# A STATE OF MIND ANALYSER


**Enrollment No. (s):** 9917103077, 9917103086, 9917103099

**Name of Student (s):** Shubhangi Mishra, Vanshika Bhatnagar, Aditya Asija

**Name of Supervisor:** Dr Pulkit Mehndiratta


**DECEMBER 2020**

**Submitted in partial fulfilment of the Degree of Bachelor of Technology**

**In**

**Computer Science Engineering**


**DEPARTMENT OF COMPUTER SCIENCE ENGINEERING & INFORMATION TECHNOLOGY2**

**JAYPEE INSTITUTE OF INFORMATION TECHNOLOGY, NOIDA**

(I)

## TABLE OF CONTENTS

(II)

## DECLARATION

We hereby declare that this submission is our own work and that, to the best of our knowledge and belief, it contains no material previously published or written by another person nor material which has been accepted for the award of any other degree or diploma of the university or other institute of higher learning, except where due acknowledgment has been made in the text.

**Place:** Jaypee Institute of Information Technology, Noida(UP)

**Date:** December 6, 2020

**Enrollment No. (s):** 9917103077, 9917103086, 9917103099

**Name of Student (s):** Shubhangi Mishra, Vanshika Bhatnagar, Aditya Asija

**(III)**

## CERTIFICATE

This is to certify that the work titled "**A State of Mind Analyser**" submitted by "**Shubhangi Mishra (9917103077), Vanshika Bhatnagar (9917103086), Aditya Asija (9917103099)**" in partial fulfilment for the award of degree of B. Tech of Jaypee Institute of Information Technology, Noida has been carried out under my supervision. This work has not been submitted partially or wholly to any other University or Institute for the award of this or any other degree or diploma.

**Name of Supervisor:** Dr Pulkit Mehndiratta

**Date:** December 6, 2020

(IV)

## ACKNOWLEDGEMENT

We would like to place on record my deep sense of gratitude to Mr. Pulkit Mehndiratta, Dept. of CSE/IT, Jaypee Institute of Information Technology for his generous guidance, help and useful suggestions.

We express our sincere gratitude to him, for his stimulating guidance, continuous encouragement and supervision throughout the course of present work.

We also wish to extend my thanks to my fellow team members and other classmates for their insightful comments and constructive suggestions to improve the quality of this project work.

Shubhangi Mishra (9917103077)

Vanshika Bhatnagar (9917103086)

Aditya Asija (9917103099)

Dr Pulkit Mehndiratta

(Dept. of CSE/IT)

## ABSTRACT

The outbreak of COVID-19 caused heavy disruption to the everyday lives of people across the globe. In a country with a large, diverse population like India, there are bound to be instances of mass hysteria and panic which are further fuelled by unreliable and sometimes inaccurate data.

Gauging the feelings/emotions of the citizens would provide insights into the public mind-set and would pave the way for the government and many organizations to address these situations by providing them with the right data and information, helping in suppressing unnecessary panic among the people. Social media acts as the bridge between the people, the government, and such organizations.

The scope of this project lies in the application of sentiment analysis to the views expressed by people on social media, twitter, in this case, to analyse the trends in the dynamic mood of the population.

## LIST OF FIGURES

## LIST OF TABLES

# LIST OF SYMBOLS & ACRONYMS

| Sr. No. | Name | Symbol/Acronym |
|---|---|---|
| 1 | Logistic Regression | LogR |
| 2 | SGD Classifier | SGD |
| 3 | XG Boost | XGB |
| 4 | Multinomial Naïve Bayes | MNB |
| 5 | Random Forest | RF |

# 1. Introduction:

## 1.1 General Introduction

In this era of flourishing technology, Social Media has become a powerful platform for the public to voice their concerns and beliefs. Among them one such platform is Twitter. Twitter has been a popular platform for micro-blogging in the past few years. In this context, Sentiment Analysis is extremely useful in social media monitoring as it allows us to gain an overview of the wider public opinion behind certain topics. Across the past few years, as the organizations and governments across the world start to adopt the ability to extract insights from social data, the applications of sentiment analysis are broad and powerful. There has been a clear implication that shifts in sentiment on social media correlate with shifts in the economics of a country and also the common notion among the public.

Due to the recent COVID-19 pandemic, there has been a wide change in sentiments of various sectors of the Indian public towards the government policies/actions. Studying the sentiment of the people on the epidemic and government decisions is very important as it acts as a sanity check for the effectiveness of the adopted government policies.

