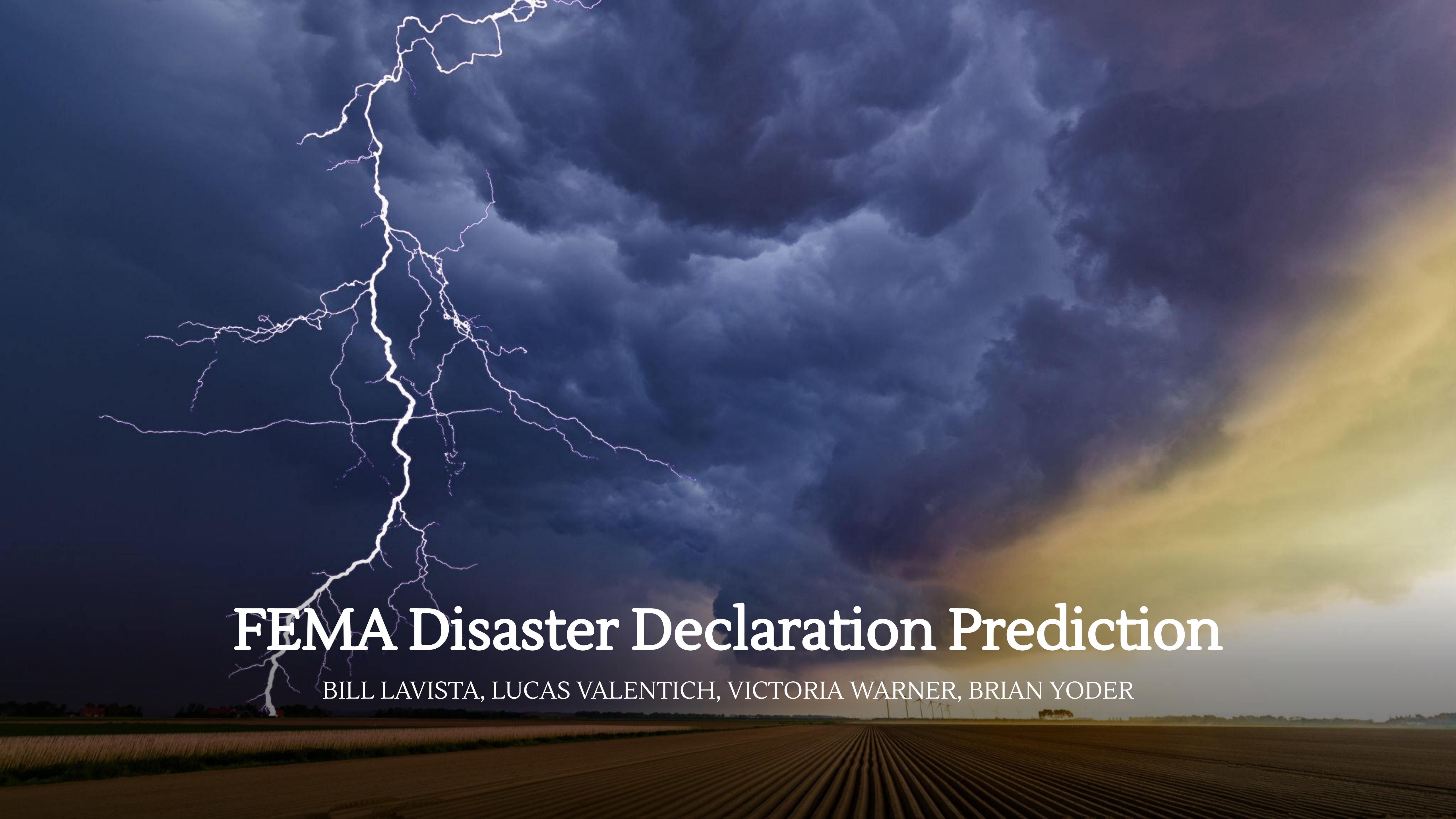
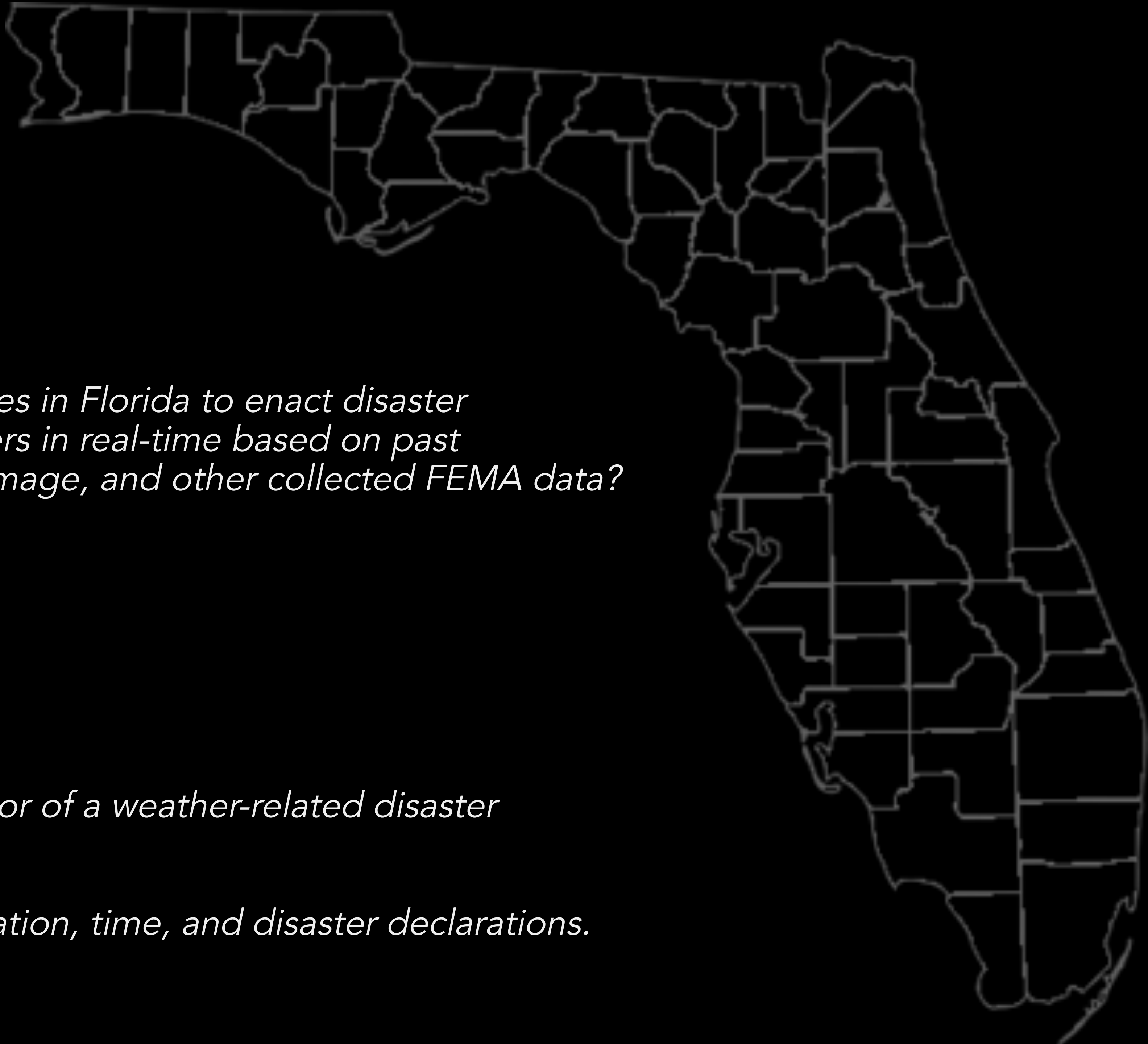


FEMA Disaster Declaration Prediction

BILL LAVISTA, LUCAS VALENTICH, VICTORIA WARNER, BRIAN YODER

Research Question

Can we predict the likelihood of counties in Florida to enact disaster declarations for weather-related disasters in real-time based on past precipitation data, concentration of damage, and other collected FEMA data?



