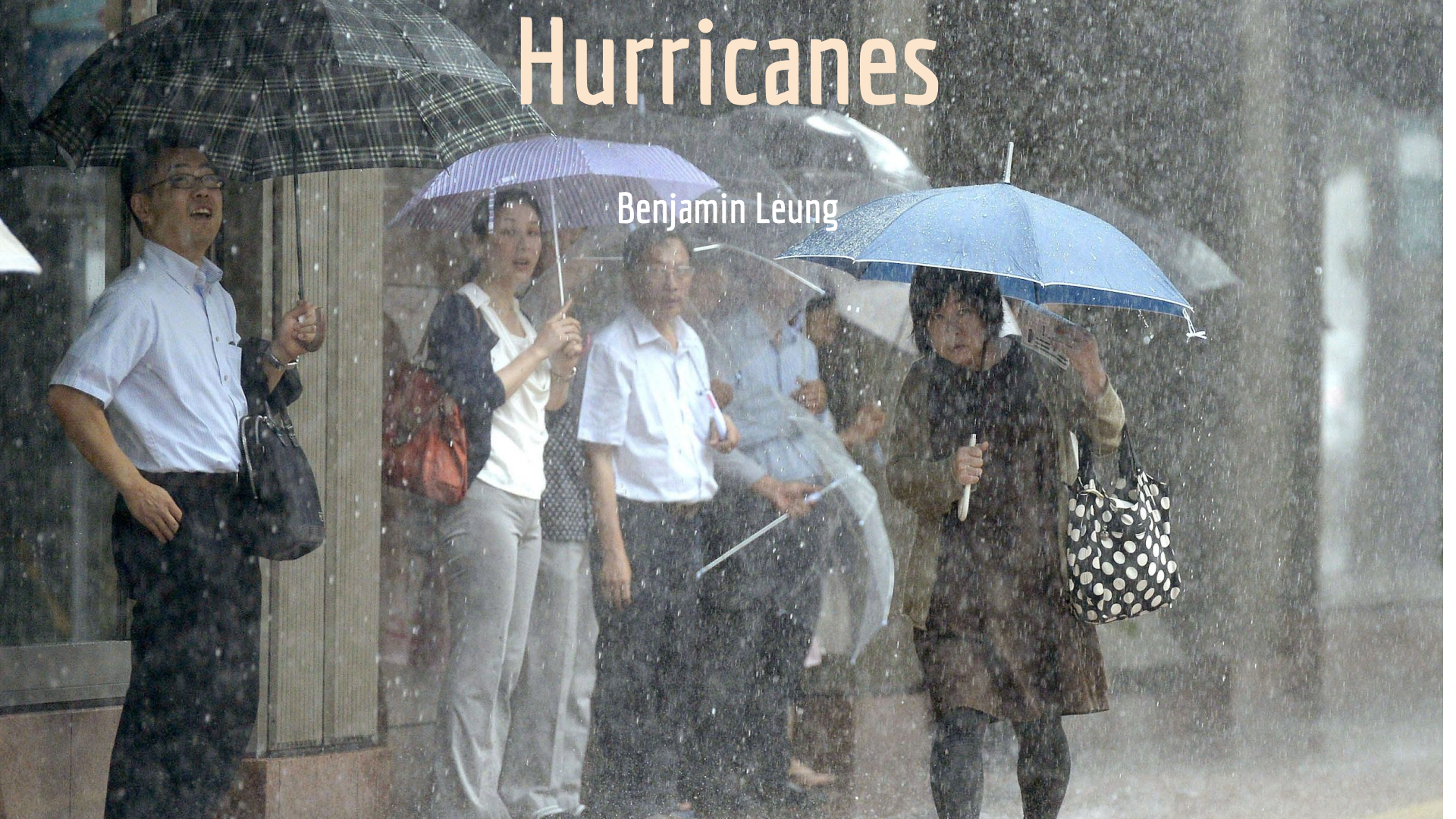
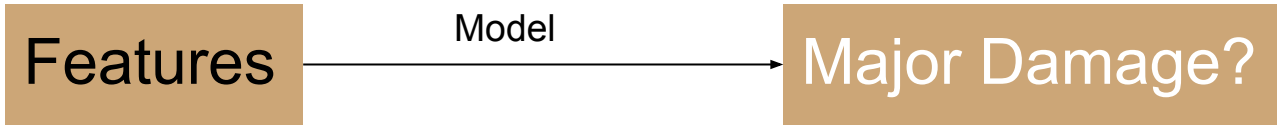
# Hurricanes