## 1.2 Problem Statement

The effect of COVID-19 pandemic is visible all over the world. National healthcare systems are facing the contagion with incredible strength, but concern regarding psychosocial and economic effects is critically growing. In a fast-moving crisis, as information swarms in from every direction, citizens look to their governments for information, guidance, and leadership. Sentimental Analysis is only the option in this current situation to understand the psychological condition/mental condition of the public. By Sentimental Analysis, the public opinion on COVID-19, regime policies, and actions can be understood. After Analysis, amendments can be made to the decisions taken by the regime policies, and the public can be fortified in such a way so as to enhance the sentiment towards a positive outlook. Not only this but also sentiment analysis will help NGOs and various organizations to come forward to help the people. Businesses can adapt their products and services to match the requirements of the people based on the real-time trending mood of the public, which will not only help businesses to grow but will also help the public meet their need of the hour. Also, this will enable the government to make business and people-friendly rules and laws to help in the betterment of the economy and the market in these untested times.

### 1.3 Significance of the problem

Due to the outbreak of COVID-19, in a country with a large, diverse population like India, there are bound to be instances of **mass hysteria** and **panic** which are further fuelled by **unreliable and sometimes inaccurate data.** Social media acts as the bridge between the people, the government, and such organizations.

•Determining the emotions of the citizens by the government would :

1. Provide **insights into the public mind-set**
2. Pave the way for the government and many organizations to **address these situations**
3. Provide **right data and information**
4. Help in **suppressing unnecessary panic** among the people

•**Acts as a sanity check for the effectiveness of the adopted government policies.**

•After Analysis, **amendments can be made to the decisions taken by the regime policies**, and can be made in such a way so as to enhance the sentiment towards a positive outlook.

•**Help NGOs and various organizations** to come forward to help the people.

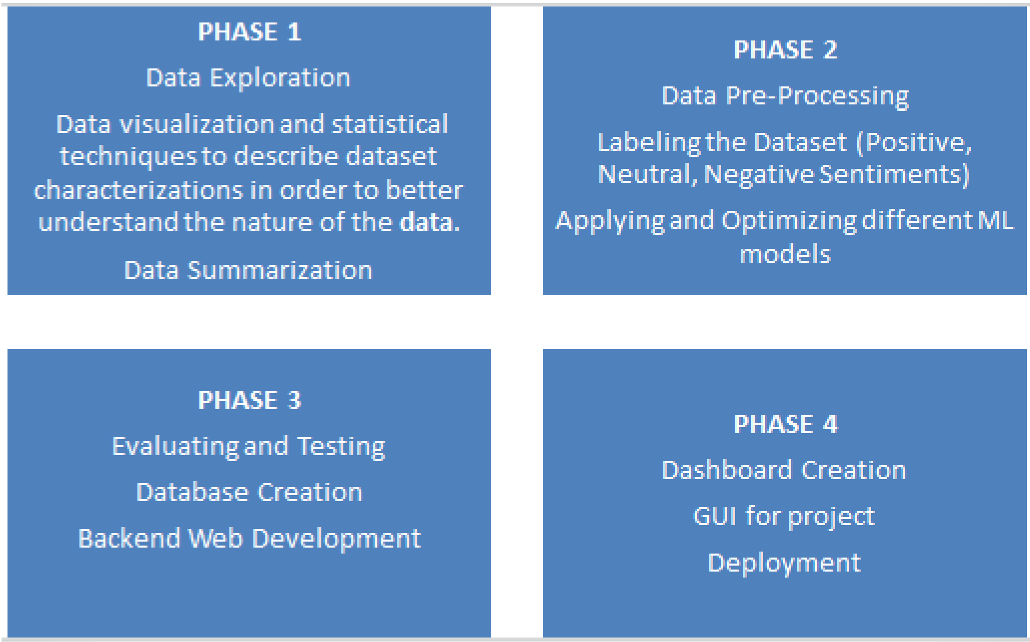
•**Businesses can adapt their products and services** to match the requirements of the people based on the trending mood of the public, which will not only **help businesses to grow but will also help the public meet their need of the hour.**

•**Shifts in sentiment on social media correlate with shifts in the economics of a country** •Govt. can thereby make **business and people-friendly rules and laws to help in the betterment of the economy** and the market in these untested times.