Hypothesis

Precipitation will be the primary indicator of a weather-related disaster declaration in Florida (i.e. hurricanes).

There is a correlation between precipitation, time, and disaster declarations.

FEMA Disaster Declaration Modeling

- Data Ingestion
- Munging & Wrangling
- Data Analysis
- Modeling
- Conclusion

Introduction

Scope Determination

- Narrowing down from the U.S. to the East Coast to Florida

Data Selection

- Choosing the variables to include in building the model from both quantitative and qualitative features

Resolution

- Create a model that can provide useful predictions for FEMA to use
- Provide a visual tool to aid in disaster relief

Data Ingestion

NASA precipitation data and FEMA disaster declaration data was ingested into a Digital Ocean server where it could be accessed, cleaned, and manipulated.

Precipitation Data

- Using a Python script, rainfall data was collected from NASA for the past 20 years in Florida, USA.
- This data included features such as time, latitude, longitude, rainfall, and more.

Disaster Declaration Data

- Disaster declarations in Florida were collected from FEMA.
- This data included incident type, incident date information, state and county codes, designated areas, amounts, and more.

Munging and Wrangling

- Given the historic weather data compressed in file type of NC4, and FEMA disaster declarations in csv
- A database was then created combining all the historic weather data within Florida for the past 20 years. The resolution of the data was scaled down due to storage limitation and computational power
- Cleaning the data consisted of extracting the necessary time and location data (month and county) and aligning it with the disaster and monetary values

Exploratory Data Analysis

Disaster Declarations Over Time

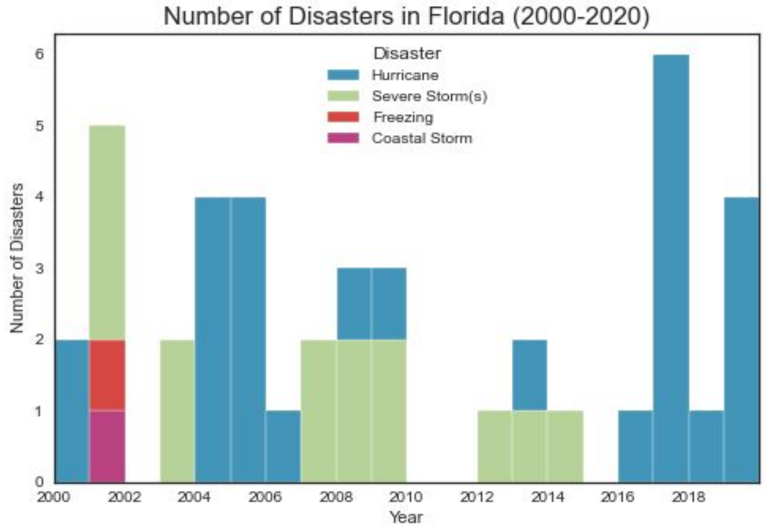
- Hurricane disasters are on the rise over time
- Severe storm(s) are on the decline
- Disaster declarations as a result of freezing and coastal storms are rare in Florida

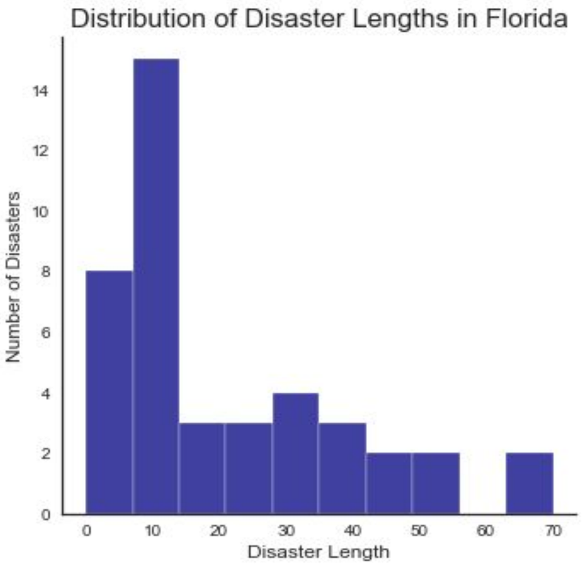
Lengths of the Disaster

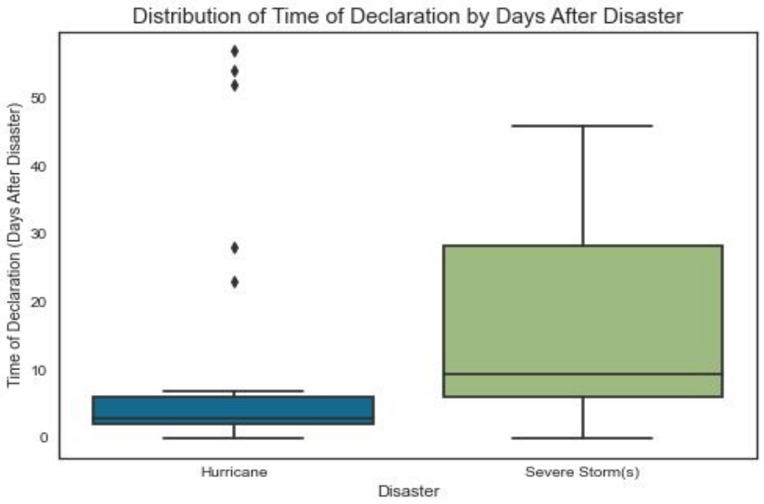
- The average length of a disaster from Incident Begin to Incident End date is 20.83 days
- The data is right-skewed (or with positive skewness) with a median of 13.0 days

Declaration Time Period

- Hurricanes tend to be declared soon after they occur, with the median at 3.0 days after initial incident date
- The distribution of Severe Storm(s) declaration time periods is larger with a median of 9.5 days after incident date