Benjamin Leung



# Objective

- Simplify: Determine if major damage will occur
- Good preliminary screening tool for citizens and observatories

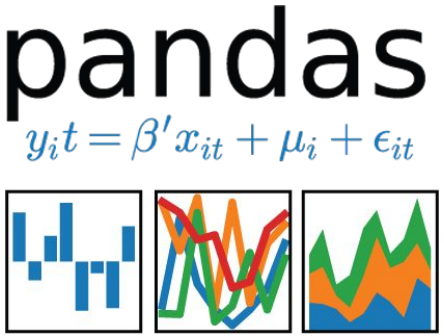



# Tools


jupyter


python


NumPy


pandas
$y_i t = \beta' x_{it} + \mu_i + \epsilon_{it}$


matplotlib


Seaborn


scikit
learn


Flask

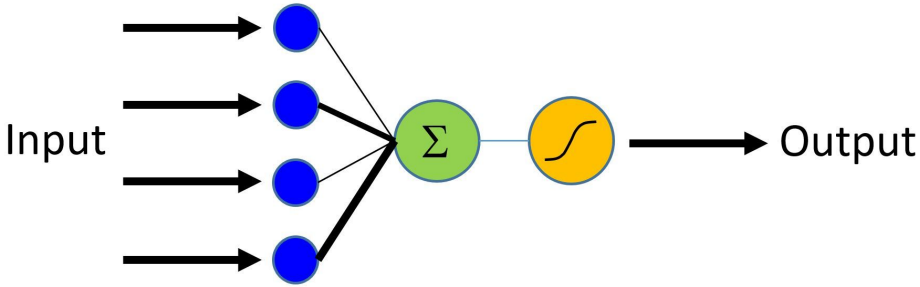
# Target and Features

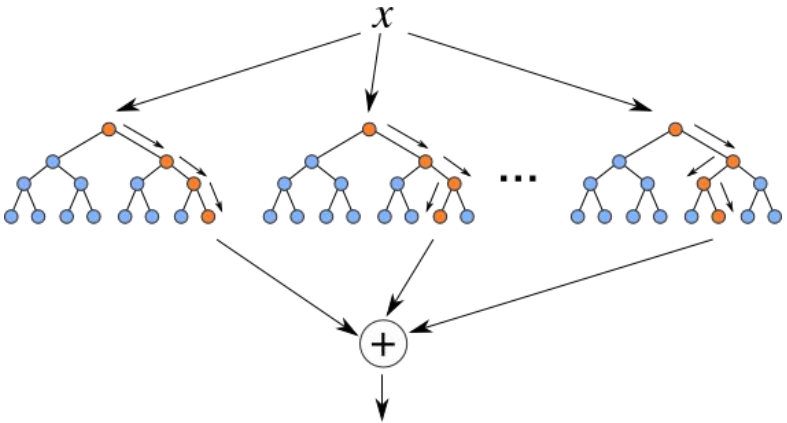
**Damage Score (USD) = Property Damage + 9M x Deaths**