### 1.4 Scope of the project

"A State of Mind Analyser" is a **web application**, in which when inputted "**Geographical Location**" and "**Time Period**" from a given available set, displays the **"overall" sentiment expressed by people** on Twitter, a popular microblogging platform. It can be **used to analyse the trends in the dynamic mood of the population.**

### 1.5 Solution Approach



## 2. Literature Survey:

Kiritchenko et al [1] conducted a large-scale computational analysis to explore how the terms associated with the state of being alone are used in online language. They presented SOLO (State of Being Alone), a corpus of over 4 million tweets collected with query terms solitude, lonely, and loneliness, which formed the basis of our project.

Jianqian et al [2] discussed the effects of text pre-processing method on sentiment classification performance in two types of classification tasks, and summed up the classification performances of six pre-processing methods using two feature models and four classifiers on five Twitter datasets.

Tyagi et al [3] extracted features using N-gram modeling technique. The sentiments then are categorized among positive, negative and neutral using a supervised machine learning algorithm known as K-Nearest Neighbor.

Kharde et al [4] provide a survey and a comparative analyses of existing techniques for opinion mining like machine learning and lexicon-based approaches, together with evaluation metrics.

Kaushik et al [5] has elaborated on Opinion Mining, a subset of Data Mining, and its sub divisions which include Supervised Opinion Mining, Case Based Reasoning and Unsupervised Opinion Mining.

Pandarachalil et al [6] presents an unsupervised method for analyzing tweet sentiments. Polarity of tweets is evaluated by using three sentiment lexicons—SenticNet, SentiWordNet and SentislangNet. SentislangNet is a sentiment lexicon built from SenticNet and SentiWordNet for slangs and acronyms. Experimental results show fairly good F-score.

Zhang et al [7] adopts a lexicon based approach to perform entity-level sentiment analysis which gives high precision, but low recall. To improve recall, additional tweets that are opinionated are identified automatically by exploiting the information in the result of the lexicon-based method. A classifier is then trained to assign polarities to the entities in the newly identified tweets. Instead of being labeled manually, the training examples are given by the lexicon-based approach. This improves the recall and the F-score, and outperforms the state-of-the-art baselines.

Kouloumpis et al [8] investigates the utility of linguistic features for detecting the sentiment of Twitter messages, evaluates the usefulness of existing lexical resources as well as features that capture information about the informal and creative language used in microblogging. They take a supervised approach to the problem, but leverage existing hashtags in the Twitter data for building training data.

Desai et al [9] presents the sentiment analysis process to classify highly unstructured data on Twitter. They discuss various techniques to carryout sentiment analysis on Twitter data in detail. Moreover, they present the parametric comparison of the discussed techniques based on our identified parameters.

Alsaeedi et al [10] study the existing sentiment analysis methods of Twitter data and provide theoretical comparisons of the state-of-art approaches.

Pak et al [11] perform linguistic analysis of the collected corpus and explain discovered phenomena. Using the corpus, they build a sentiment classifier that is able to determine positive, negative and neutral sentiments for a document.

Saif et al [12] introduce a novel approach of adding semantics as additional features into the training set for sentiment analysis. For each extracted entity from tweets, they add its semantic concept as an additional feature, and measure the correlation of the representative concept with negative/positive sentiment. They find that semantic features produce better Recall and F score when classifying negative sentiment, and better Precision with lower Recall and F score in positive sentiment classification.

Sarlan et al [13] reports on the design of a sentiment analysis, extracting a vast amount of tweets. Prototyping is used in this development. Results classify customers' perspective via tweets into positive and negative, which is represented in a pie chart and html page.

Deshwal et al [14] combine many feature extraction techniques like emoticons, exclamation and question mark symbol, word gazetteer, unigrams to design more accurate sentiment classification system. This paper presents empirical comparison of six supervised classification algorithms.

Madhuri [15] proposed a framework for discovering sentiments from tweets of Indian Railways. This is a domain specific framework which leverages business intelligence through different classifiers such as C4.5, Naive Bayes, SVM and Random Forest. An evaluation procedure with measures like precision, recall, F-Measure and accuracy is provided.

Go et al [16] use different machine learning classifiers (Naive Bayes, Maximum Entropy, and Support Vector Machines) and feature extractors (unigrams, bigrams, unigrams and bigrams, and unigrams with part of speech tags). They build a framework that treats classifiers and feature extractors as two distinct components which allow trying out different combinations of classifiers and feature extractors.

Forman [17] presents an empirical comparison of twelve feature selection methods evaluated on a benchmark of 229 text classification problem instances. The results are analyzed from multiple goal perspectives—accuracy, F-measure, precision, and recall. The results reveal that a new feature selection metric called 'Bi-Normal Separation' (BNS), outperformed the others by a substantial margin.