Data Transformation

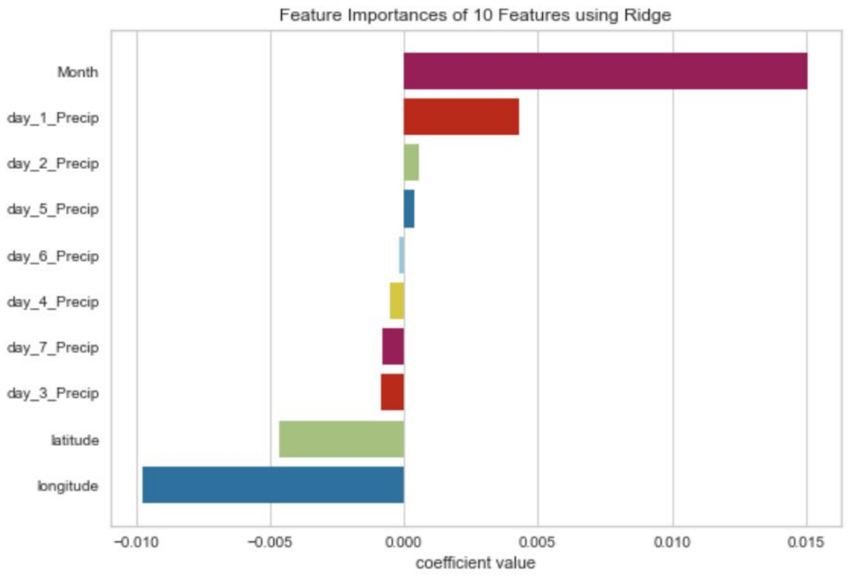
- Extract useable data from dataset
- Create column of the month the precipitation happened
- Prepare column to merge (county plus time)

- Remove non-precipitation related disasters
- Create binary variable for disaster/ non-disaster days.
- Prepare column to merge (county plus time)

Feature Analysis

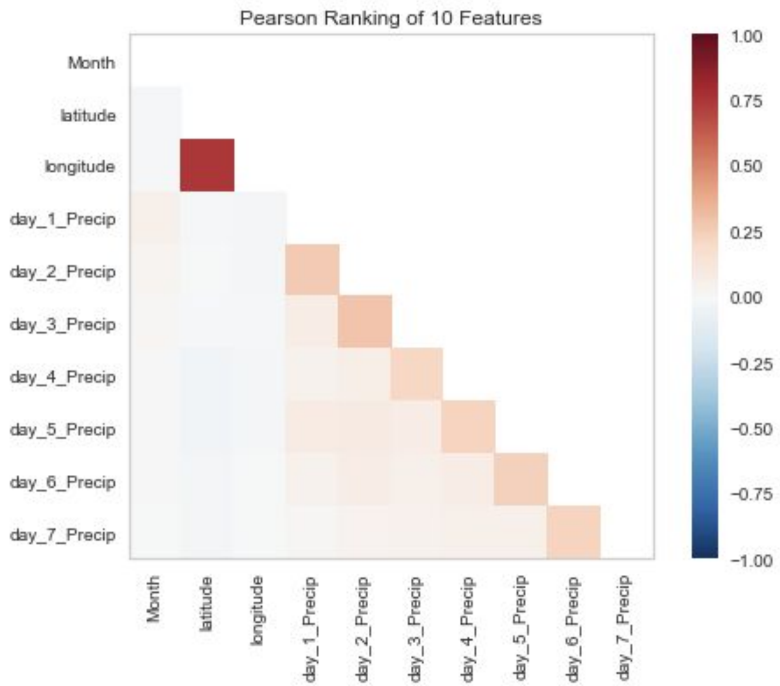
Feature importances

- Most relevant feature “Month”
- Not relevant, but important to our study “latitude” and “longitude”



Rank2D

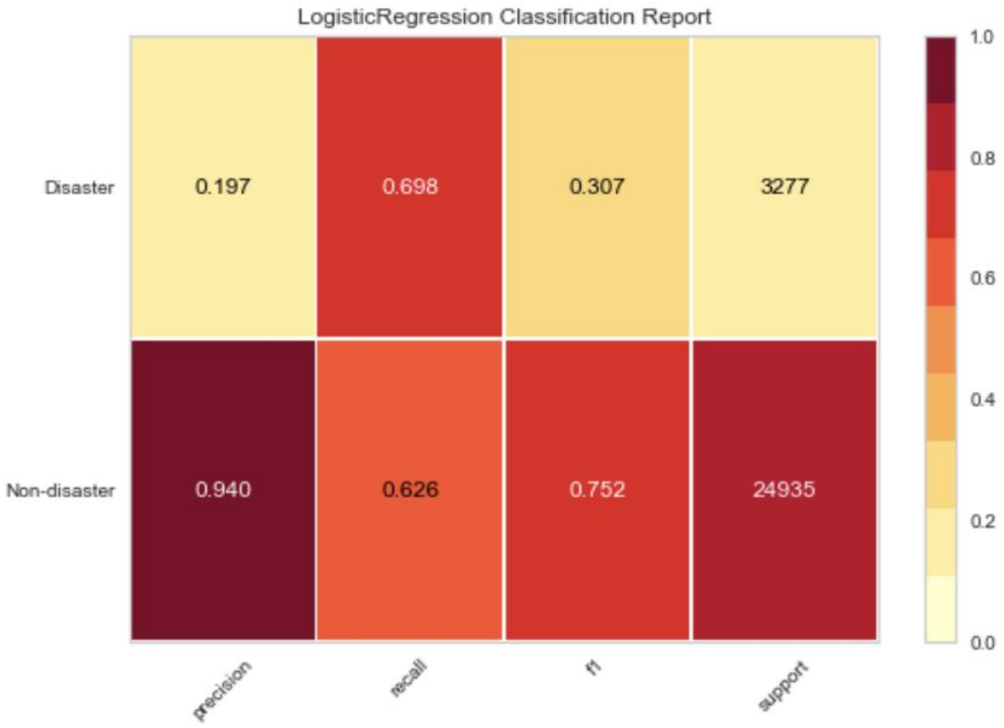
- There is a strong correlation between latitude and longitude.
- There is a soft correlation between precipitation of the next day.

