titled "weather_detail" claims that only 100/ 7725 NFL games in the last thirty years featured rain. That number appears to be too low and including it could show us false conclusions a more robust dataset would disprove. It is possible to gauge the effects of extreme weather with temperature and wind speed data. That being said, precipitation data is not terribly difficult to find and it would make a useful addition to the data set.

In order to acquire precipitation data, I scraped the website, almanac.com. To do this, I needed the date and zip code for the almanac url. While we had dates for all the games we did not have zip codes so I created the 'zipcode' method which took the names of the stadiums and found where they were located and what the zip code was. Additionally, while we had all the dates, they were in the wrong format so I created the 'dateSwitch' method which took the dates as they were formatted in the original data set and reformatted them to work in the url of almanac.com. The final step was to create a for loop to launch a request for each row in the dataframe and return the amount of precipitation for that day in that area. While this loop was relatively simple to construct it became obvious that almanac.com was not going to allow seven thousand consecutive requests. This problem perplexed me for a while but the simple solution was to have the program pause whenever it appeared it was over requesting the website. This was easy to judge because the website would throw a nonetype causing the program to crash. To solve this, the program simply checks if it received a nonetype and if so, it pauses and tries again.

My next step was to eliminate rows where essential information was missing. This information included, point and over/under spreads, temperature, wind speed, and points scored by each team. I cleaned these rows using a boolean series and the notnull method. The result included only rows where all the data was present. While reviewing the data after this step, I remembered the 1987 NFL strike where the owners continued the season with replacement players. While this period was only a handful of weeks, I believe the resulting changes in the league warrant eliminating data from before this point. This left us with 7725 full rows of data, but some adjustments still had to be made.

For my analysis I wanted a column that was the score of the game relative to the spread. To do this, I decided to make a column, "results", that would indicate the point difference in the game between the two teams. It would be negative if the favorite won and positive if the underdog won. This would make it easy to use in formulas with the spread column, which is always negative, to indicate the favorite. To create this column took a few steps. First, I needed to change the abbreviations in the home and away team columns to match the 'spread_favorite_id' column so the 'spread_favorite_id' would match either the home or away team. This was complicated by the fact that historical teams, such as the Houston Oilers, were associated with the same "TEN" team id, as in the Tennessee Titans. To make the data easier to analyze, I converted all historical teams to the abbreviations of their current iterations. The names of the teams do not really matter but this was essential in further analysis. To do this I wrote a method called teamSwitch. This method takes the long hand name for a team and reduces it to their 3 or 2 letter acronym. Once the acronyms matched I made a column to show whether the home team was favored and created the ‘results’ column.

With the results column created, I was able to create a few more columns for analysis. The first being the ‘results-spread,’ which will be used to determine how far off the spread was for a given game. For example, if the home team is favored by three points and they win by six, the recorded value in ‘results’ is negative six, because the home team won by six, and the ‘result-spread’ is equal to negative three, because the result was three points lower than predicted by the spread. Conversely if the home team is favored by three and they lose by three then the recorded value for ‘result’ is positive three and the value for ‘spread-result’ is equal to positive six because the road team outperformed the spread by that many points. This construction will be useful to see if weather factors favor underdogs or favorites. We will use a similar construction on Over/Under (O/U) lines. That is to say if the O/U line is forty and the total scored by each team is thirty-five, the column ‘over_diff’ will show a value of negative 5. Additionally, the absolute values of these columns (‘resultsmspreadabs and over_diffabs) can be taken in order to analyze what types of weather causes the most pronounced changes in gambling lines, regardless of direction.

When checking my data for outliers I realized that while some values in the column, point spread, exceeded 3 standard deviations from the mean they have value because the most important factor is how the spread relates to the outcomes. Additionally, their does not seem to be much value in eliminating observations from something that is so random as football.

