Predicting Tornado Intensity in the Contiguous United States

Ben Royce and Charles Stringfellow

Research Questions:

1. *Where in the US do the most intense tornadoes occur?*
    - The most intense tornadoes in the US occur in the Midwest and South US.

2. *To what extent can we accurately predict a tornado's intensity based on the provided features?*
    - Using a multilayer perceptron: 58.0% (model score)
    - Using a decision tree classifier: 63.8% (accuracy score)
    - Using a decision tree regressor: 0.466 (mean squared error)

3. *How does the intensity and prevalence of tornadoes vary with the time of year?*
    - Prevalence of tornadoes greatly varies with the time of year. There are far more tornadoes in the summer months than in the winter months.

Motivation:
The aim of this project is to predict the intensity of tornadoes in different parts of the US at different times of the year. Tornadoes of high intensity necessitate evacuation from their paths, but tornadoes of low intensity might not garner the same response. It is important to predict where tornadoes are likely to be intense so that those cities in those regions can make appropriate disaster-response plans, and so higher governmental bodies can determine how to assign funding to their different regions for disaster preparedness. Additionally, data related to weather events is a good fit for the work of a data scientist, as there are many variables surrounding weather forecasting and we as humans cannot piece them together as well as machines can.

Dataset:
Our dataset comes from the Storm Prediction Center (SPC). The SPC records observations on every tornado to touch down in the US from 1950 to 2021. The dataset contains 68,000 rows and 23 columns and records features such as time, date, location, and intensity. Tornado intensity is rated on the Enhanced Fujita (EF) scale, which goes up to EF5. Tornadoes are rated based on the damage they cause, which is then used to predict their wind speed. Relevant

features for our project will include year, month, day, time, state, magnitude, starting latitude and longitude, path length, tornado width, number of injuries, and number of fatalities.

The link to download is here:
https://www.spc.noaa.gov/wcm/data/1950-2021_all_tornadoes.csv

Challenge Goals:

Machine Learning:

For the machine learning goal, we will include the training of a neural network in our performance comparison on the task of predicting tornado intensity. This task is considered an "ordinal classification" or "ordinal regression" problem, since we have distinct classes that are aligned on a scale from low to high. However, this task sits between a classification problem and a regression problem, and sklearn does not have a standard way of factoring in this ordering. As a result, we will train a decision tree classifier, a decision tree regressor, and a neural network and see which of these three is best for the task. We'll vary the hyperparameters of these three types of models until we get an optimal test accuracy, which will be the final value we use to determine what works best.

Result Validity:

The challenge goal result validity works nicely with our project because it will allow us to assess more fairly how effective our trained models are. This is especially important for a model to be used for natural disaster prediction, as a model that's proven to be effective can be trusted in the process of deciding what measures need to be taken by local governments and weather prediction centers. For this project, we will implement 5-fold cross validation, using the KFold method from sklearn. After obtaining results from use of the entire dataset, we will compare those results to results on new models developed during the cross-validation step.

Method:

1. To examine patterns in the occurrences of tornadoes in the contiguous United States, we will create several MatPlotLib plots using a subplot that show where tornadoes of different intensities fall. There will be six different subplots: one subplot will show all the tornadoes that have struck ground in the US since 1950. Each plot will be a map of the contiguous United States. These tornadoes will be presented as Point objects where the color represents the intensity of the tornado on the EF scale. Then, we will create 5 more plots of the contiguous United States where respective plots will show points representing the tornadoes of the plot's respective EF intensity.

2. This question will be answered through the comparison of three machine learning models with varying hyperparameters. We will begin by splitting the data 80/20 into a training set and a test set. Then, we will create a decision tree classifier of depth 1, determine the test accuracy, and repeat this for depths up to 15. We will compare all the test accuracies to find the depth that worked the best. We will repeat this process for the decision tree regressor and determine an optimal depth. Finally, for the multilayer perceptron, we will come up with a list of hard-coded options corresponding to the learning rate of the model and the shape of the hidden layers for the network. We will then test every combination of these hyperparameters and see which yields the highest model score for the neural network.
3. To answer this question, we will create a histogram comparing the time of year with the number of tornadoes that struck ground in that month since 1950. This method is primitive, but it is extremely useful for answering the important question of when tornadoes occur in the US.