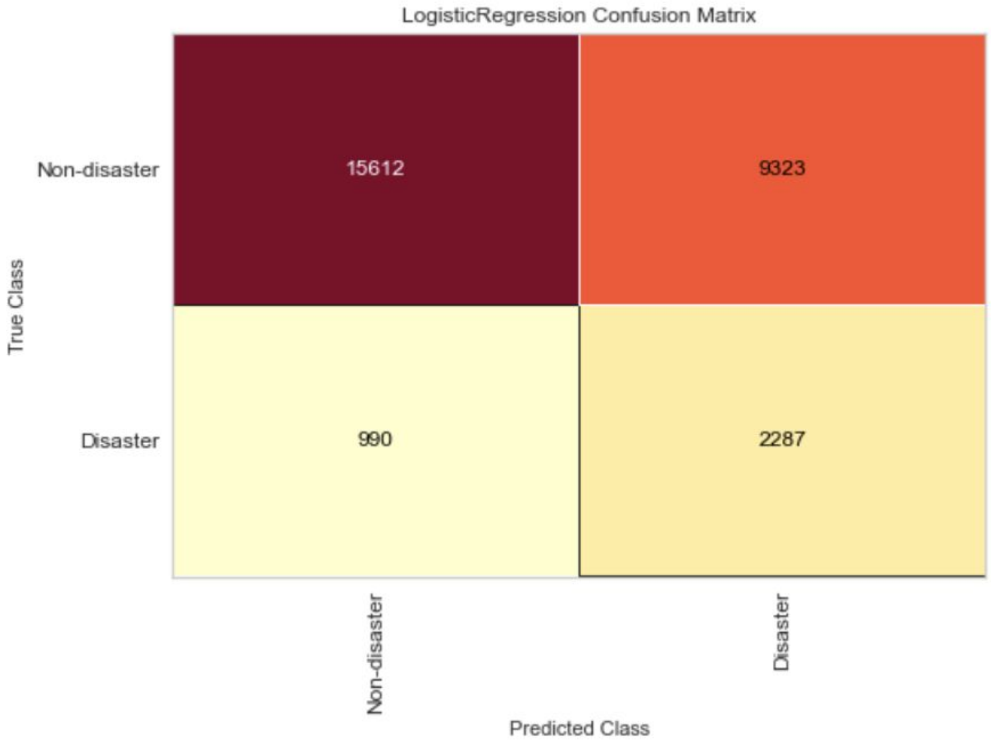


Logistic Regression

```
logreg = LogisticRegression(class_weight="balanced")
```

```
Accuracy of Decision Tree classifier on training set: 0.63
Accuracy of Decision Tree classifier on test set: 0.63
```

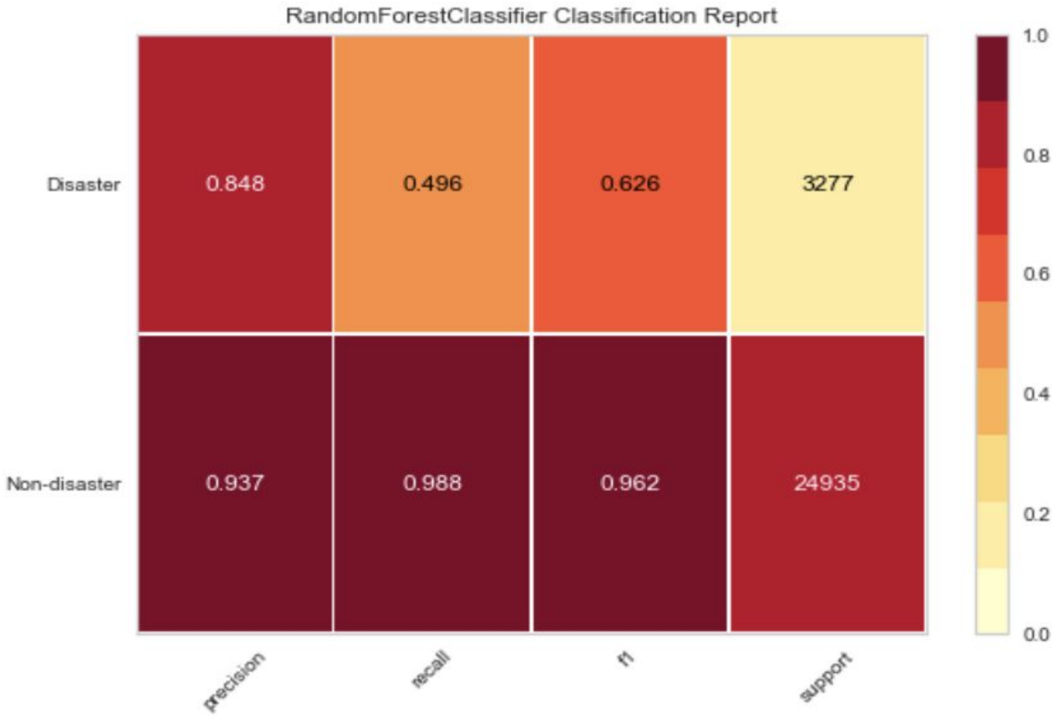


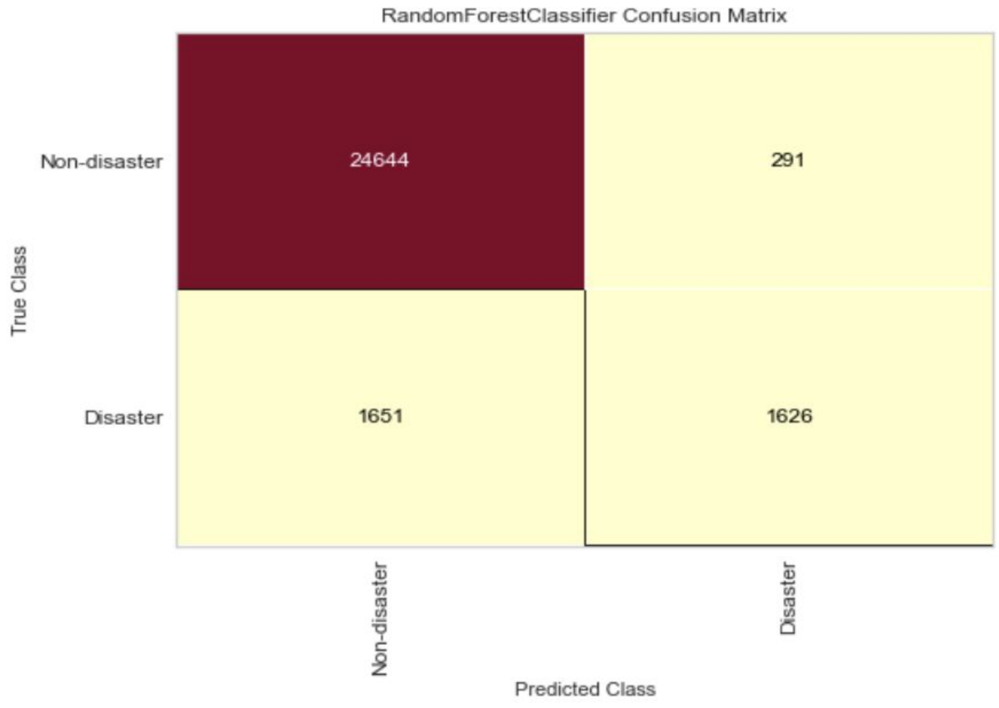


Random Forest

```
forest = RandomForestClassifier(random_state=0, n_estimators=3, max_depth=12)
```

```
Accuracy of Decision Tree classifier on training set: 0.943
Accuracy of Decision Tree classifier on test set: 0.931
```

