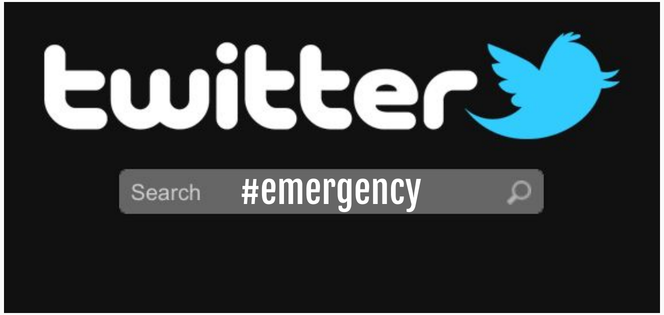
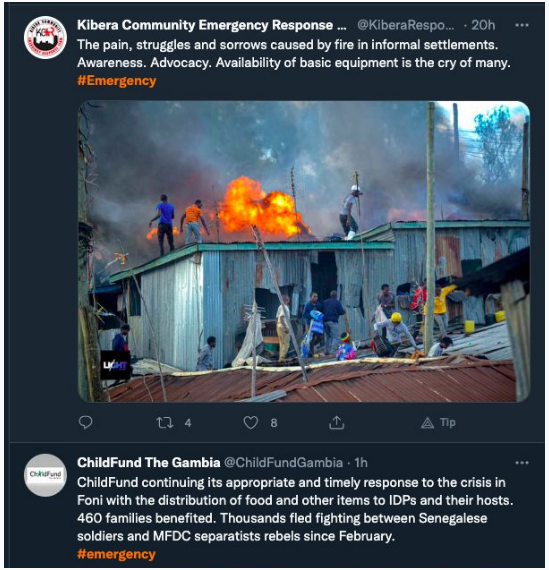
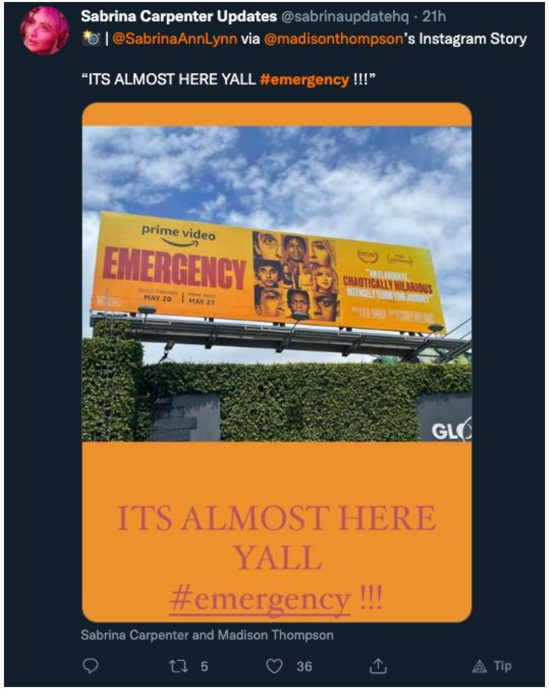
Disaster Tweet Prediction using NLP

By Tyusha Sarawagi & Sarah Wright




Kibera Community Emergency Response ... @KiberaRespo... · 20h
The pain, struggles and sorrows caused by fire in informal settlements. Awareness. Advocacy. Availability of basic equipment is the cry of many.
#Emergency

ChildFund The Gambia @ChildFundGambia · 1h
ChildFund continuing its appropriate and timely response to the crisis in Foni with the distribution of food and other items to IDPs and their hosts. 460 families benefited. Thousands fled fighting between Senegalese soldiers and MFDC separatists rebels since February.
#emergency


Sabrina Carpenter Updates @sabrinaupdatehq · 21h
📸 | @SabrinaAnnLynn via @madisonthompson's Instagram Story

"ITS ALMOST HERE YALL #emergency !!!"

