Disaster Help Management Platform

Tweets Classification:

- With the use of **NLP/ML** we are building a platform where people can visit to get the stats as well as reports regarding a disaster.
- The displayed data/stats will be that of tweets that users have posted online regarding the calamity.
- Also we will be collecting photographs of people, which combined with facial recognition can be used to know their status.

Dataset:

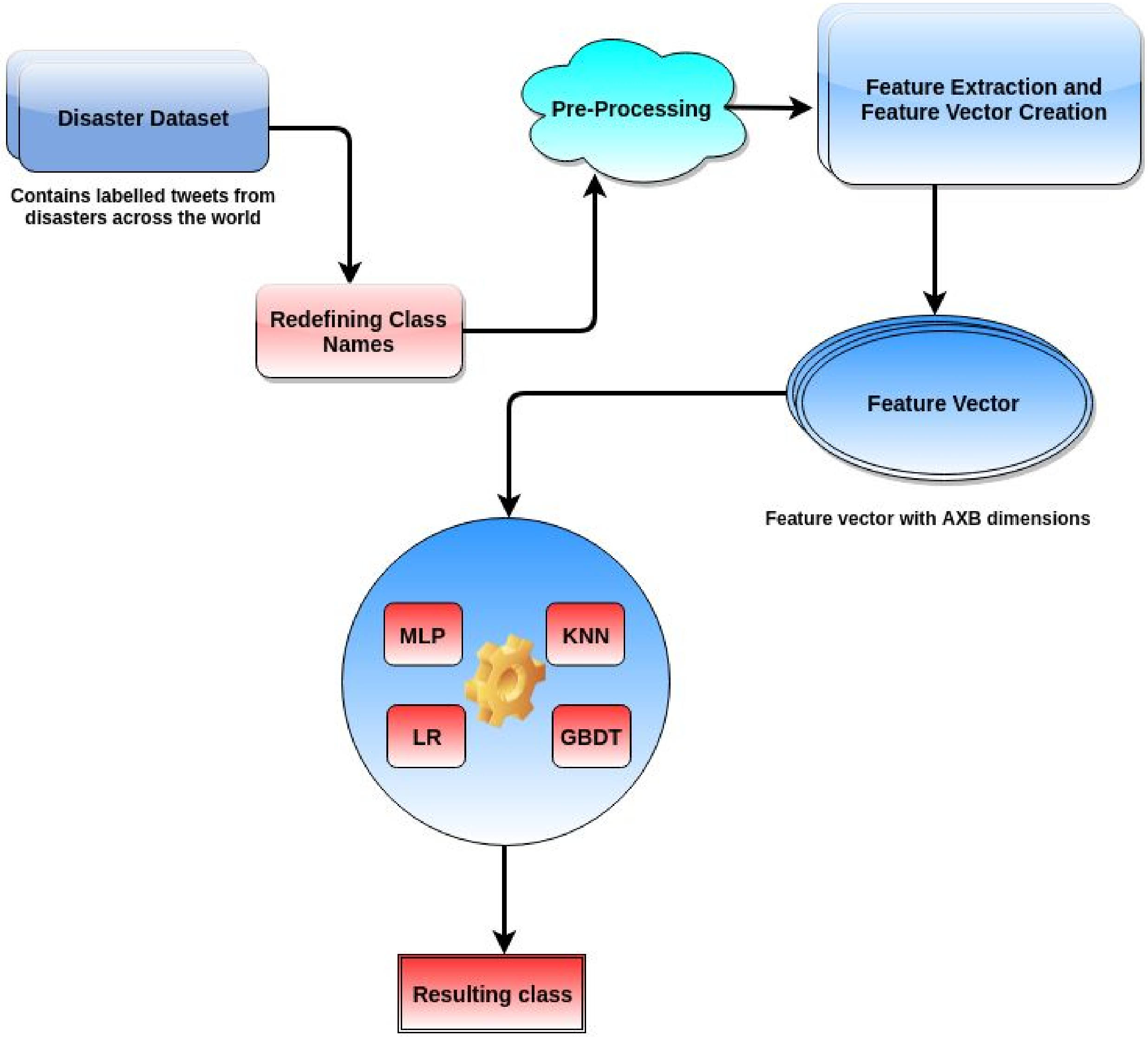
1. We gathered the previous disaster related twitter data from **here.**
2. This data is used to train our model for the upcoming new tweets to be classified into proper classes.
3. We converted our **15 class** dataset into **8 classes** based on type of information they consisted.
4. We had a total of **25,159** labelled tweets.

Pre-Processing:

1. Removed **Mentions (‘@’), Stopwords, Punctuations** and **URLs** from the tweet.
2. We have used an Out-Of-Vocabulary dictionary to replace the slangs, abbreviation, common misspellings, etc.
3. Extracting words from hashtags using Camel-Case splitting.
4. Training Word2Vec on the remaining words, and then using it to represent a sentence by averaging all the words vectors of that sentence.

Experiments:

1. We experimented with **Gradient Boosting, K Nearest Neighbours (KNN), Logistic Regression** and **Multi layer Perceptron (MLP)** individually.
2. Ensemble of these 4 methods provided the best results with an accuracy of **70%.**


Disaster Dataset
Contains labelled tweets from disasters across the world
Redefining Class Names
Pre-Processing
Feature Extraction and Feature Vector Creation
Feature Vector
Feature vector with AXB dimensions
MLP
KNN
LR
GBDT
Resulting class

Training Model Architecture

System Flow:

- Our pre-trained model will be there to predict the classes of the incoming tweets from any happened disaster.
- The tweets with their predicted class will be displayed in our application along with an option for a user to vote for which class they think it should belong to.
- Once a sufficient number of tweets are fetched and voted by various users, we train our classifiers again and hence this process keeps running and improving performance.
- **This display of tweets and classes will show which kind of help is in high demand for that disaster and can help agencies know what to provide.**

Face Recognition:

- **This will be very helpful in case of disasters for finding lost people.**
- People can upload their photos or their loved-ones photos, our model will learn their face representation.
- So, afterwards if someone wants to know their location/condition, they can search by submitting some other/same picture.
- State-of-the-art face recognition models are used for this task.

Dataset :

- The dataset for testing the performance of our models is taken from here.
- It contained photographs of 153 subjects, among them 113 males, 20 females, 20 malestaff.

System Flow :

- When an image is uploaded, it is first compared with all the images in our database -
    - If a match is found, there condition as well saved details are returned (sorry could not be implemented).
    - If a match is not found, then the uploader is notified and the image is then used to learn the encoding for that face.
- For face 100 dimensional encodings are learned.

Help Volunteer :

- **The tweets classified as help category will be list here and people/organisations can take it up.**

Technologies :

We have used Django Web Development framework to implement our system. The following feature are implemented -

1. Tweet classification algorithm is continuously predicting tweets fetched from tweepy api. Our system also allows users to classify the tweets, giving us best of both worlds, if the algorithm does not classify the tweet correctly. We use the information by the users to further train our models for better performance.
2. Python libraries such as sklearn, numpy, nltk, gensim are used to implement tweet classification system.
3. We have used facial_recognition python library, which uses state-of-the-art dlib face recognition and deep learning.
4. Github is used for version control.
5. Microsoft Azure is used for deployment.