Results:

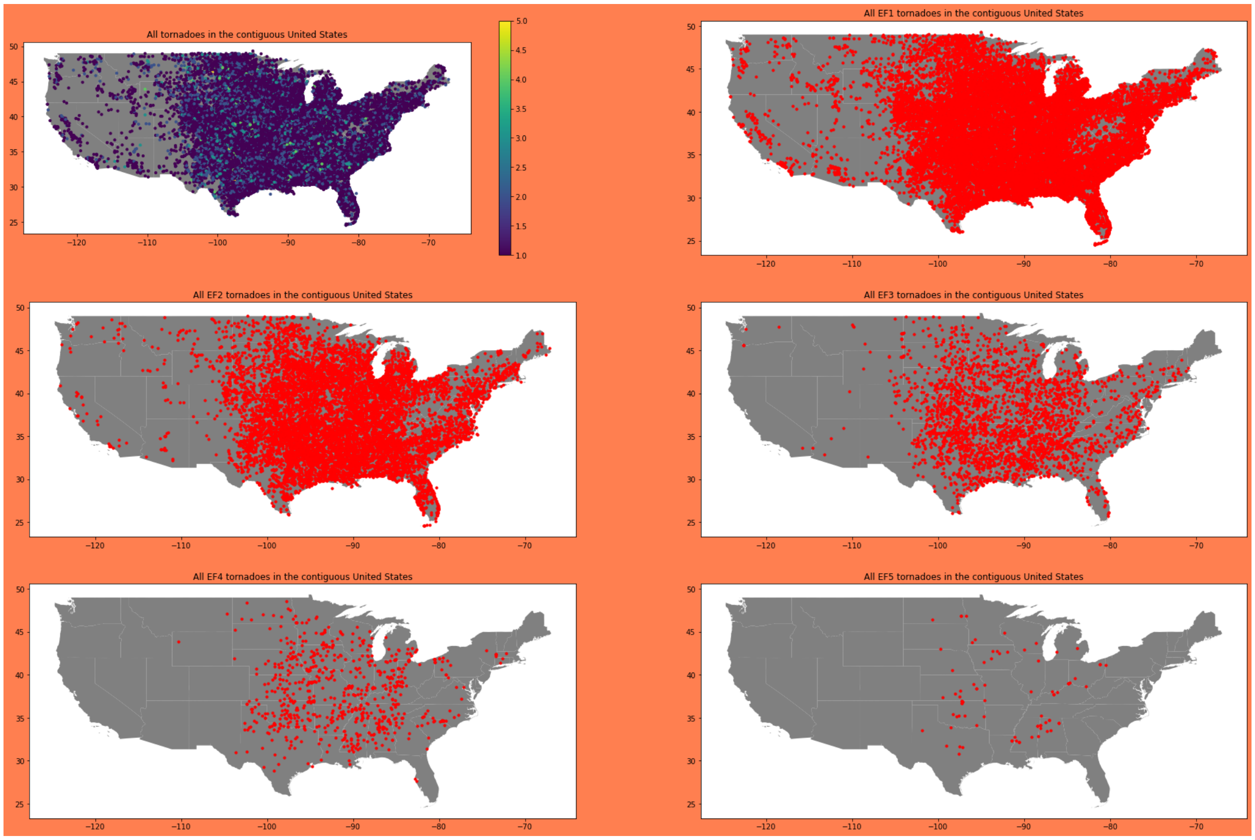

Figure 1.

1. *Where in the US do the most intense tornadoes occur?*
    a. Interestingly, our data processing revealed that tornadoes occur all across the United States in considerable frequency. In the top left subplot of figure 1, the tornadoes in the US are so prevalent that they are indistinguishable, and since lower-intensity tornadoes are more common, it deemphasizes high-intensity tornadoes when in reality these higher-intensity tornadoes should be emphasized greatly considering their destructive power. This necessitates the creation of additional plots to more carefully examine where intense tornadoes occur. However, even with this first plot, it is clear that tornadoes are far more prevalent in the East than the West. Additionally, there seem to be some areas in the Northeast where there are fewer tornadoes. However, when we create more plots that examine the tornados in the US by intensity, we can see the true tornado hotspots.
    b. Lower intensity tornadoes (EF1 and EF2 tornadoes seen in the top right and middle left graphs respectively clearly occur with considerable frequency around the Eastern United States with a particular density in the Midwest and Southern United States. However, the most intense tornadoes (EF4 and EF5 tornadoes)

form two tornado zones. Particularly, two clusters are revealed with the graph of the EF5 tornadoes.