**CUTOFF = 1M USD**

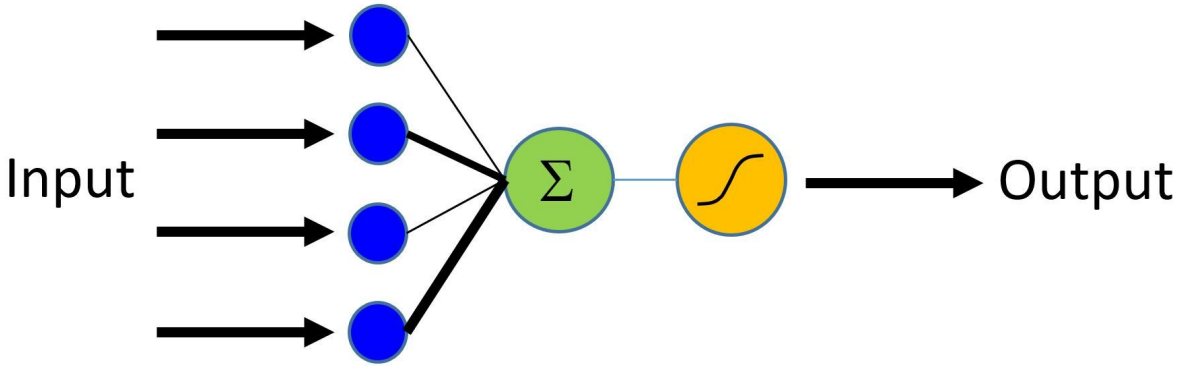
- Features:
  - Year, Month, Temp, Precipitation, Wind Speed, Latitude, Longitude, State, System Type