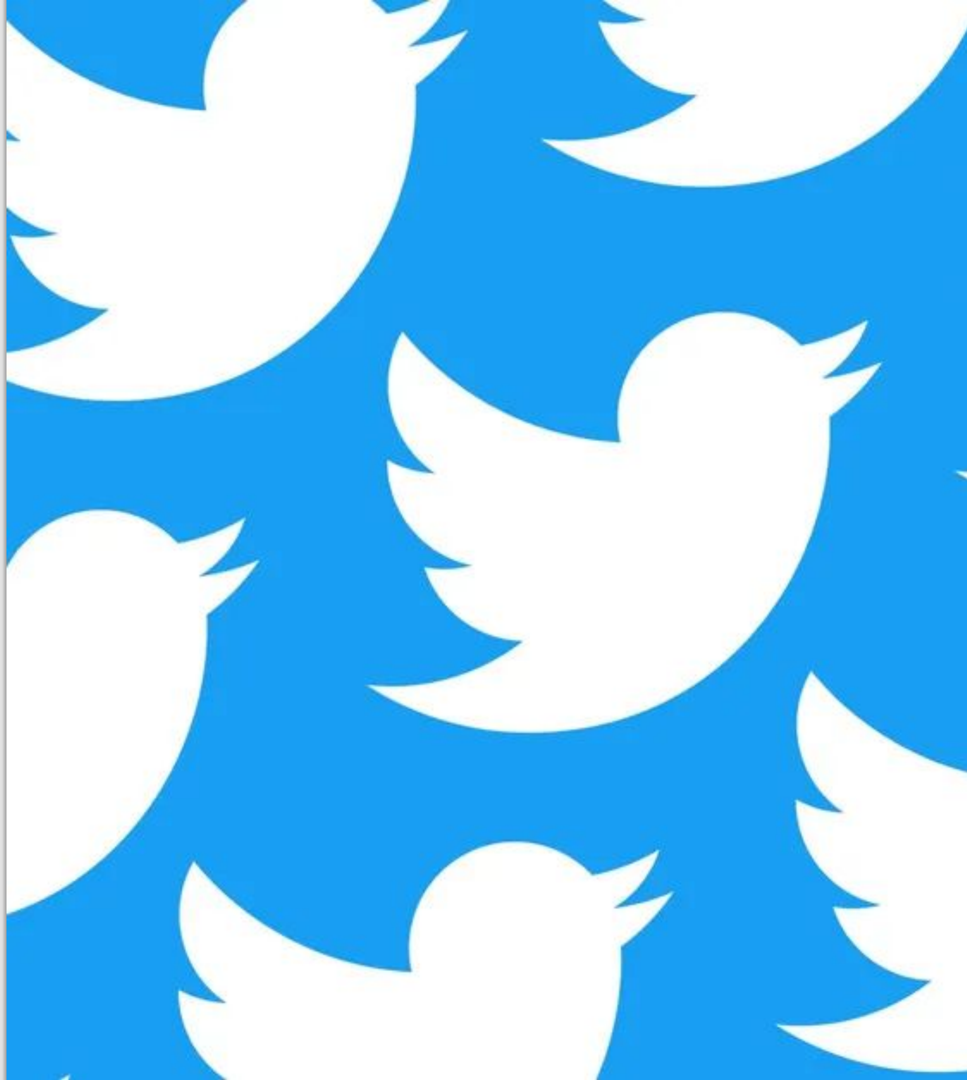
Sabrina Carpenter and Madison Thompson

We have seen in recent years how powerful Twitter can be as both a resource for finding urgent information and as a tool for communicating useless information about sales and petty gossip.