2. *To what extent can we accurately predict a tornado's intensity based on the provided features?*
    a. As it turns out, most of the models we trained for this task were not very effective at predicting tornado intensity. All in all, it seems that a regressor is best suited to a task dealing with ordinal labels when you cannot directly implement an ordinal classifier. Although there's no mathematical comparison between the accuracy score of a decision tree classifier to the MSE (mean squared error) of a regressor, we feel that an MSE of 0.466 is was the best performance we achieved across the different models. Ideally the MSE is as close to 0 as possible, and 0.466 is quite respectable. The regressor achieved its lowest MSE at a depth of 9.
    b. Our decision tree classifier also settled on a depth of 9, which produced a model capable of 63.8% accuracy. Finding the optimal depth for this type of model followed a predictable pattern. With very small trees it had poor performance, then it gradually peaked around 63.8%, and fell off again as the depth increased due to overfitting. It's unsurprising that this model performed worse than the regressor, as it does not take into account the relationship between the classes it's trying to place tornadoes into. It simply sees the ratings from 1 to 5 as distinct buckets to place observations into, whereas the regressor is trying to fit the predictions to an increasing scale.
    c. The performance of the neural network was disappointing, to say the least. Regardless of the layout of the hidden layers or the learning rate of the algorithm, we could not train a multilayer perceptron that would score over 60%. We theorize that neural networks may only be effective when you have a large amount of information to feed into the model, such as for image processing where each *pixel* is considered a feature. Instead of thousands of features, we had 10 for this task. The results from this investigation tell us that neural networks are not a miracle cure for any machine learning task, as simple decision trees outperformed our multilayer perceptron.
    d. The result validation portion of the machine learning work went very well.. For the decision tree classifier, the average test accuracy across the five folds of data was nearly identical to the result we achieved with the entire dataset (63.4%). This allows us to be confident in our initial model's ability to predict tornado intensity at that accuracy. However, our mean squared error for the decision tree

regressor returned a decidedly lower average value across its five folds with 0.390. One reason for this could be that the dataset as a whole has some bias that misleads the regressor, and chunking it up into smaller groups of observations helps eliminate this bias. This is an important takeaway from our analysis, as future analyses may want to use smaller portions of the data to similarly dissipate this bias.

