

# DISASTER MESSAGE CLASSIFIER

**IMPROVING ACCESS TO CRITICAL INFORMATION  
DURING CRISIS**

**SUELEM LEE  
OCTOBER 19, 2021**

## Project Overview

Natural disasters affect almost every part of the world. In 2018, Indonesia faced the highest number of deaths in the world, due to the earthquakes and tsunami that occurred in September. In the United States that year, most of fatalities from natural disasters came from tropical cyclones, wildfires, heat, and drought.

Social media is being explored as tool for disaster management by developers, researchers, government agencies and businesses. The disaster-affected area requires both cautionary and disciplinary measures. The need for decision-making system during emergencies and in real time poses problems classifying emergencies.

## PROBLEM STATEMENT

**Can we explore Social Media tools during times of crisis to help people make more informed and better decisions?**

## DATASET

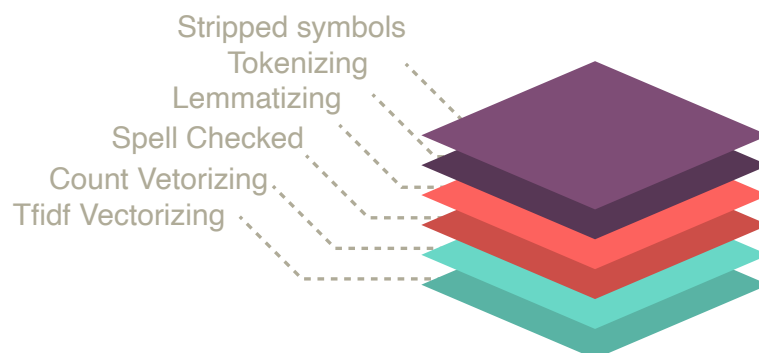
This dataset available at Kaggle provided by Figure Eight, contains 15,000 messages drawn from events of earthquakes, floods, hurricanes and news articles spanning a large number of years and 100s of different disasters. Each target was already binarized. Since it was an overwhelming number of target variables I've selected a few to prioritize when comparing scores.

15000 messages  
36 Target Variables

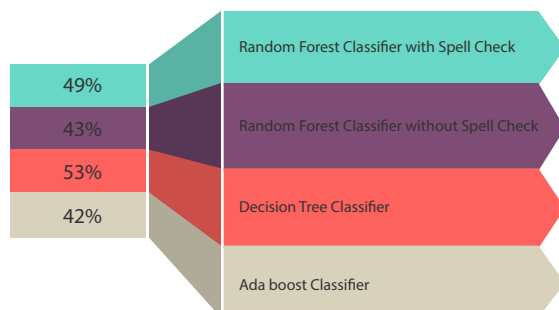


## MULTIOUTPUT CLASSIFIER

### Preprocessing



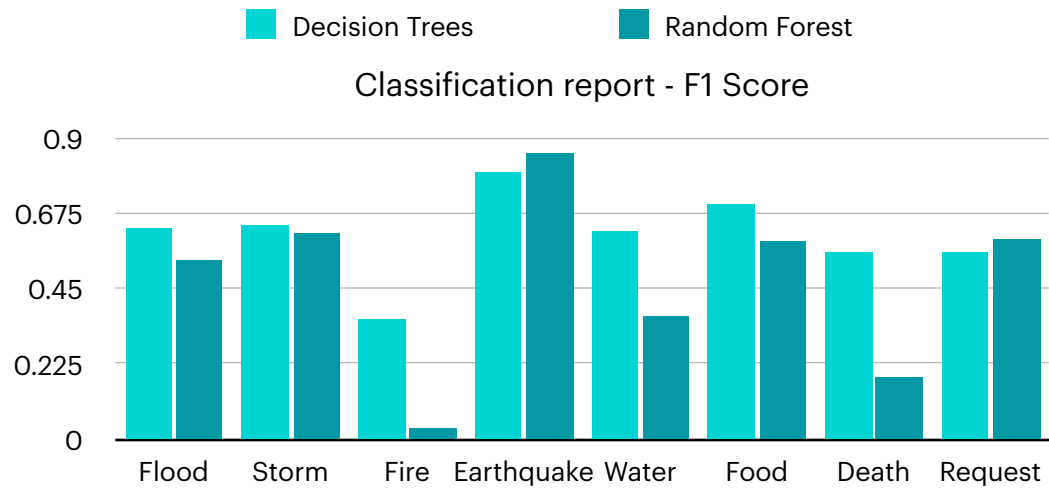
### Model Evaluation



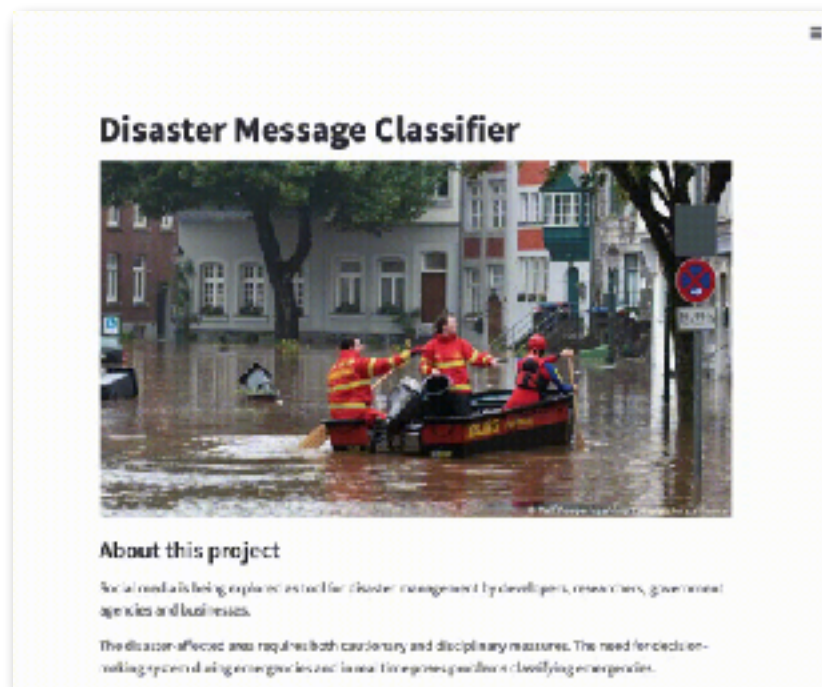
I chose evaluating for micro average, since my data is pretty unbalanced the micro-average will adequately capture this class imbalance, and bring the overall precision average down.

Although my Decision Tree Classifier has a better

Micro Average F1 Score, Random Forest Performed better in more features, between other features too. I would pick the appropriate model according to your preference of category.



## APP



## CONCLUSIONS

Attempt to retry Resampling, under-sampling and BalancedBaggingClassifier to normalize data.

Attempt for class\_weight hyper-parameters by putting less weights on the majority class instances.

Aggregate realtime twitter messages for temporal location information.

Aggregate to Google Earth Engine to identify and outline public satellite image to illustrate disaster affected area.