Hashtags can be useful tools for sifting through the nonsense, however, oftentimes they are misused

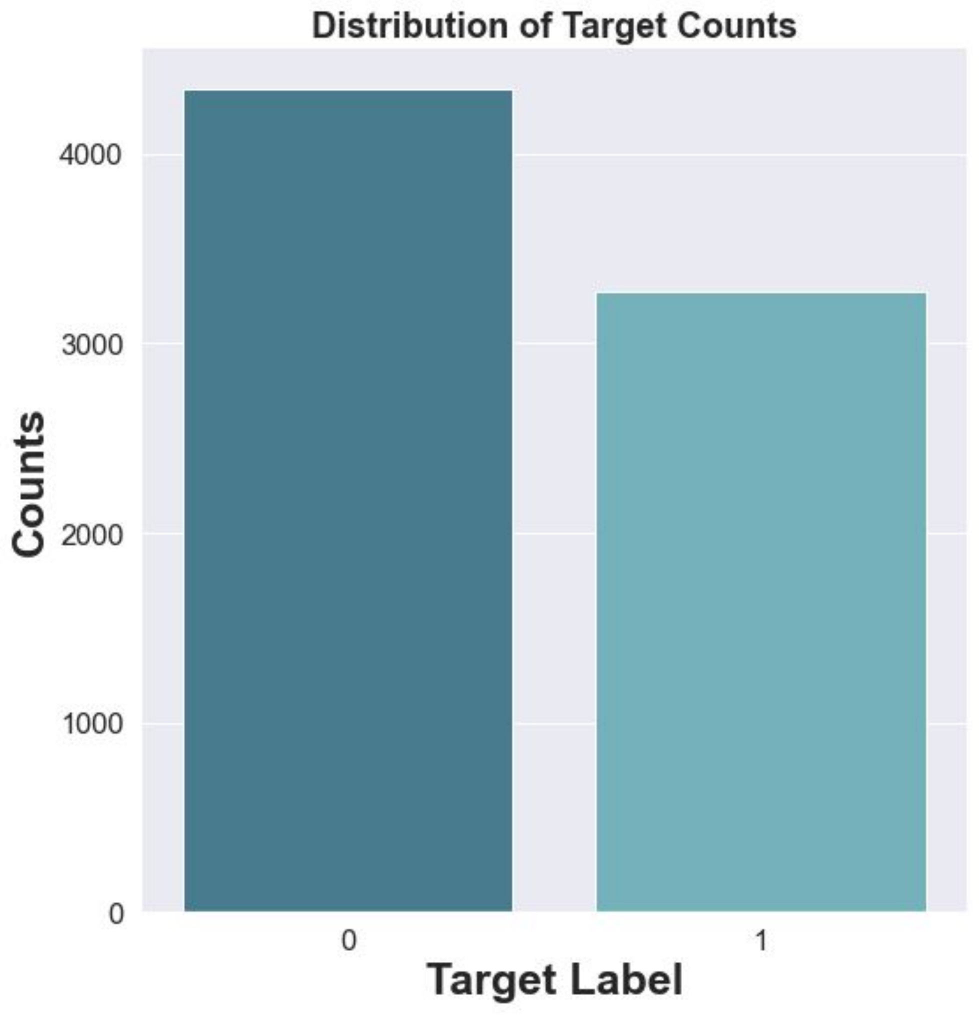
Disaster Tweet Prediction

I. Data
II. EDA
III. Methods
IV. Modeling
V. Results
VI. Conclusion & Purpose



What do we know?