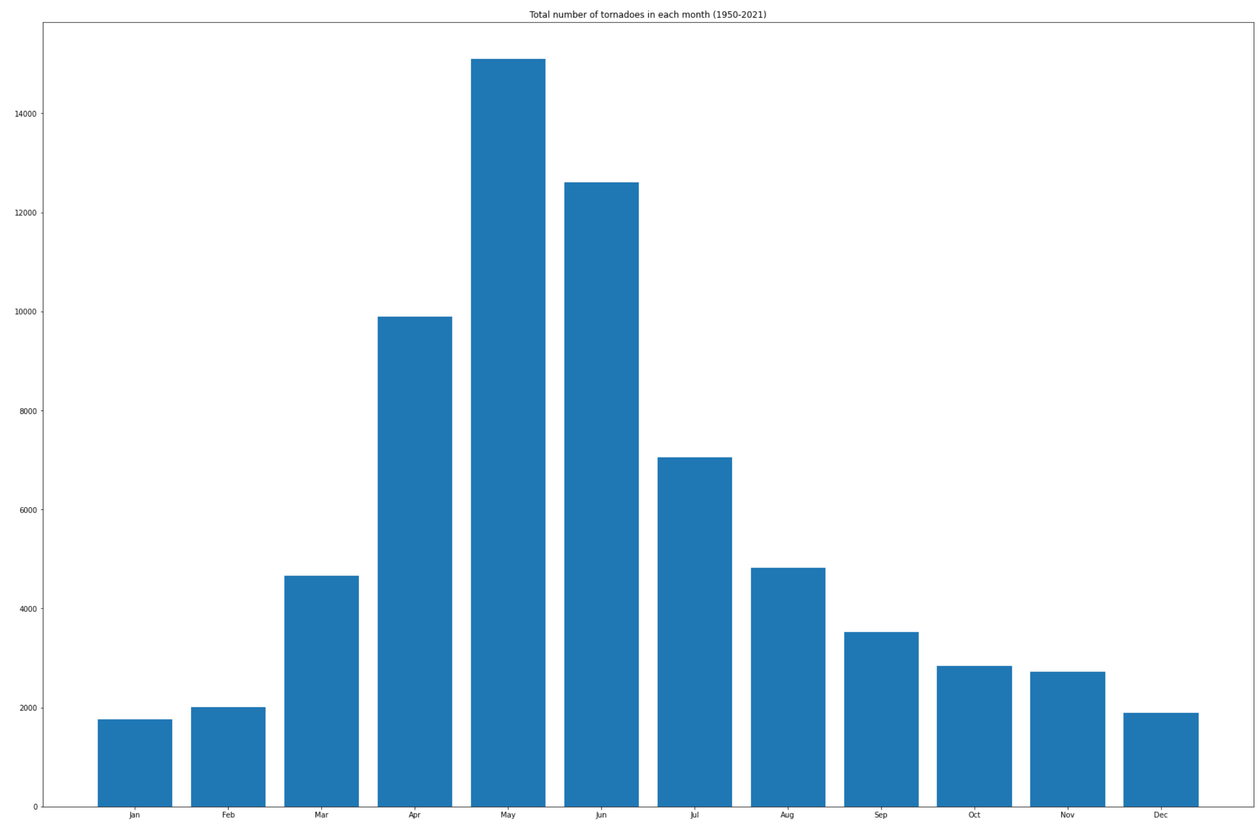

Figure 2.

3. *How does the intensity and prevalence of tornadoes vary with the time of year?*
    a. This was particularly interesting for this project. The tornado season in the United States is extremely pronounced. As can be seen in figure 2, since 1950, there have been over 14,000 tornadoes in May. However, in January, since 1950 there have been fewer than 2000 tornadoes. We can see from the graph that this follows a normal distribution over the months of the year, with especially active months between April and August. As a result, budgetary meetings during this time should give tornado relief/preparedness a higher priority during the summer.

Impact and Limitations:

From our results, we can see that certain areas of the US have far more need to be prepared for tornadoes than others. For example, in Washington there is no cause for concern over tornadoes. As can be seen in figure 1, since 1950, there have been 0 EF4 or EF5 tornadoes. Since 1950, there have only been two EF3 tornadoes. Despite the considerable impacts of those earthquakes, they are incredibly infrequent, so there is no need for Washington to invest in tornado-related disaster relief. This conclusion may lull individuals living in states like Washington into a false sense of security, falsely concluding from the visualizations that there is no reason to ever think about tornadoes. They can still occur in every single state and cause damage, and even if it's not a common event, it's good to know what to do just in case. Oklahoma, on the other hand, has considerable reason to invest in tornado-related disaster relief. Since 1950, as can be seen in figure 1, Oklahoma has experienced 9 EF5 tornadoes, and far more incredibly destructive EF4 and EF3 tornadoes. These tornadoes can level cities if building codes don't account for the stresses of tornado winds. Although an understanding of tornado season may be important for a Washintonian, it is incredibly important for an Oklahoman. An Oklahoman might be inspired to assemble supplies in preparation for a tornado-related disaster in the summer months. However, this project also has many limitations. Firstly, this project only considers tornadoes in the contiguous United States. Alaska and Hawaii were not included in the dataset, so any appropriation of funds from a federal level must take into account that they do not have information about the tornado-preparedness needs of those states.

Additionally, the results of the machine learning models we trained tell us that predicting tornado intensity may not be feasible from the given features. There's no causal relationship between any of the features and high-intensity tornadoes, so it's difficult for the models to achieve high accuracy. If one were to use a model to predict intensity, we'd put forth the decision tree regressor, but it should be kept in mind that it's still imperfect. Perhaps more research needs to be done into what data we can actually gather for tornado events, as there may be atmospheric measurements we can take that will produce stronger machine learning models.

Work Plan Evaluation:

Our initial work plan was ambitious. In the end, the members of our team spent much more time than initially expected preparing code. A lot of time was spent researching the different types of machine learning models that might be best suited for an ordinal classification task, but ultimately we decided to make that portion of our project dedicated to a comparison between three well-known models taking on the same task. The code was straightforward in theory, although we ran into difficulties with implementing spatial joins for the visualizations and the cross-validation portion of the machine learning section. One massive mistake was caught at the very end of preparing the report, which was that we initially assumed a high mean squared error was optimal and erroneously compared it to the scores of the other two models. We had to quickly rewrite parts of the code to search for the lowest mean squared error, and then decide how it fared against the scores of the other two models. Ultimately, our work plan did not anticipate the difficulties we had with actually implementing the code.

Testing:

Testing was largely only required for the machine learning portion. For each type of model and each set of hyperparameters, three trials of training were done whose test accuracies were then averaged to reduce bias. Additionally, the result validity portion of the code served as a high authority on our results, and backed them up as described in the Results section. K-folds are a widely used method of cross-validation, and our dataset had enough observations to support a 5-fold implementation. Thus, we are confident that the results we see are accurate.

Collaboration:

For this project, we did not consult external human resources, but we did use a publicly available tornado dataset. Additionally, for the visualizations of the United States, we used a GeoDataFrame from the course that includes the Polygon objects that represent the shape of the United States and its states. And as with any project involving code, StackOverflow was crucial in smoothing out bugs in our implementation.