# Models Used

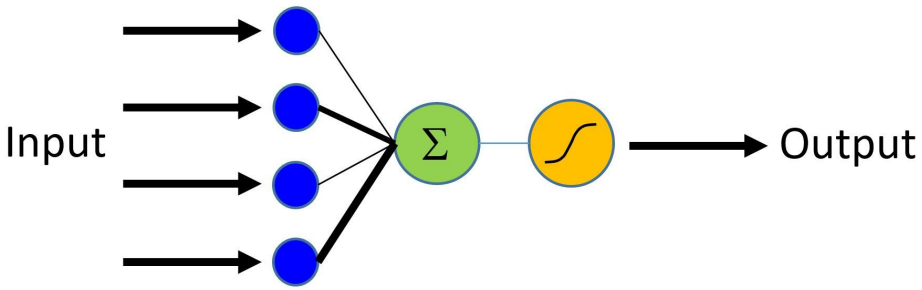


# Logistic Regression

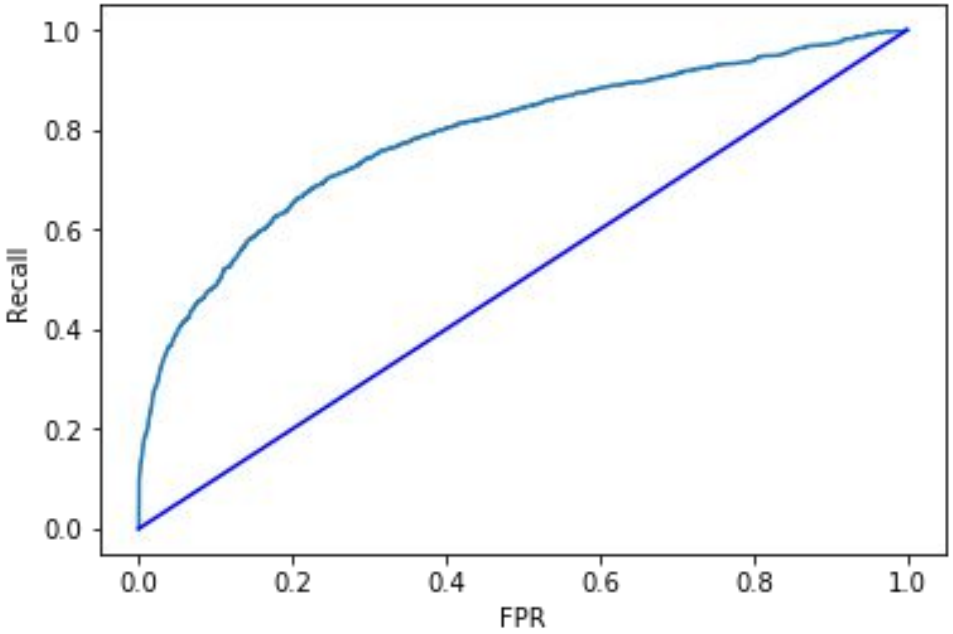


# Optimized Scores and Parameters

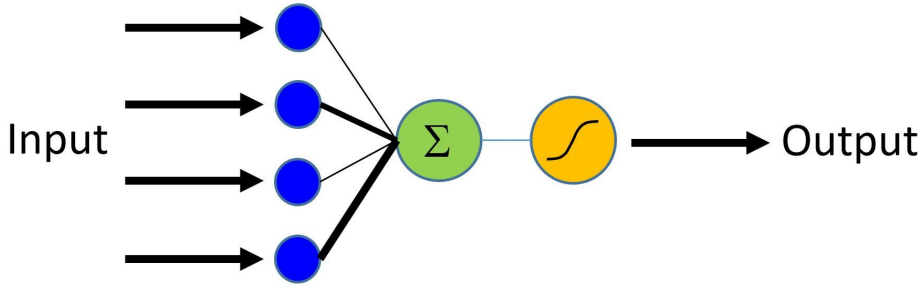



| Optimized Score |
| --- |
| F2 Score |
| F1 Score |
| F1 Score |
| Recall |
| Recall |

# Logistic Regression ROC Curve



# Logistic Regression Highlights

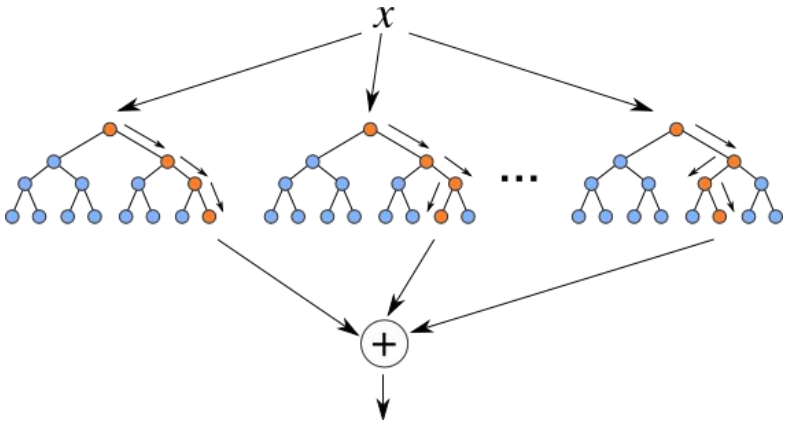



| Optimized Score | Balanced Weighting? | | |
|---|---|---|---|
| F2 Score | No | | |
| **Recall** | **Precision** | **AUC** | **Threshold** |
| 0.46 | 0.10 | 0.79 | 0.0379 |
| 0.29 | 0.20 | 0.79 | 0.065 |

# Coefficients/Features of Note

| Feature | Coefficient | Comments |
|---|---|---|
| Wind Speed | 0.05 | Lower than expected |
| Most Continuous Features | | Lower than expected |
| Louisiana | 1.10 | Unfortunate |
| Storm Type (eg. Hurricane) | >> 1 | As Expected |