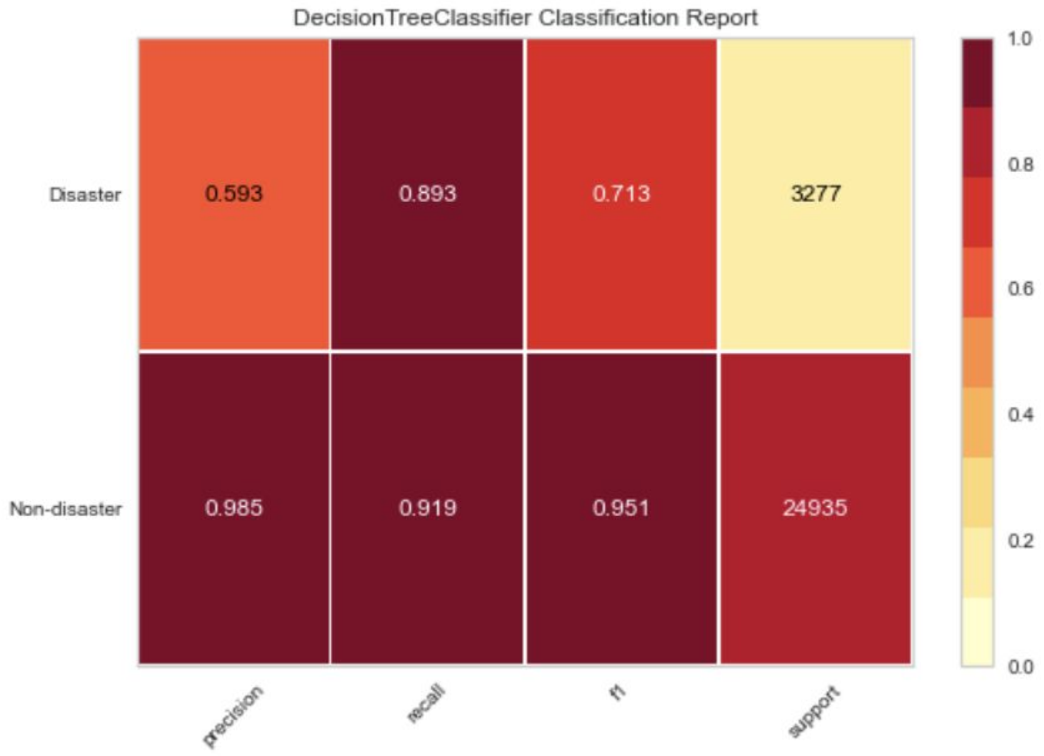


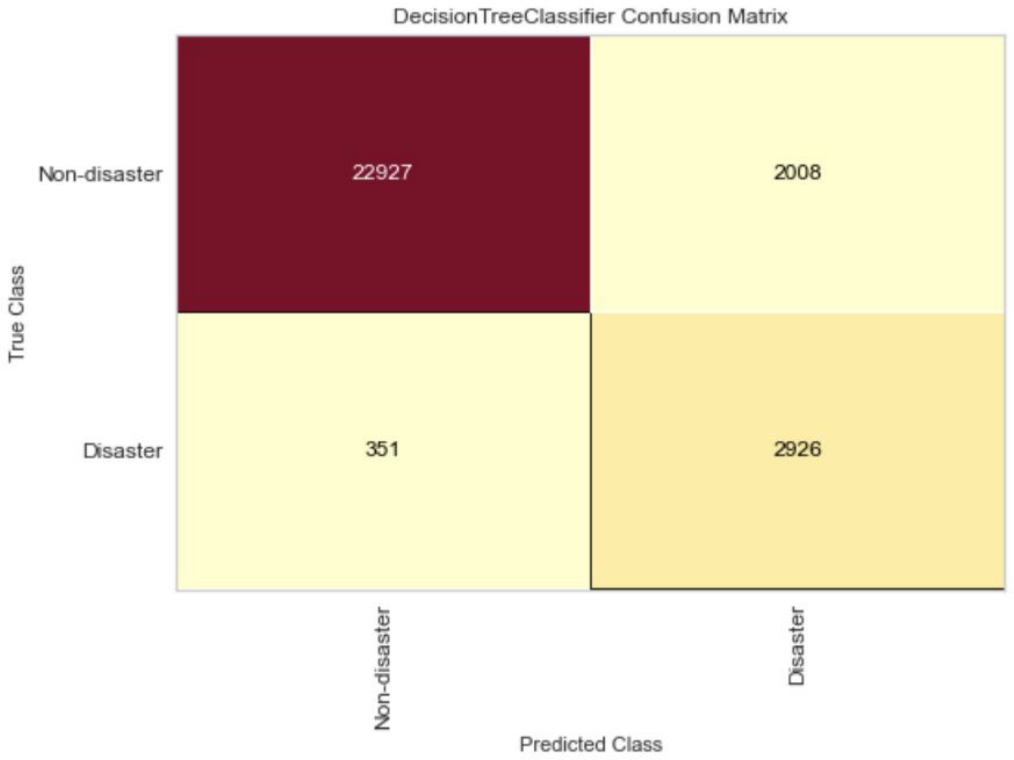


Decision Tree

```
clf = DecisionTreeClassifier(max_depth=18, class_weight="balanced")
```

```
Accuracy of Decision Tree classifier on training set: 0.94
Accuracy of Decision Tree classifier on test set: 0.92
```