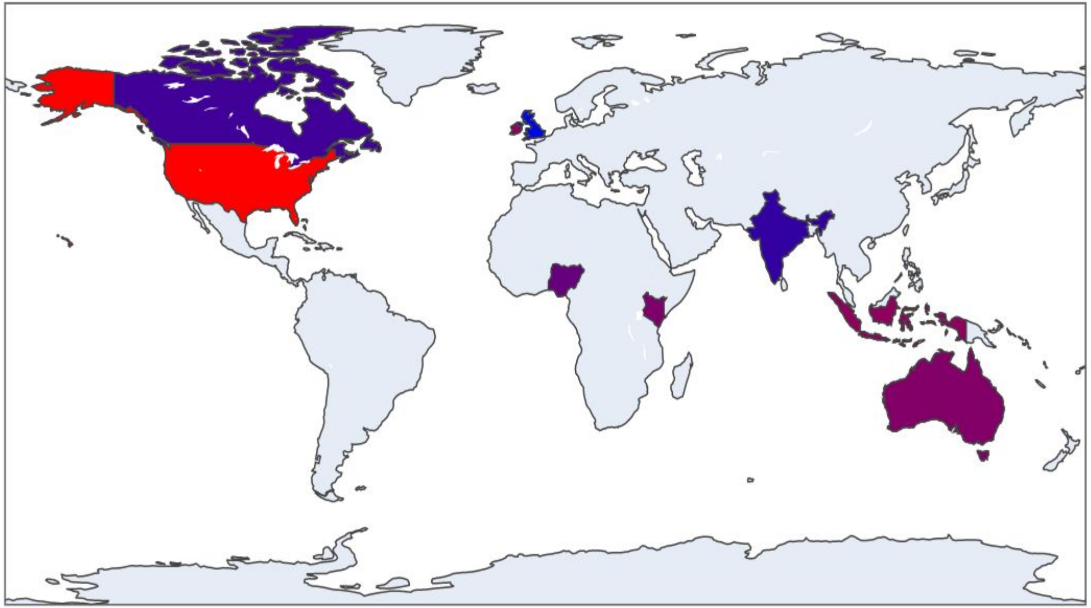
- 7613 tweets, each with its twitter-id, location, a keyword chosen from the tweet, and a target indicating whether the tweet is a disaster tweet or not in the training set.

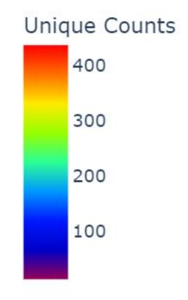


What do we know?

- There are, 4521 unique locations in the dataset
- Top ten countries from the dataset