# Random Forest Highlights

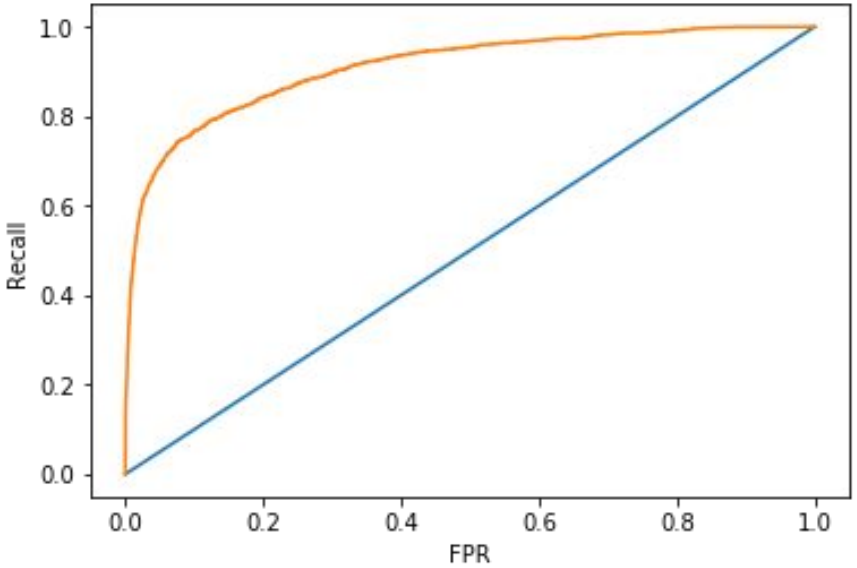


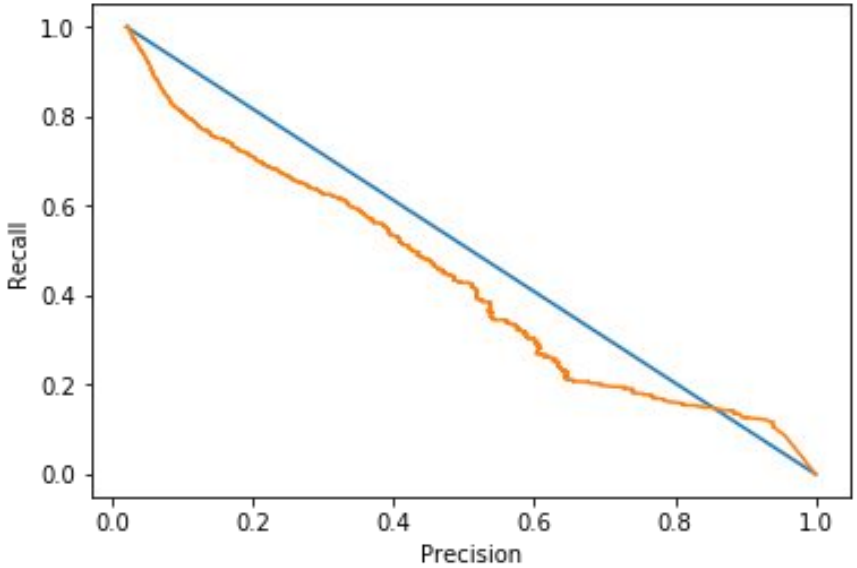

| F2 Score | | | 200 Trees |
|---|---|---|---|
| Recall | Precision | AUC | Threshold |
| 0.65 | 0.24 | 0.91 | 0.5 |
| 0.50 | 0.40 | 0.91 | 0.64 |

# Most Important Features

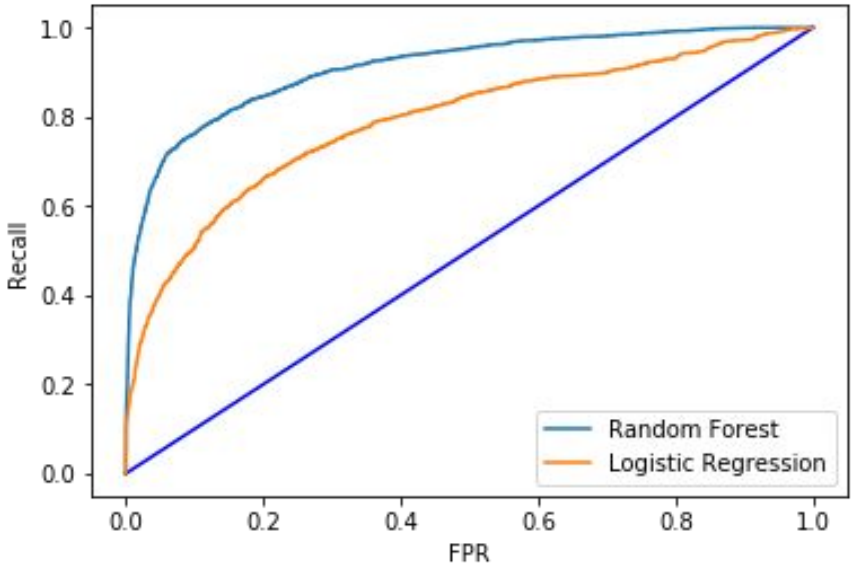

| Feature | Importance |
|---|---|
| Wind Speed | 0.264061 |
| Year | 0.133199 |
| Longitude | 0.104564 |
| Latitude | 0.099290 |
| Temperature | 0.073899 |
| Precipitation | 0.070128 |