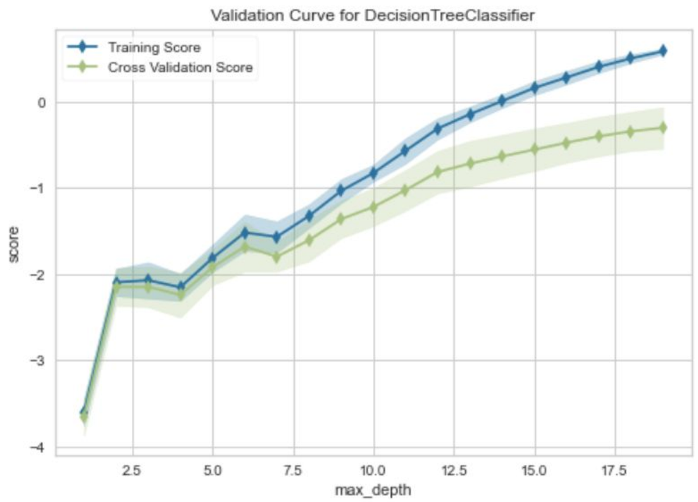


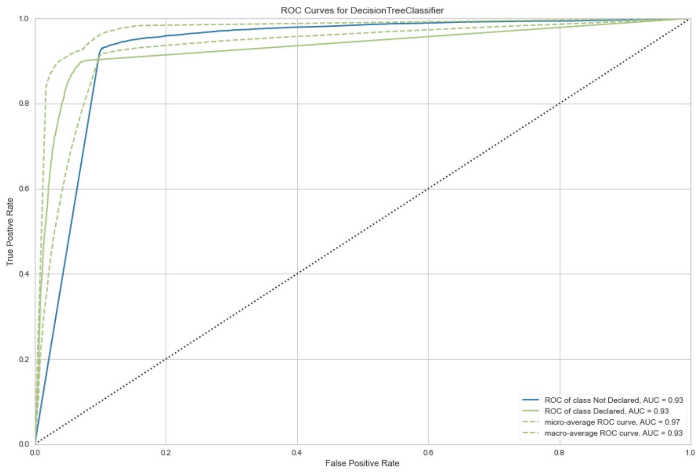


Model evaluation

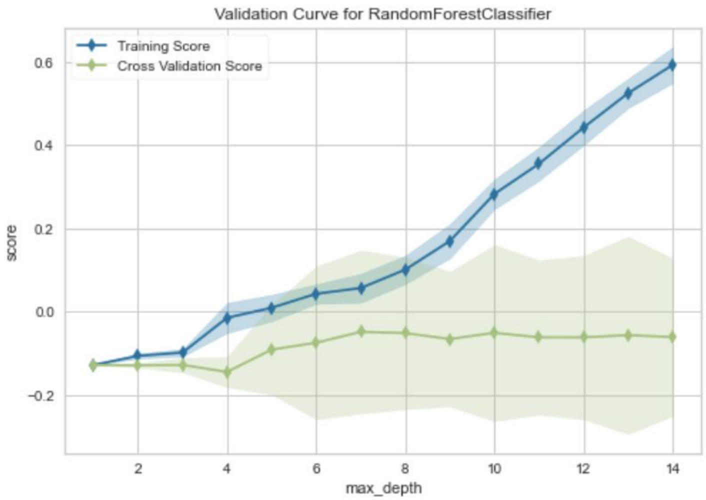
Decision Tree	Random Forest

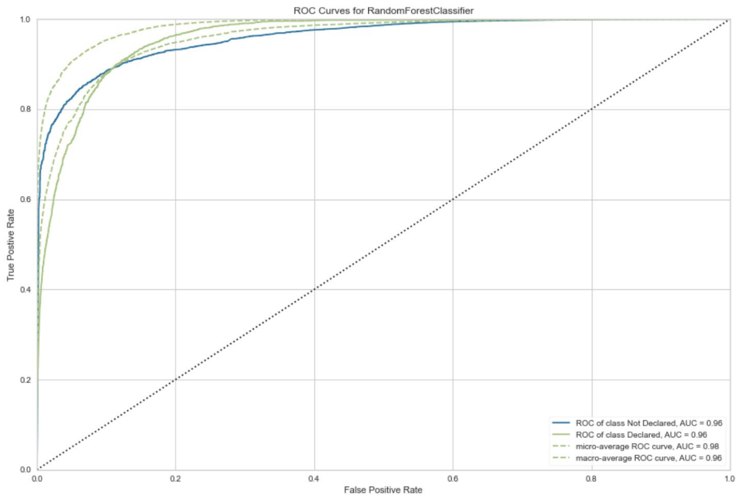
Decision Tree





Random Forest





Individuals and Households Program Declaration

The IHP Housing Assistance Provision includes both financial housing (i.e. rental assistance, home repair) and direct housing assistance (i.e. temporary housing units, permanent housing construction).

Model Creation

1. **Feature Creation**
   Calculated the length of the disaster and the number of days after the start of the disaster for the official declaration

2. **Dataset Merge**
   Retrieved the NASA precipitation data based on county and the initial Incident date (beginning of the disaster)

3. **Data Transformation**
   Used a Label Encoder to assign a unique value to each of the Disaster Types (i.e. hurricane, severe storm(s), coastal)

Feature Selection

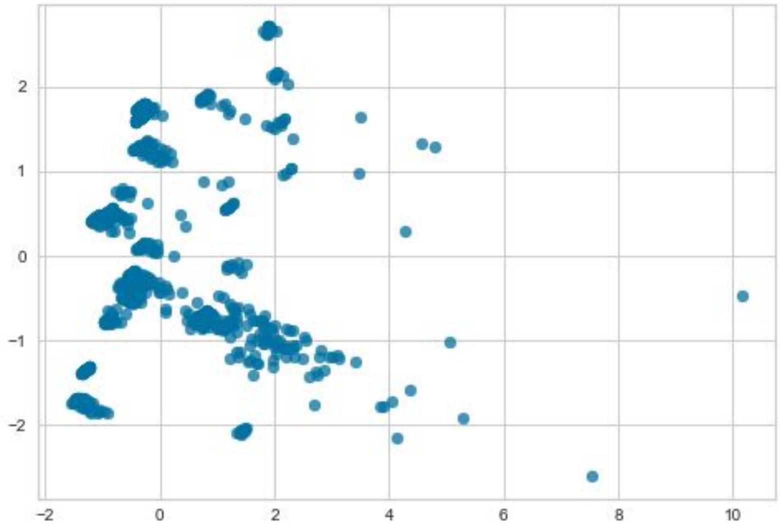
Top features given regularization techniques and transformer models