# 3. Requirement Analysis:

## 3.1 Software Requirements

- Anaconda Platform (Jupyter Notebook)
- Python 3 or higher

## 3.2 Hardware Requirements

- Microsoft Windows 10
- Processor: Intel ® Core (TM) i5 -6200U CPU @2.30GHz 2.40GHz
- Ram : 4 GB and above
- Disk Space : 1 TB

## 3.3 Functional Requirements

- Appropriate data set to work on.
- Eligible software to implement our ideas.
- Suitable libraries for using algorithm in the source code.

## 3.4 Non-functional Requirements

- Validation: Applying benchmark functions for checking the accuracy and performance of algorithms.

## 3.5 Libraries Used:

- Numpy
- Pandas
- matplotlib
- seaborn
- sklearn
- stratifiedKfold
- display
- imblearn
- joblib

- counter
- preprocessing
- pickle
- tweet-preprocessor
- nltk
- tfidfvectorizer
- tfidftransformer
- train_test_split
- confusionmatrix
- RandomForestClassifier
- SGDClassifier
- XGBClassifier
- LogisiticRegression
- StandardScaler
- MultinomialNB
- accuracy_score
- clone

# 4. Detailed Design:

## 4.1 Exploratory Data Analysis (EDA):

- **Missing value determination-** Missing values are explored in the data and imputed, along with they are plotted to get a clear visualization.
- Determining the unique values and plot them
- **Data visualization through bar plots and word cloud-** Word Cloud is a data visualization technique used for representing text data in which the size of each word indicates its frequency or importance. Significant textual data points can be highlighted using a word cloud.
- Plotting tweet count based on location.
- Plotting tweeter sources based on tweet count

## 4.2 Data Pre-processing:

After collecting the data, it needs to be preprocessed and cleaned, thus allowing feature extraction from the input data. Data preprocessing is one of the major phases within the knowledge discovery process. Despite being less known than other steps like data mining, data preprocessing actually very often involves more effort and time within the entire data analysis process. Raw data usually comes with many imperfections such as inconsistencies, missing values, noise and/or redundancies. Performance of subsequent learning algorithms will thus be undermined if they are presented with low-quality data. Thus by conducting proper preprocessing steps we are able to significantly influence the quality and reliability of subsequent automatic discoveries and decisions. There are certain steps involved in preprocessing which are as follows:

a. Noise and Outlier Detection

b. Missing Feature Values

c. Normalization

d. Discretization

e. Instance Selection

f. Tokenization

g. Part of speech tagging (POS tagging)

h. Removing the hash tags, URLs, user tags, retweets, slang, incorrect spellings

i. Removal of stopwords

j. Stemming

k. Lemmatization

l. Feature Selection

### 4.3 Training the classifiers:

1. • Training data cleaning and preprocessing
2. • tf-idf vectorizer: Transform text into a meaningful representation of numbers which is used to fit machine algorithm for prediction.
3. • Balancing the dataset by k-corss validationg(k=5) to avoid overfitting (using SMOTE over-sampling and Random over-sampling)
4. • Splitting the dataset for training (75%) and testing (25%) data
5. • Training the classifiers (LogR, SGD classifier, XG Boost, MNB, Random Forest)
6. • Evaluation of performance of classifiers (Comparing Accuracy, Confusion Matrix, graphs)

# 5. Implementation:

## 5.1 Description about dataset:

Dataset used for the project has been rightfully incorporated from Appendix 1. At present, it contains 1260 entries. On a later deployment stage, all of the data from Appendix 1 which is updated in real-time will be accessed and trained accordingly. This will provide a comprehensive insight on micro-blogging trends and its sentiment analysis in times of COVID-19. [Appendix 1]

## 5.2 Exploratory data analysis:

Exploratory data analysis is an approach of analysing data sets to summarize their main characteristics, often with visual methods. Variable identification, missing value identification along with its treatment, plotting correlation matrix and different bar plots along with WordCloud has been done as a part of exploratory data analysis.

## 5.3 Data Pre-processing:

It is a method used to convert the raw data into a clean dataset. Pre-processing is applied using different techniques such as removing the stop words, stemming and lemmatization. After cleaning the data we have saved it for training purpose.

## 5.4 Training the classifiers:

After cleaning the dataset we have trained it using different Machine Learning classifiers in order to make predictions. In total we have used five classifiers namely, LogR, SGD, XGB, MNB and RF. Along with this we have compared the training and testing accuracy for each of these models. Dataset has been split into 75%-25% ratio for training and testing respectively.