# ROC Curve and Precision-Recall Curve



# ROC Curve Comparison Between Models



# Metric Comparison Between Models

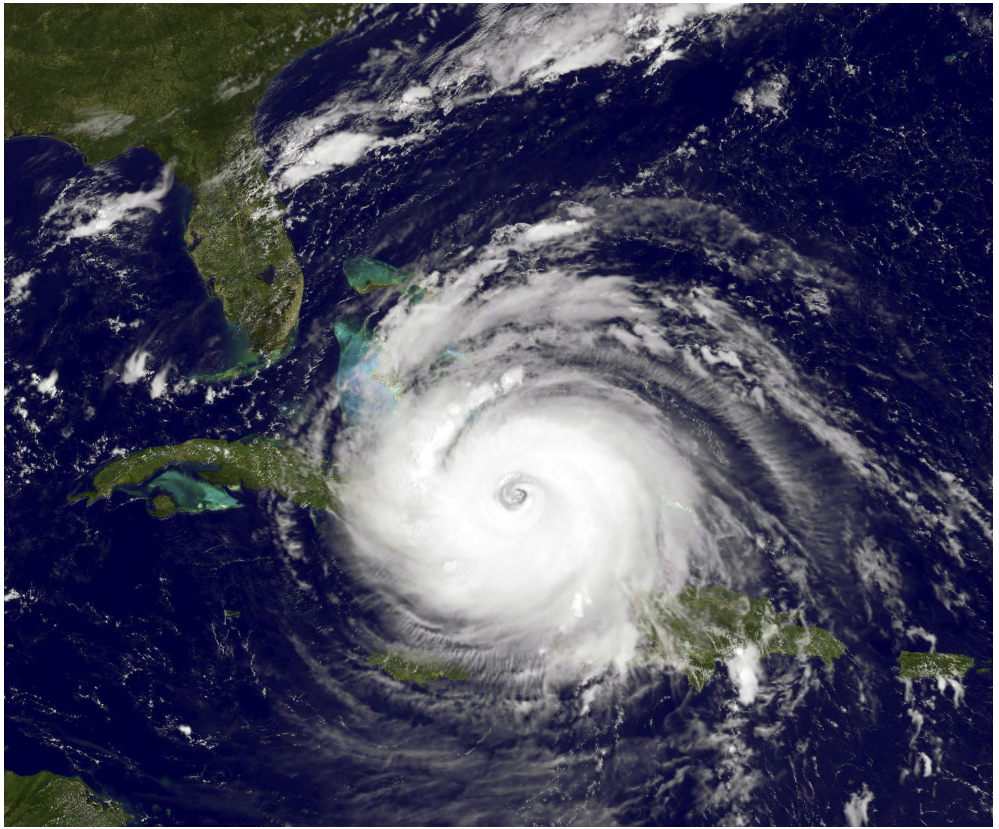
| Logistic Regression | | | |
|---|---|---|---|
| Recall | Precision | AUC | Threshold |
| 0.29 | 0.20 | 0.79 | 0.065 |
| Random Forest | | | |
| Recall | Precision | AUC | Threshold |
| 0.50 | 0.40 | 0.91 | 0.64 |

# Final Chosen Model

| Logistic Regression | | | |
|---|---|---|---|
| Recall | Precision | AUC | Threshold |
| 0.29 | 0.20 | 0.79 | 0.065 |
| Random Forest | | | |
| Recall | Precision | AUC | Threshold |
| 0.50 | 0.40 | 0.91 | 0.64 |