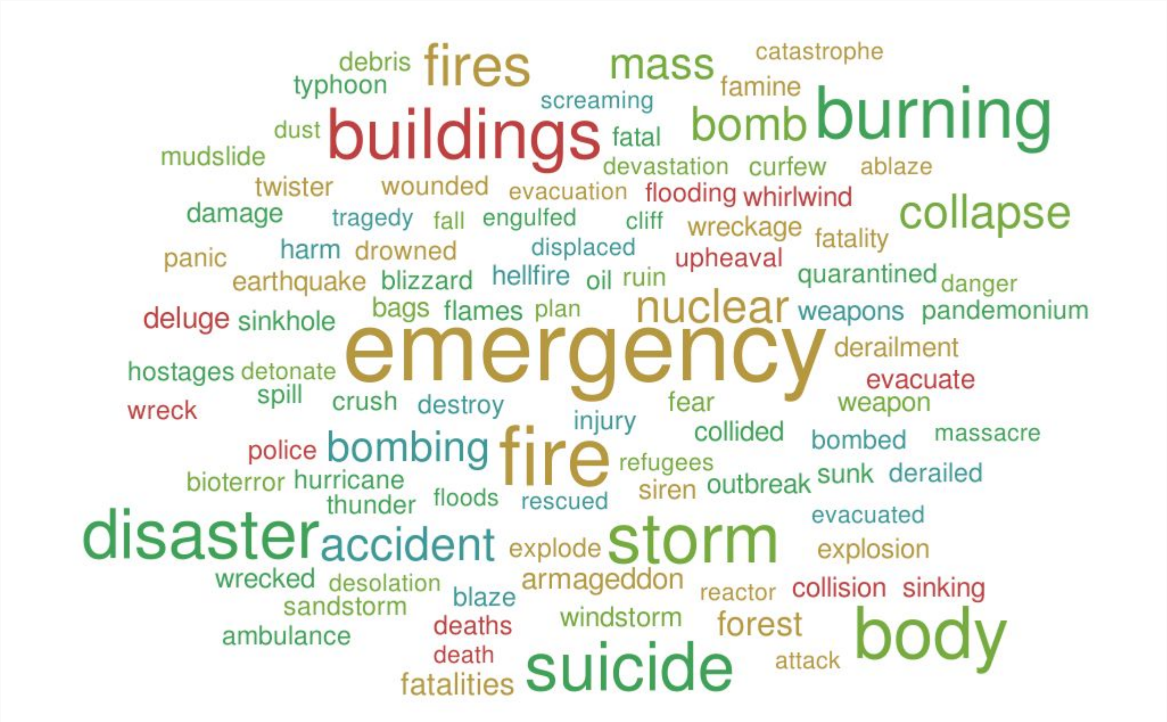
Tweet Locations of Top 10 Countries

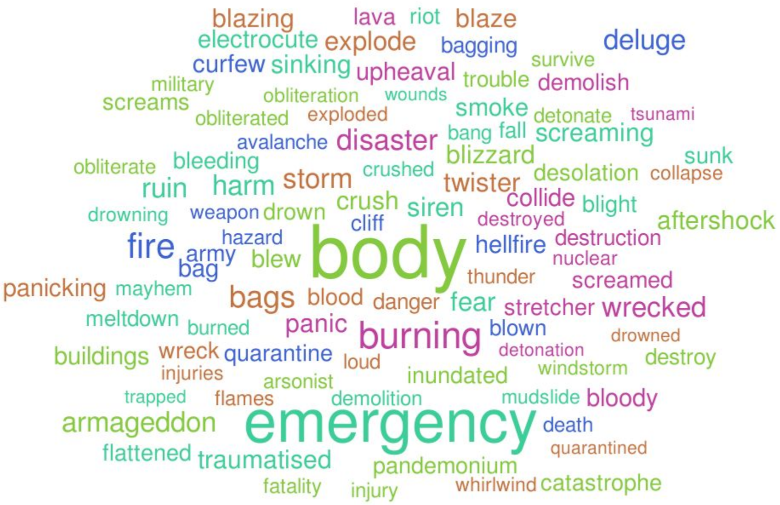


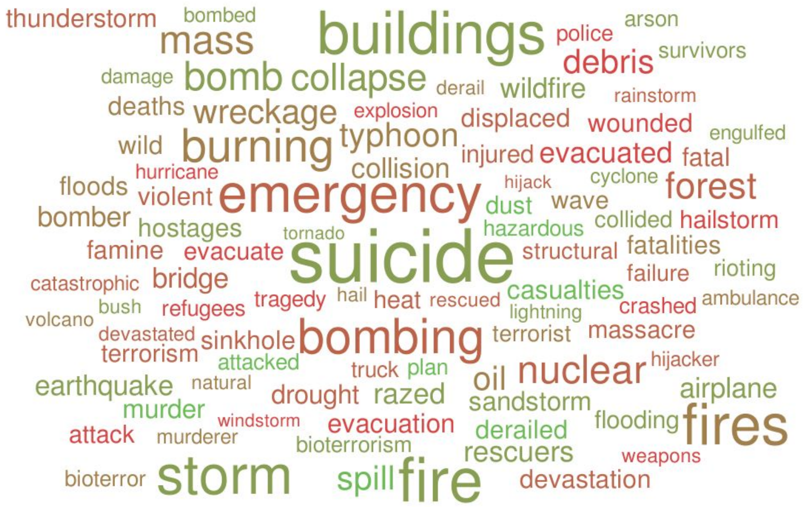


What do we know?

❖ The words “emergency”, “disaster”, “fire”, “storm”, “suicide”, “burning” occurred more frequently in the text data as compared to others.







What do we know?

- ❖ Non-disastrous tweets (only).
  - ➢ Most Frequent Words: “body” and “emergency”.