## 6. Experiment Results:

- Missing values are detected and treated with appropriate means.
- Wordcloud of hashtags and tweets have been created for better visualization.
- Bar plots and wordcloud creation for manipulating tweets and other features based on location.

**Fig 6.1Top 20 Locations by Number of Tweets:**

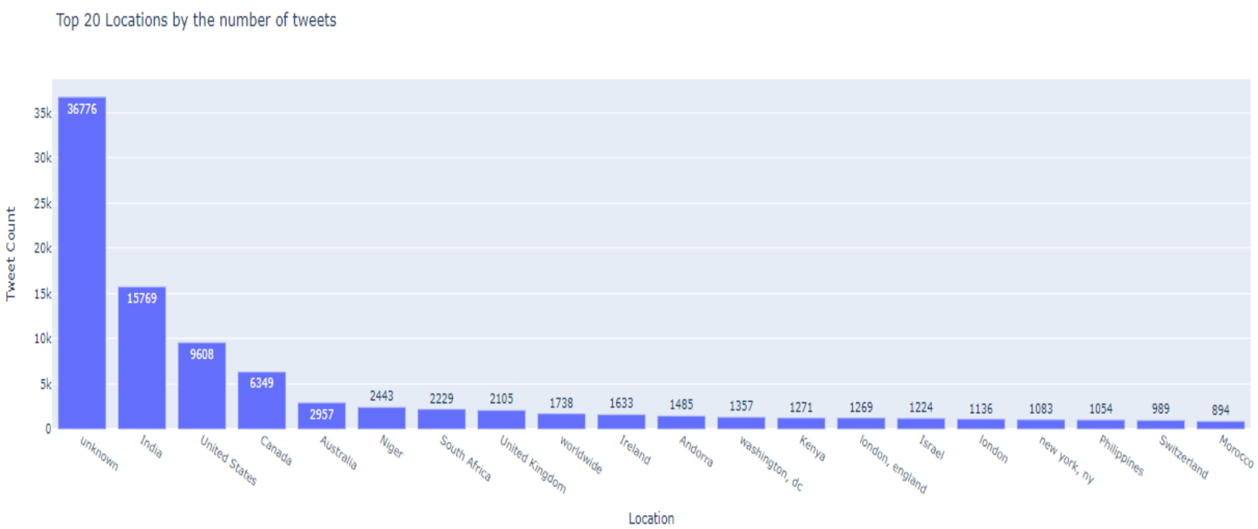



**Fig 6.2 Top 20 Sources by Number of Tweets:**

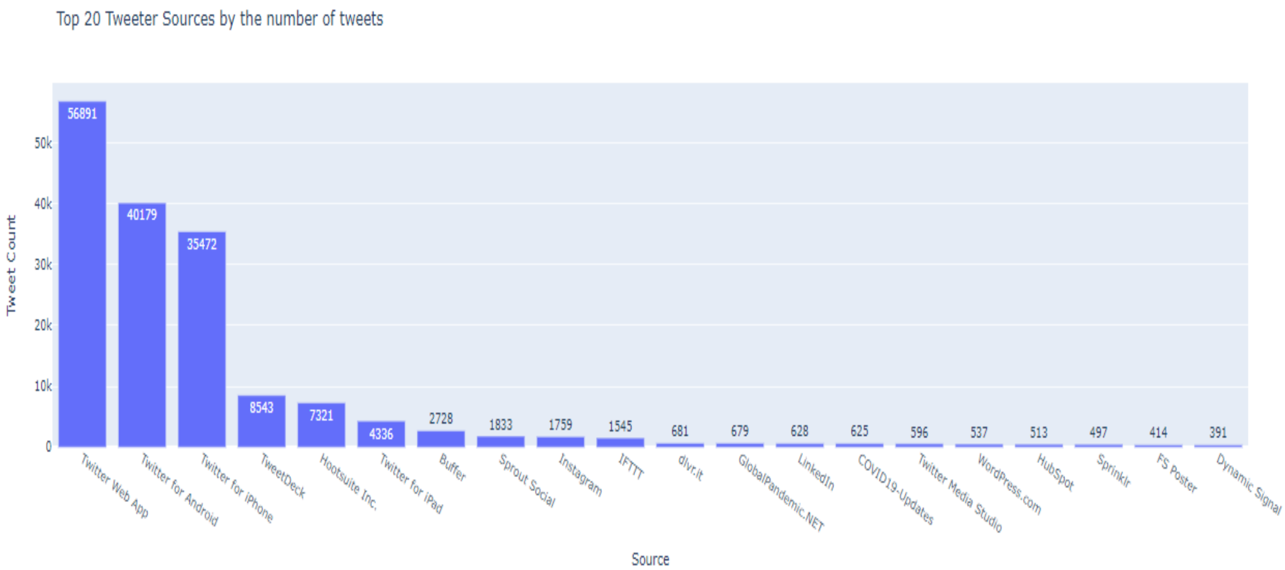
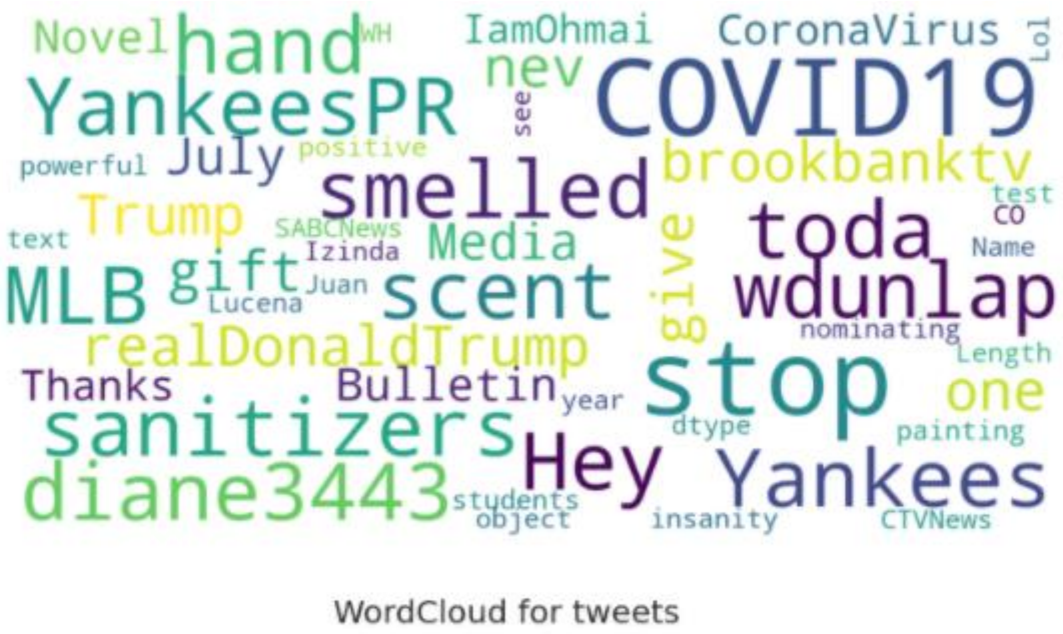


**Fig 6.3 Word Cloud for Prevalent Words:**




| Classifiers | No Sampling | | SMOTE over-sampling | | Random over-sampling | |
|---|---|---|---|---|---|---|
| | Training Accuracy | Testing Accuracy | Training Accuracy | Testing Accuracy | Training Accuracy | Testing Accuracy |
| LogR | 84.71 | 82.72 | 83.36 | 74.4 | 82.02 | 72.67 |
| SGD | 85.37 | 82.84 | 84.11 | 73.41 | 82.56 | 73.5 |
| XG Boost | 90.65 | 84.91 | 90.91 | 80.7 | 90.86 | 81.04 |
| MNB | 81.55 | 80.47 | 76.39 | 71.24 | 75.37 | 69.77 |
| Random Forest | 97.83 | 88.98 | 96.89 | 86.5 | 95.91 | 85.48 |

**Table 6.1 : Accuracy of Classifiers (refer to Appendix 2)**

| Classifiers | Training Accuracy | Testing Accuracy |
|---|---|---|
| LogR | 84.79 | 83.99 |
| SGD | 85.8 | 85.1 |
| XG Boost | 90.55 | 85.19 |
| MNB | 81.96 | 81.42 |
| Random Forest | 97.82 | 90.25 |