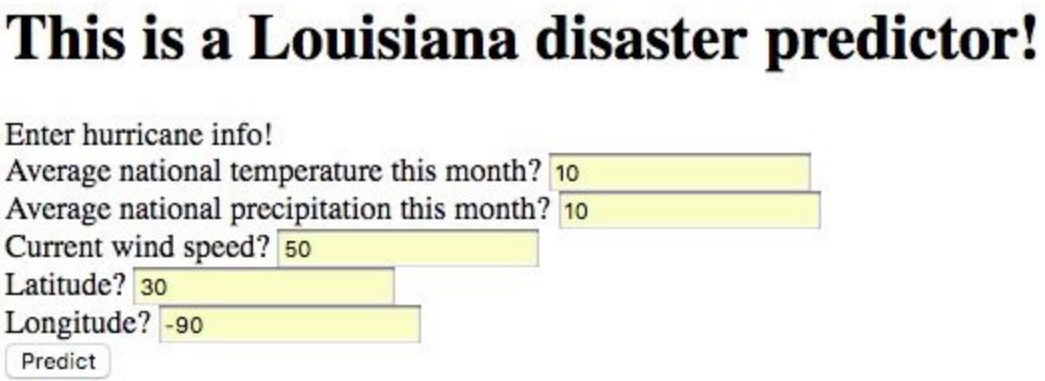
# Next Steps

- More features
- More models
- Better data
- Hurricane path prediction
- Better cutoff for labeling

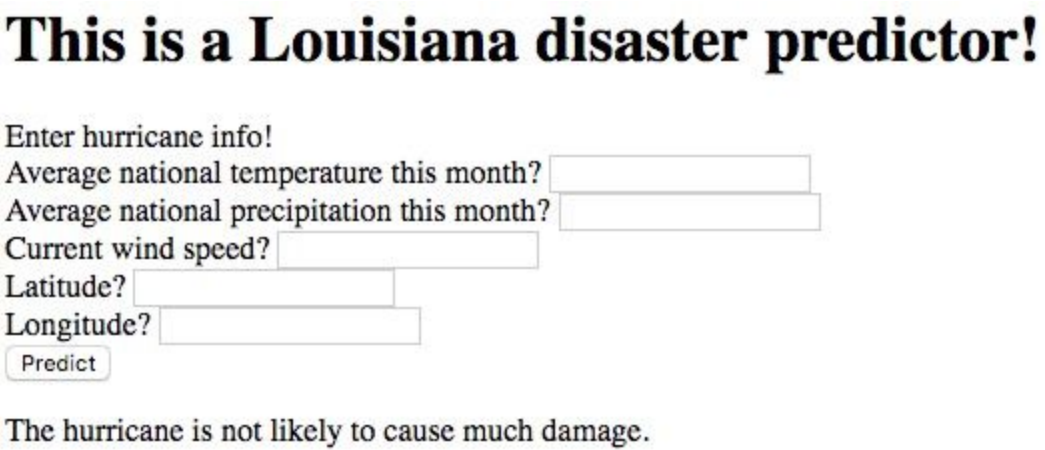



Thank
you!