- ❖ Disastrous tweets (only)
  - ➢ Most Frequent Words: “suicide”, “emergency”. “Bombing”, “storm”, “fire”
  - ➢ There is some overlap, such as with the word “emergency”.

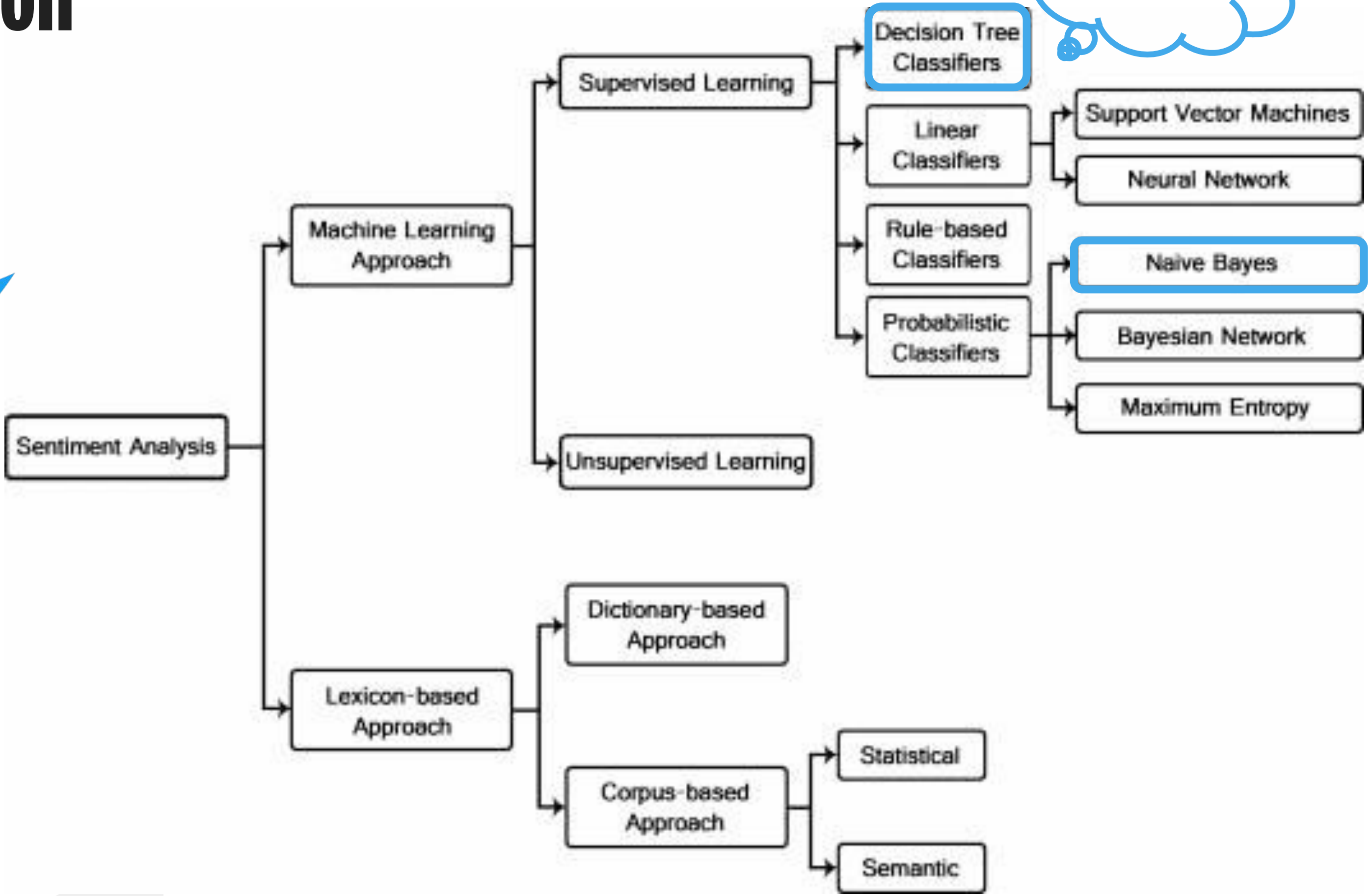
Which messages on Twitter can be classified as Natural Disasters, and how can we do that?