**Table 6.2 : Confusion Matrix Accuracy of Classifiers (refer to Appendix 3)**

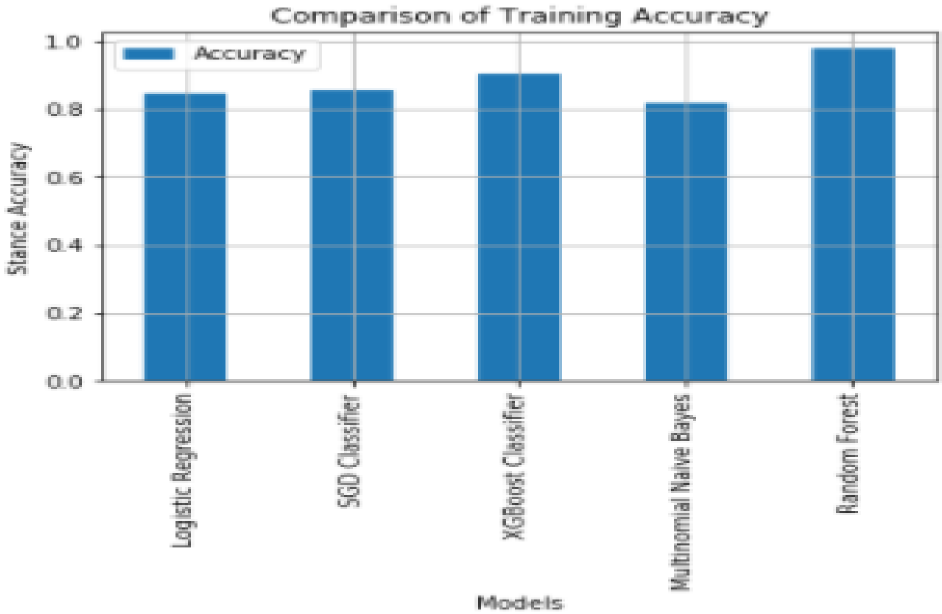



**Fig 6.4 Comparison of Training Accuracy**

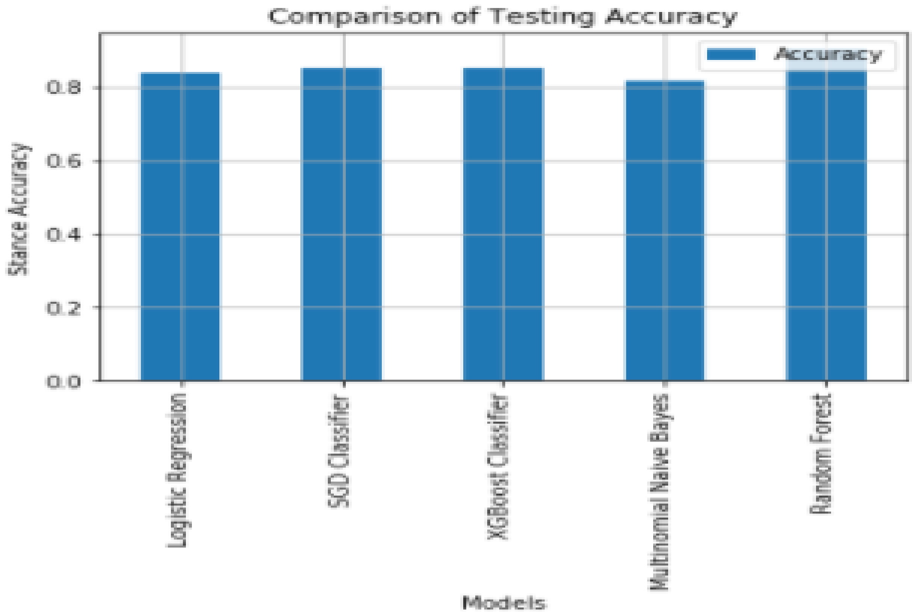



**Fig 6.5 Comparison of Testing Accuracy**

- The highest Training and Testing accuracy has been achieved when the dataset was classified using the Random Forest classifier without sampling with accuracies 97.85 and 90.25 respectively.