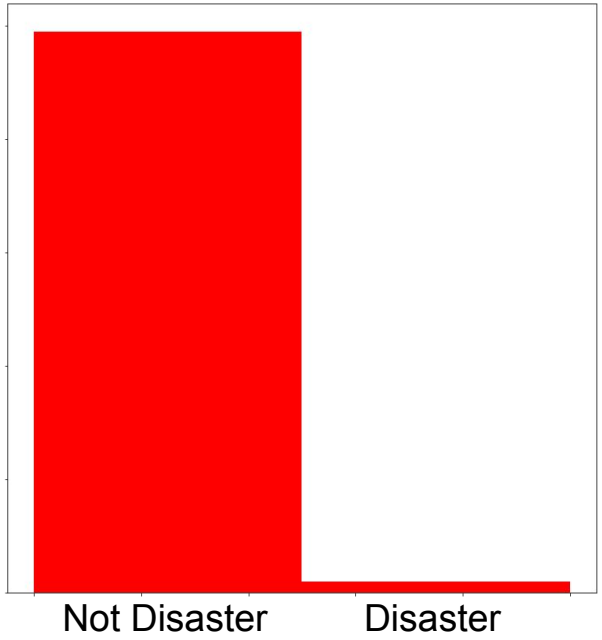
# Flask Backup

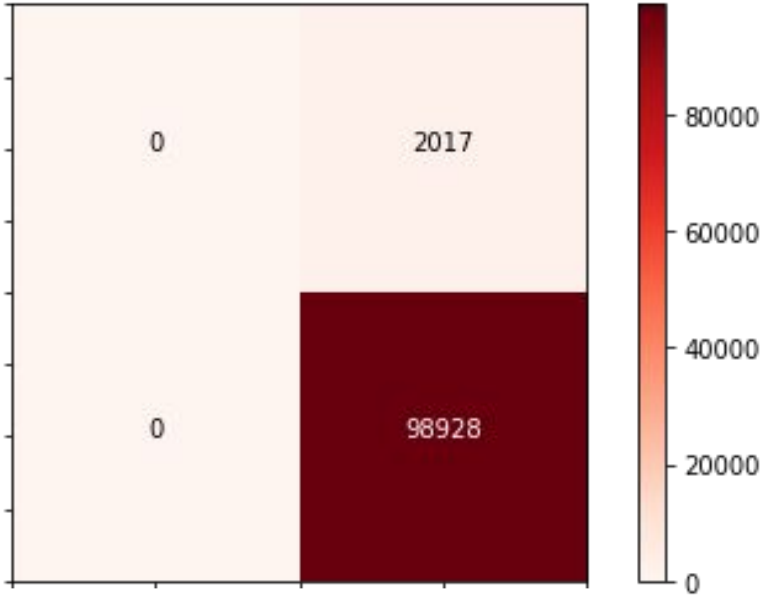


# Flask Backup

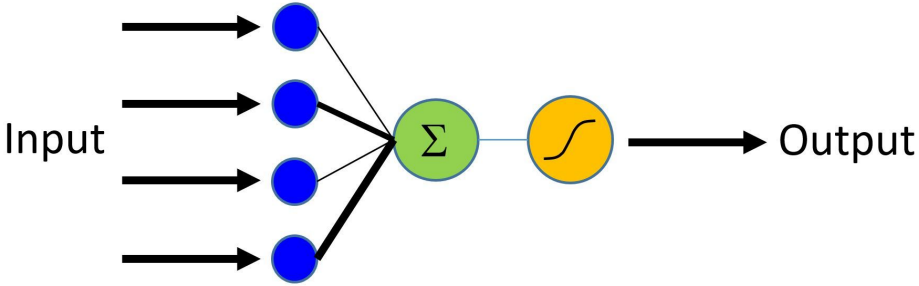


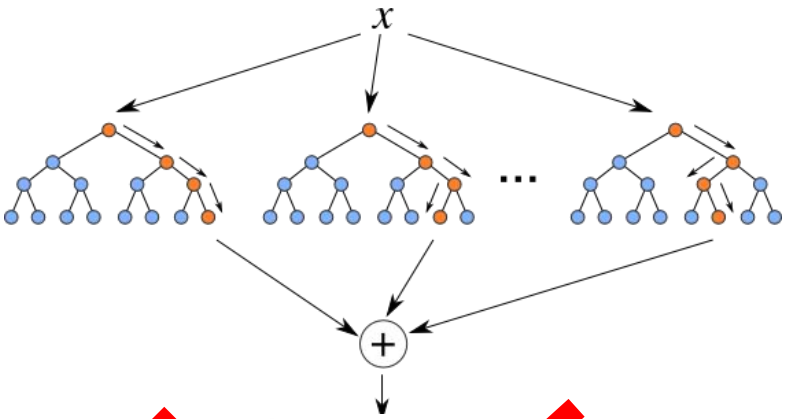
# Classes Very Imbalanced!



# Models Used




$$P(A \mid B) = \frac{P(B \mid A)P(A)}{P(B)}$$