Natural Language Processing (NLP)

Natural language processing (NLP) refers to the branch of computer science concerned with giving computers the ability to understand text and spoken words in much the same way human beings can.

- Pre-Processing
- Text Speech Processing
  - Tokenization
  - Stop Word Removal
  - Lemmatization
- Morphological Analysis
  - Bag-of-Words/ Count Vectorization
  - Term Frequency Inverse Document Frequency (TFIDF)
  - Part-of-Speech Tagging

Model Selection



Sentiment Analysis

Machine Learning Approach

Supervised Learning

Decision Tree Classifiers

Linear Classifiers

Support Vector Machines

Neural Network

Rule-based Classifiers

Probabilistic Classifiers

Naive Bayes

Bayesian Network

Maximum Entropy

Unsupervised Learning

Lexicon-based Approach

Dictionary-based Approach

Corpus-based Approach

Statistical

Semantic

*Random Forest:
Which combines
several Decision Trees

Model Selection

1. Random Forest: 78% accuracy
   a. Random forest is a supervised machine learning algorithm that is commonly used to solve classification and regression issues
   b. Random Forest Algorithm is that it can handle data sets with both continuous and categorical variables

2. Multinomial Naive Bayes: 80% accuracy !?!
   a. Multinomial Naive Bayes is commonly used in multinomial event models such as bag-of-words, which is a way of representing a document as vector space by counting words.

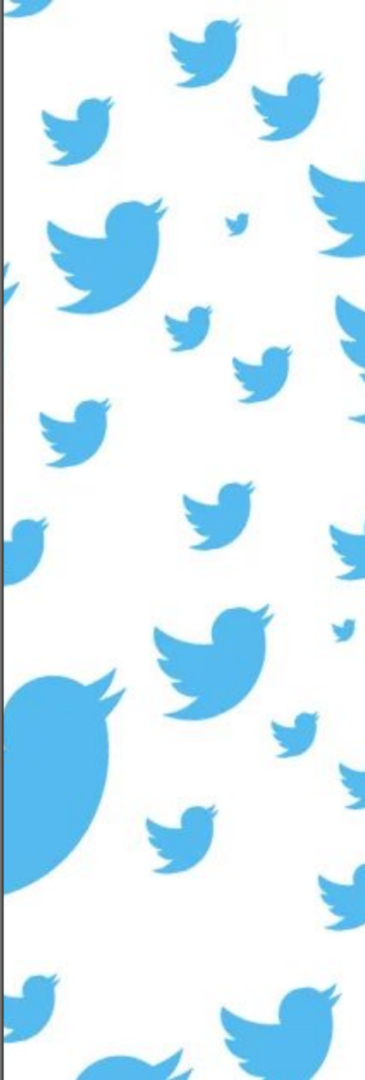
80.6%

ACCURACY

Which messages on Twitter can be **classified** as Natural Disasters, and how can we **use** that?

How can we use that?

- Twitter is a crucial channel of communication.
- Crisis information shared on social media has the potential to save thousands of lives by informing others and allowing them to take preventative action.
- Many agencies are attempting to examine tweets programmatically to detect disasters and emergencies.
- This type of effort can benefit millions of people who have access to the internet and can be notified in the event of an emergency or tragedy.



How can we use that?

- News organizations and disaster relief organizations are attempting to monitor tweets in real-time in order to detect calamities.

- Millions of ordinary people would benefit from this, as it would help them avoid potential disasters.

- People could take preventative action if they were notified in real-time about the occurrence of disasters in a specific place.

- Before the situation becomes out of control, government agencies can carry out an evacuation too.

Questions & Comments