**IMPLEMENTATION:**

**MODULES:**

* **User**
* **Admin**
* **Data Preprocess**
* **Machine Learning**

**MODULES DESCRIPTION:**

**User:**

The User can register the first. While registering he required a valid user email and mobile for further communications. Once the user register then admin can activate the customer. Once admin activated the customer then user can login into our system. User can search tweets based on hashtag. The first 100 tweets will get from twitter database and displayed to the user. At this time we are using geo code to identify the user location and tweet location. Most of the time user will not provide coordinates of his identity in the twitter account. So we are taking that as label class. This all tweets and geo code will stored in the database. Later we can apply the machine learning algorithms to test prediction result. The y\_pred and y\_test will displayed on the console. By help of sklearn.model\_selection we can split the data into trainandtest. here we taken 80% of data for training and remiaing 20% for the testing.

.

**Admin:**

Admin can login with his credentials. Once he login he can activate the users. The activated user only login in our applications. The admin can set the training and testing data for the project dynamically to the code. After user operated the algorithms on provded dataset. The admin can view the results of naivebayes, svm and Decision tree results on his screens.

**Data Preprocess:**

Extra characters are removed from tweet text. Capitalize all words to find for geo location. Here we are using geography python library to get the exact latitude and longitude points of the users. Remove the tweet if user home location not mentioned. Mention home location in tweet location, if user tweet location is null

Removes tweets if no location is mentioned in tweet text. Final extract geodata from tweet text. Last step is to assign float value to the locations by its latitude and longitude values.

**Machine Learning:**

**Naive Bayes Classification**

Naive Bayes classifier is the most popular and simple classifier model used commonly. This model finds the posterior probability based on word distribution in the document. Naïve Bayes classifier work with Bag Of Words (BOW) feature extraction model, which do not consider the position of word inside the document. This model used Bayes Theorem for prediction of particular label from the given feature set. The dataset is split into trainset and test set. Upon test set, NB\_model is applied to find the location prediction.

**Support Vector Machine**

Support vector machine is one of most common used supervised learning techniques, which is commonly used for both classification and regression problems. The algorithm works in such a way that each data is plotted as point in ndimensional space with the feature values represents the values of each co-ordinate.

**Decision Tree**

Decision tree is the learning model, which utilizes classifications problem. Decision tree module works by splitting the dataset into minimum of two sets. Decision tree’s internal nodes indicates a test on the features, branch depicts the result and leafs are decisions made after succeeding process on training.