Visual and Inferential Analysis:

Once we started analyzing the data, it became clear that our two main variables would be over_diff and result-spread because these variables are measurements of the error that the line setters were making. The first objective of the data story was to show these errors were increasing risk for the sports books on a yearly basis. While we do not have data on casino risk we can see on average how close the lines they set are to reality. Obviously the further off the lines are the more risk we can retroactively place on the sports book. When comparing the yearly means for the spread differences it appeared that bad weather substantially increased the variance of the line accuracy. In bad weather the standard deviation for all the yearly spread mistake means was about three times what it was in nice weather. The same result held for the O/U line. On a yearly basis, bad weather appears to have an effect on the variance of the accuracy on lines. I wanted to be sure these results were not simply do to time of year, which is obviously very related to weather, so we examined how the mean difference changed as the season went on. The week of the season does not appear to have a meaningful effect on either variable. Interestingly though, the variables skewed in unexpected directions. Later in the season and in bad weather, it appears that favorites tend to do better against the spread. This is surprising because we would expect weather to be a somewhat equalizing factor so maybe some teams simply perform better in the rain. Additionally, we also see that teams usually outperform the O/U line late in the season and in bad weather. Maybe this is the result of line setters reading too much into weather as the season goes on. In any case, the most useful information gained from visualizations was that weather may not have an absolute effect, but instead it may favor certain situations, such as being the favorite.

Before continuing the analysis, it is important to know how our dependent variables were distributed. With this in mind, we made cdfs for over_diff and results-spread and overlaid them with a normal distribution with the same means. The result was both distributions were quite normal. Next, I wanted to see if the difference between the means of the two variables was due to randomness or because the weather was having a real effect. To do this I created a bootstrap test with mean difference as my test statistic. Because the nice weather and bad weather test sets were not the same length I had to use a permutation test. I redrew the data ten-thousand times and found that while the difference in bad and good weather results-spread was replicated in twenty percent of samples, the mean difference for the over_diff in good and bad weather was under one half of one percent. This means that Over_diff will be more interesting for further analysis because the result was most likely not due to randomness.

We also wanted to find out how individual weather aspects affected our dependent variables. To do this we measured the correlation between each aspect of weather and the variables. It appears that each aspect on its own has only a weak correlation. It seemed reasonable that this occurred because weather typically interacts amplifying its effects. Being cold is bad but it becomes a lot worse when the wind is blowing or it is precipitating. To measure this we gave each row a weather rating from one to four. One is our current 'nw' dataset. Two will be situations in which (exactly) one of our three indicators can be considered 'extreme' or above the 80th percentile. Three will be (exactly) two of our three indicators are 'extreme' and four will of course be all three. The correlation between the weather rating and the variables was extremely weak so it appears breaking the data into just the two sections is best for analysis.

Going forward with machine learning, it appears our most interesting line of questions appears to be what is it about the weather that causes the difference in the mean error in O/U spread.

Machine Learning Techniques:

For this situation, I believe a regression analysis is called for rather than a classification model. While classification models are often useful, we are looking for a numeric relationship between weather and gambling spreads. Our analysis aims to show how a certain amount of weather affects the results of the games compared to the spread and how that effect varies based on changes in the weather.

With the lack of meaningful correlations found in the statistical analysis I decided it would be best to use Lasso regressions to determine if I overlooked relationships between each individual form of weather. Even though I specified that we should narrow our focus onto over_diff I see no harm in applying some of these techniques to result-spread as well. Especially considering the lack of previous correlations of any kind. Lasso regression is an interesting and useful tool because it takes different features and determines which among them explains the most variance in the target variable while reducing the coefficients of other variables to zero. The Lasso regressions confirmed the results of the statistical exploration by showing that none of the individual weather effects explained a significant amount of variance in either over_diff or result-spread. This was expected but using Lasso regression was especially useful in this situation because it regularizes the variables. Regularization is the process of reducing noise in your dataset by converting variables to their standard form. This was ideal because the variables were on very different scales. A difference of one unit is not that big a deal for temperature but it certainly is for rainfall. Regularization can be very useful in situations like this but can be rather pointless in other contexts.

