

A collage of five images representing different types of natural disasters. On the left, a large vertical image shows a powerful tornado touching down. To its right is a vertical image of a bright, fiery volcanic eruption. Further right is a horizontal image of a hurricane's eye from space. Below the hurricane is a horizontal image of a forest fire with trees silhouetted against the flames. On the far right is a vertical image showing a road that has collapsed into a deep ravine, with people standing on the remaining section of the road.

Project Cymopolea: Natural Disaster detection from tweets

Image source: www.alrasub.com

Author: Terence Yep



Introduction

- On average, approximately 60,000 people globally died from natural disasters each year
- Low frequency, high impact events such as earthquakes cause significant but preventable loss of human life
- Why Twitter?
 - real-time microblogging platform
 - 186 million daily active users as of 2020
 - 150 million users worldwide, 36 million from US



Why Bother?

- Total \$210 billion dollars in losses world-wide in 2020
 - Up 26% since 2019
 -
- Preventable loss of human life
 - Last year's natural disasters claimed approximately 8,200 lives
- Fraudulent insurance claims
- Twitter is being used over 50 countries





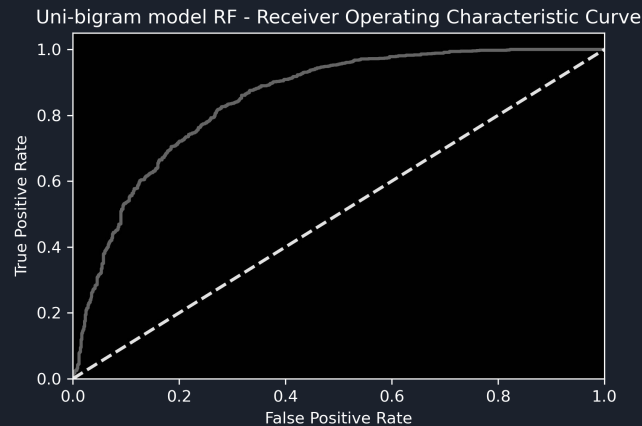
Objectives

- Detect natural disaster tweets
- Detect natural disaster images from attached images

- know world earthquake atomic bomb force man flooding rising
 nuclear disaster watch city injury
 charged monitoring bridge collapse road update latest home
 flood one via
 northern california oil spill family sue high tank
 time damage 4 year new
 life today another crash attack burning building
 death say
 near back building people
 evacuate confirmed mh370 report obama declares
 two storm emergency
 look got army collided bomb home razed dead first
 structural debris found family affected

Results

- Best model based on accuracy
 - RandomForest classifier
 - ~78% accuracy
 - AUC ~0.85
- Worst model
 - Unigram BoW with Gradient-Boost classifier
 - ~65%
 - AUC ~0.85





Limitations

- Class imbalance
- Bias introduced during pre-processing
 - Suicide, war were excluded
 - Locations and keywords were excluded



Future Directions

- Include location and hashtags/keywords in modeling
- Improve accuracy of model (95% target)
- Word embeddings and deep learning (LSTM)
- Mobile app that tracks and sends email to subscriber



Fin