County

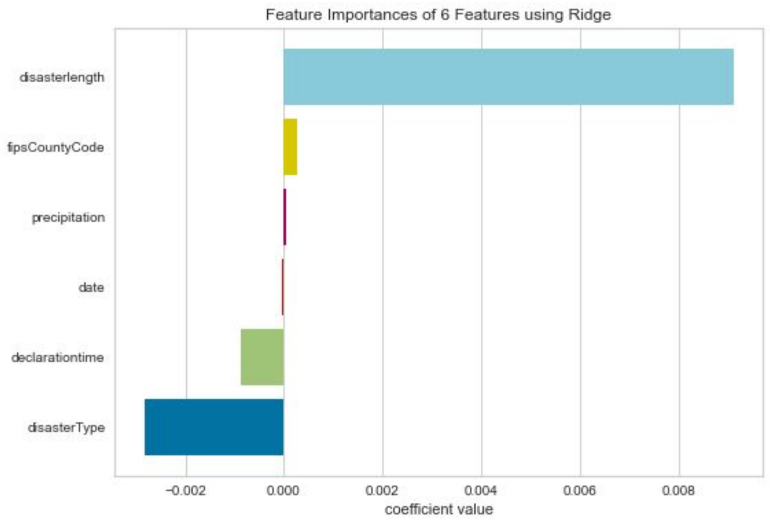
Disaster Type & Precipitation

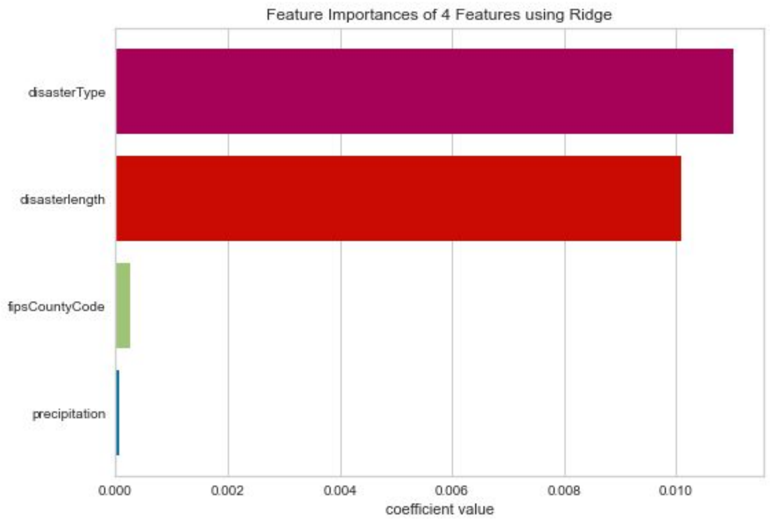
Length of Disaster

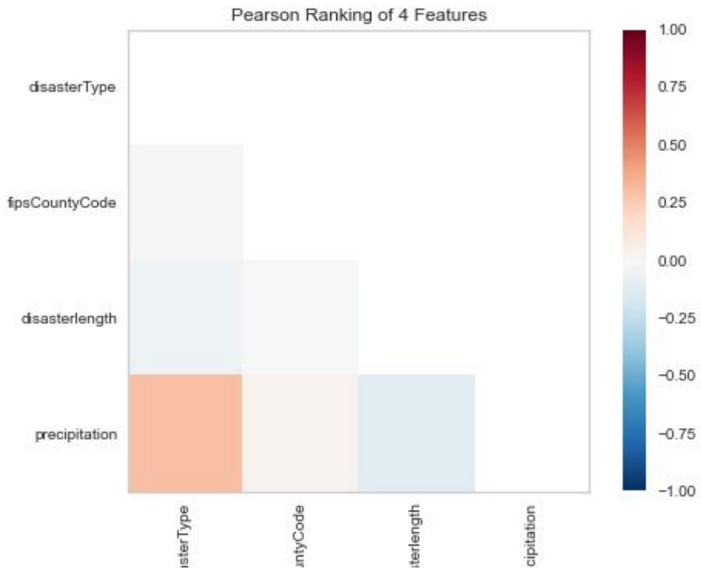
Principal Component Analysis



IHP Declaration Feature Selection

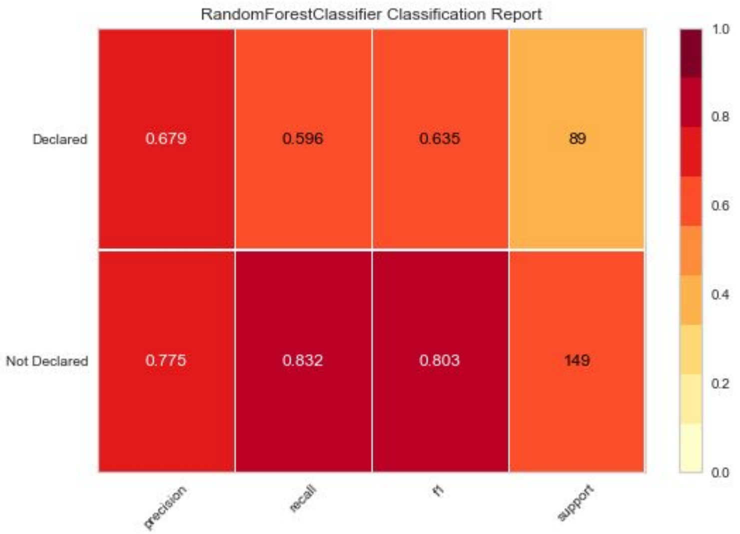


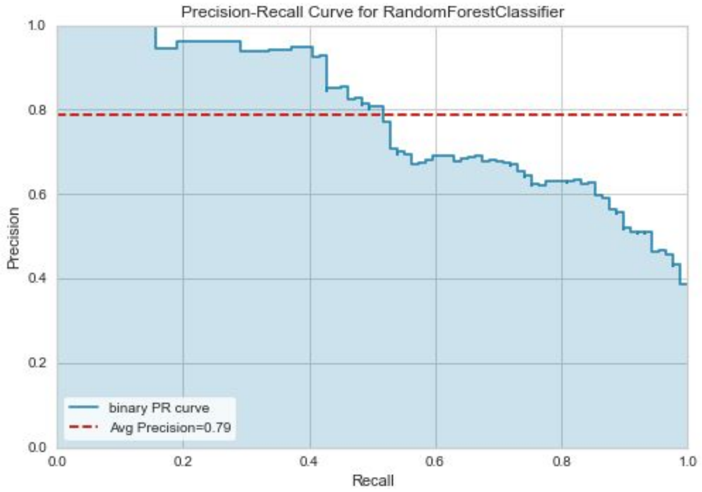




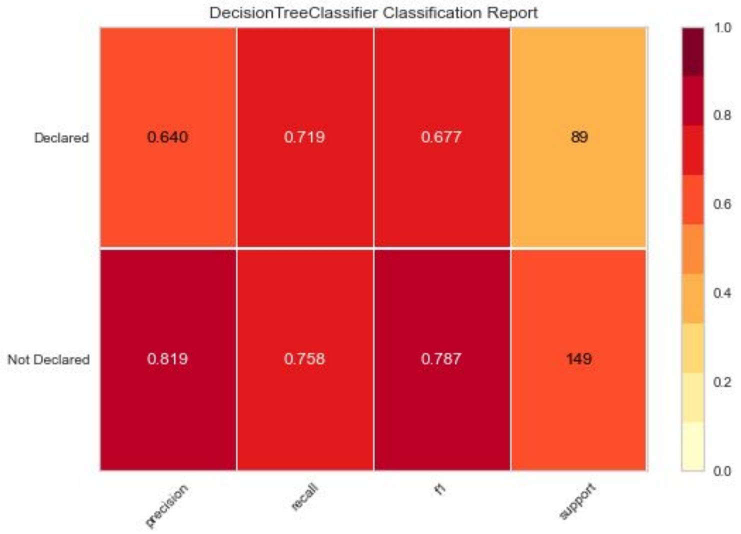
IHP Declaration Prediction

Random Forest

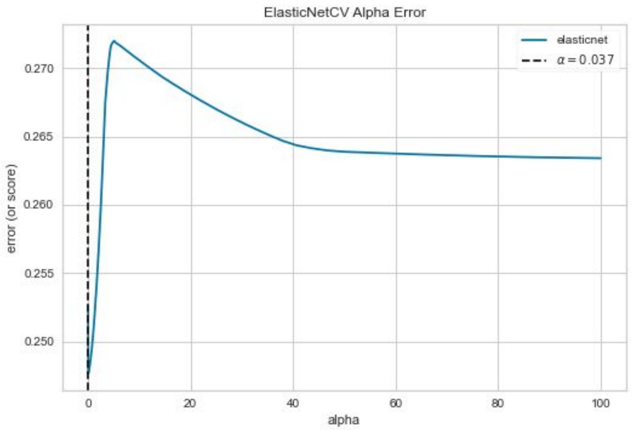


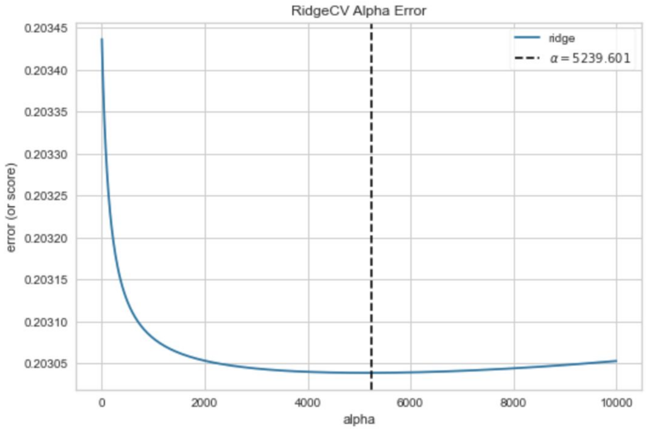


Decision Tree



Hyperparameter Tuning

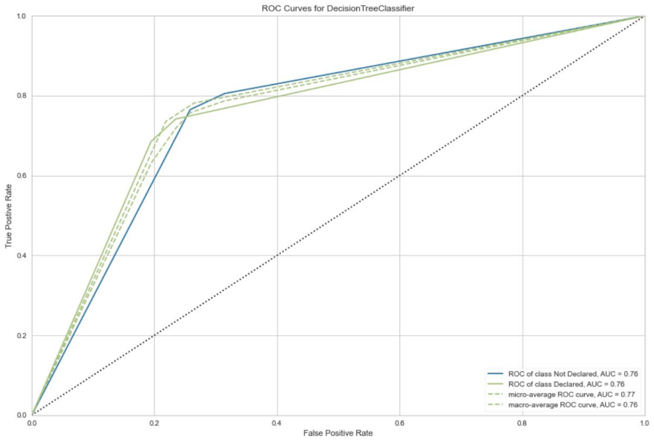




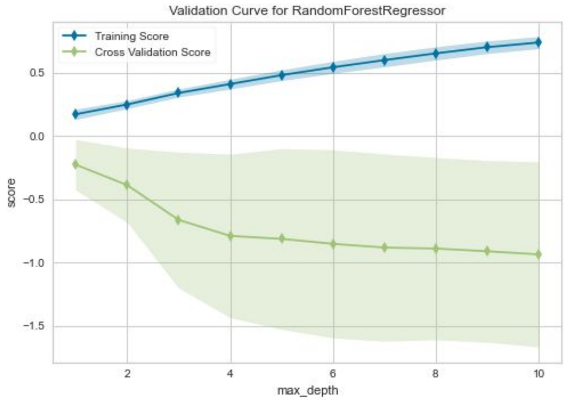
IHP Declaration Prediction

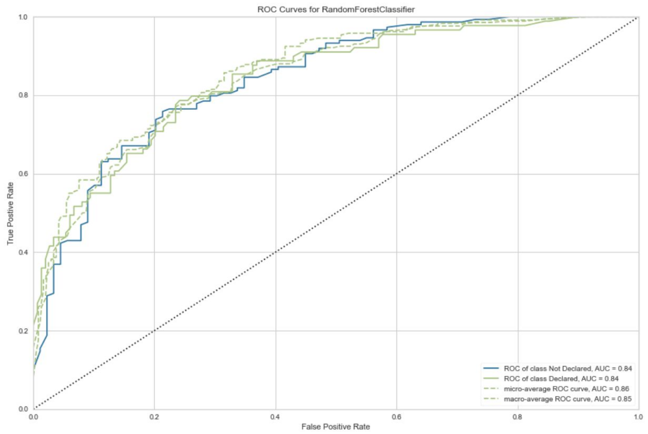
Model Evaluation

Decision Tree



Random Forest





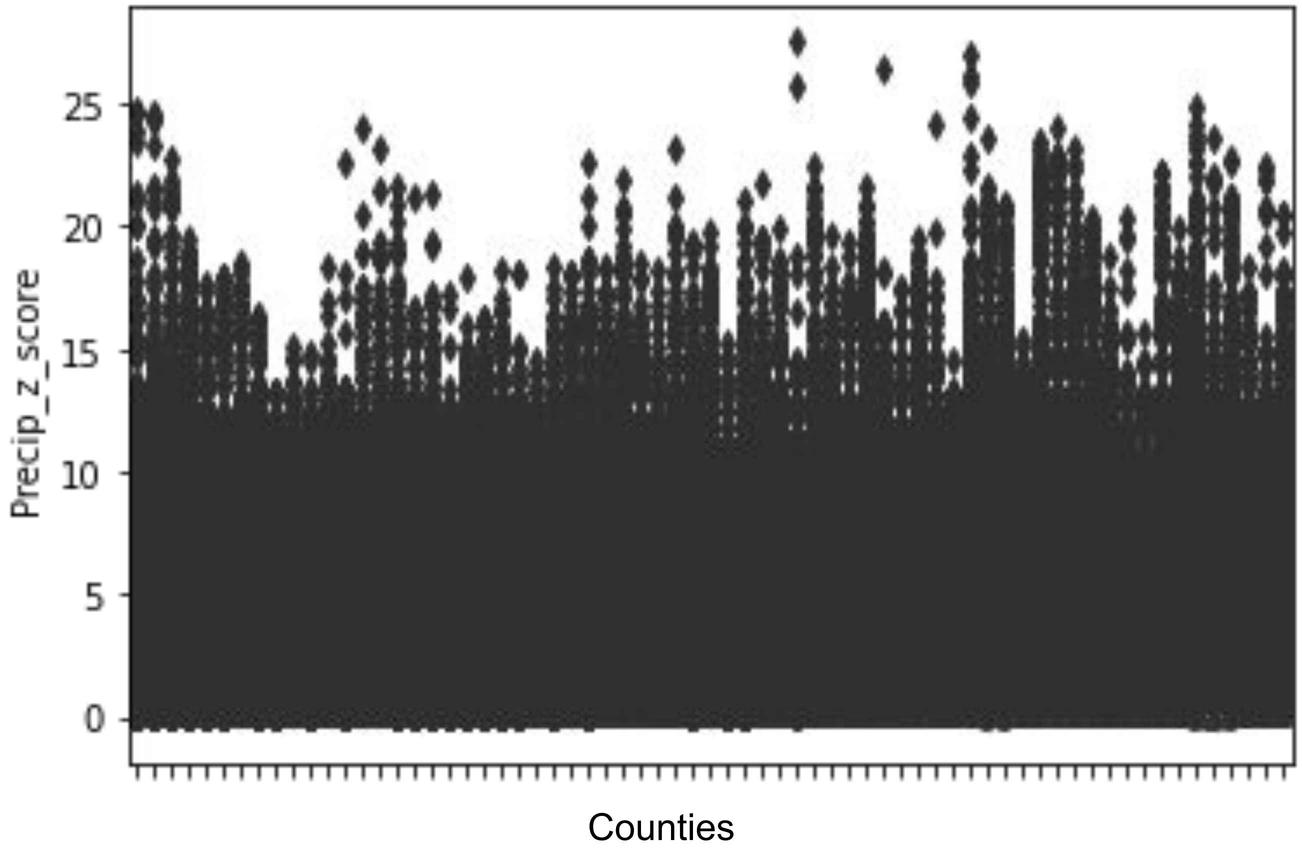
Next Steps

Expand Scope
Include the East Coast or entire US in this prediction model

Non-Random Residuals
Identify missing variables that will eliminate patterns in the residuals

Model Overfit
Tune the model more through cross validation, expand scope, ensure no overlap in train and test data

Anomaly Detection



Reviewed available libraries for conducting anomaly detection.

Ended up writing code to convert precipitation values for each lat / lon to a z-score to detect anomalous levels of precipitation.

Used different z-score cut-offs and approaches to re-coding scores to see which produced the best model scores.

Models using identified anomalies performed as well or worse than precipitation.

Takeaways

Data Storage

- Having enough drive space for the minute analysis took up more space than expected
- Organizing the data appropriately made modeling easier

Class Imbalance

- Select a random sample of non-disaster days around three times the number of disaster days.
- Use a parameter class_weight = "balanced", to balance the data.

Feature Selection

- Changing the data from dates to months helped the model identify seasons
- 7 day averages of Precipitation yielded best results
- Counties were replace by geographical for geographical sizing

Data Product

134.122.119.130:5000

Specification for final product

- Visualization of disaster declarations
- Accessible through mobile devices
- Real Time Analysis of weather data

Questions?