With Lasso showing no single weather attribute predicting any variance in the target variables, I moved on to testing the only meaningful correlations discovered by my statistical analysis, between bw and over_diff. Additionally, I ran regressions on over_diff with weather rank. I did the same for result-spread because the lack of previous actionable results. I considered continuing my analysis with a regularized ridge regression, but I determined that was not necessary and a normal OLS linear regression would be suitable. A ridge regression is useful in multivariable regressions where the independent variables may be correlated, referred to as multicollinearity. Additionally, the cross_val_score method was used so that each regression could be run multiple times, in this case five, and we could take the mean of the R squared of these five trials. The results of these regressions were as expected. Little to no variance was predictable by these variables. One interesting thing was that all the R squares were negative. While the statistic is called 'R squared,' implying a squared value, it can still in fact be negative. A negative R squared implies that a straight horizontal line would better explain the variance of the target variable than the fit regression line. With this it appeared the model had failed to predict any variance in the target variable.

The lack of success of the previous analysis led me to consider using a classification model. While the model is useless, because it relies on knowing the scores of games in advance of them happening, I thought it may be interesting and good practice for future analyses. The results surprised me and I am not completely sure how to interpret them. The model was able to predict with between 78% and 86% accuracy whether or not a game had been contested in bad weather only knowing the difference between the O/U and the total points and the difference between the spread and the final score of the game. I found this interesting and slightly encouraging, but completely unactionable. The weather can be known hours or even days before the game but the result cannot be known before the game ends. Furthermore, if the result of the games were known beforehand, sportsbooks would not need any help setting their lines.

Conclusion:

This analysis failed its stated goal of predicting how the severity of weather affects the size of the error between the gambling spreads and the results of the games. There are a few possible reasons for this, both conceptual and methodical. The next few paragraphs will address these reasons and attempt to explain their effects and alternate solutions for further research.

The first and most obvious explanation for the models inability to predict the differences between gambling lines and results is that the line setters accurately account for weather. This study using the spread and O/U to show the relative abilities of both teams but, if those numbers are already adjusted for the weather than it stands to reason that a model would not be able to interpret differences due to weather because they are already baked into that number. Additionally, gambling lines often reflect the knowledge of the crowd and are adjusted as money and wagers come in on each side. While these adjustments are usually relatively small they could account for some issues in the model.

One solution to these problems could be use another measure on the ability of each team. Football Outsiders has a stat called DVOA (Defense adjusted Value Over Average), which measures the teams offense defense and special teams in terms of the value they provide compared to a league average unit in that situation. The reason I did not use this number in my original analysis is because it is only available on a yearly basis and I felt it would simply be noise in the model in many situations. For example, if a team has a high season ending DVOA due to an excellent player or unit that substantially improved their play over the course of the season than this would skew results from their early season matchups where they were underperforming. Conversely, if a season DVOA is lower due to injuries than the DVOA will be understated in early season matchups. These issues led me to not include DVOA, but if the number could be acquired on a week to week basis than it could be very useful.

Another potential issue could be all the information condensed into just our three measures of weather. Weather is continuously changing and so one number which measures the whole game may be insufficient. For example, if the weather is extreme in the first half but then clears up in the second, our data has no means of differentiating the two which could affect the results of our model. This issue could be resolved using hour by hour weather data, but that was not available to my knowledge. Even with that data, it still would have been a challenge to code it into the dataset in a meaningful way.

Another potential issue with my methodology is that different players and teams are affected differently by weather. Weather may be an advantage to certain play styles and teams. Additionally, it could serve to amplify teams strengths while simultaneously mitigating their weaknesses. For example, a team that primarily runs the ball may not be as affected by changes in weather as a team that mostly passes the ball. Going off this line of reasoning, weaknesses in a teams pass defense may be more difficult to expose in inclimate weather. While these points are interesting I believe more research should be done into the exact relationships between weather and how it affects different position groups before this information is actually useful. While the relationships above are obvious others may be more difficult to quantify. It is possible that passing the ball in rainy, but not windy weather may be advantageous because quick changes of direction may be more difficult for defensive players on wet fields. The necessary data to fully understand these relationships is currently unattainable because it would require measuring each individual player on each individual play while also having data for each play on the temperature, wind, and field conditions.

In conclusion, there are clearly ways to improve this analysis. While some of them may not be possible in the scope of this study, I believe adding DVOA may have created a more robust dataset, even if it could have also increased the amount of noise in the data. One lesson from this analysis is that gmabling lines are not a great measure of the relative strength of two teams and that other measures should be included to give a predictive model and opportunity to provide actionable results.