## 7. Test Plan

I. Tweets based on particular keyword and geolocation

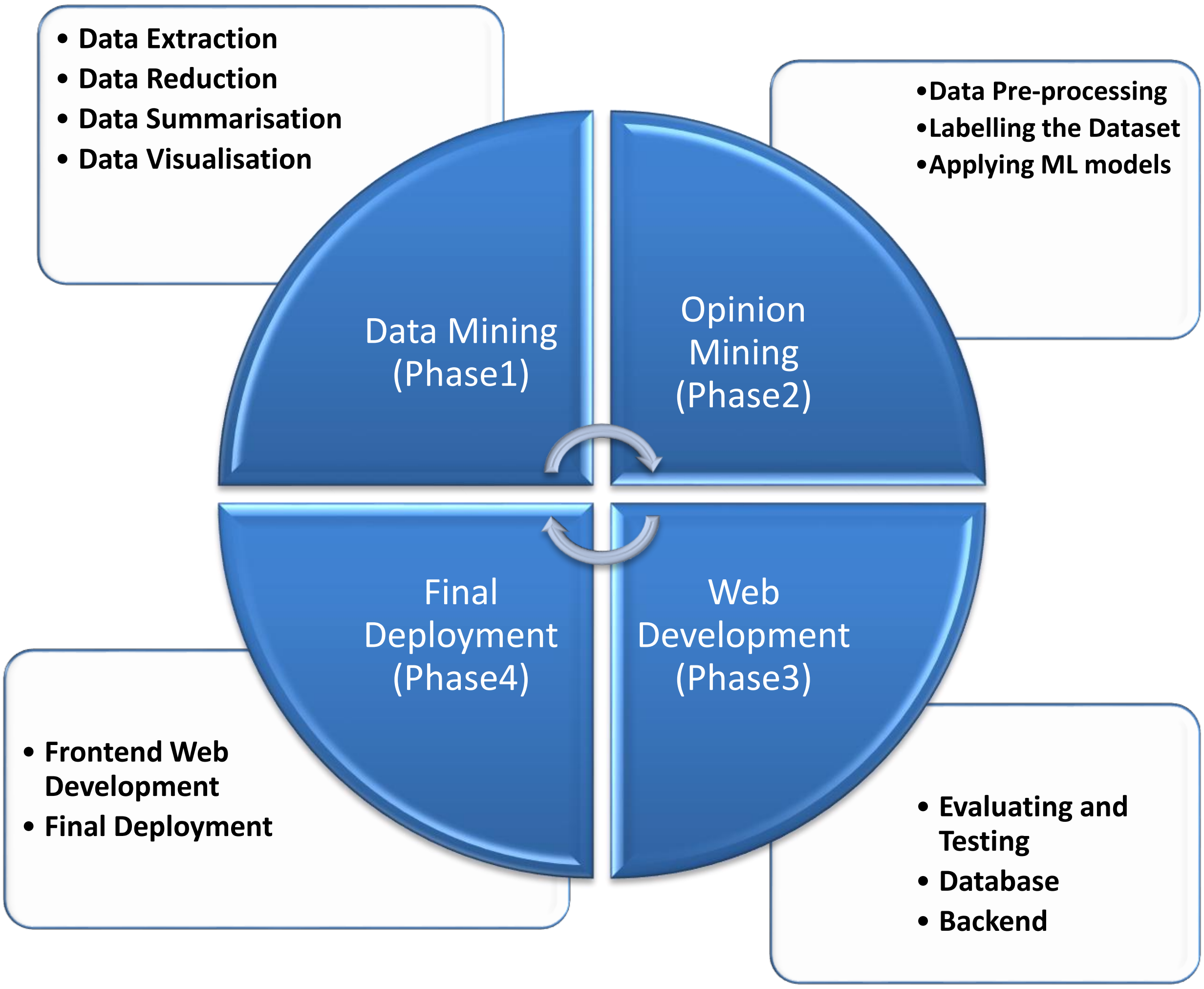
II. Data pre-processing (especially labelling the dataset) and cleaning the text by normalization and removing the user tags.

III. Train the data along with applying and optimizing different machine learning models.

IV. Data Visualization- Different plots for most frequent words, hashtags and wordcloud representation for both positive and negative tweets.

V. Making the Backend, getting live data and connecting everything.

VI. Creation of GUI (Front End)




:

## 8. Risk Analysis and Future Scope:

- There is difficulty to implement sentiment analysis because of the ambiguity of natural language and also the characteristics of the posted content. The analysis of tweets is an example of this, for they are usually coupled with hashtags, emoticons and links, creating difficulties in determining the expressed sentiment.
- There is a need for automatic techniques that require large datasets of annotated posts or lexical databases where emotional words are associated with sentiment values.
- Sentiment analysis faces a problem in recognizing human aspects of a language like irony, sarcasm, negotiations, exaggerations, and jokes - the sorts of things humans wouldn't face many problems in understanding. Machines sometimes fail in recognizing these aspects, which leads to skewed and incorrect results. The project can be further enhanced by training the model about the human aspects of a language and making it more accurate in cases where sarcasm, irony, and other aspects are used.
- In the context of project, for example, the terms "fight" and "positive" are used in a negative and positive context respectively, but we observe a role reversal in this situation. The identification of such terms and their usage according to the context would be an essential part of the project.
- For this particular project, taking the actions of all the ministries into consideration while gauging the sentiments of the public can also make the analysis more detailed and sector-specific, which would help in the analysis of the area of development required in those sectors
- There is still a long way to go before sentiments can be accurately detected from texts, because of the complexity involved in the English language, and even more when other languages like Hindi are considered.
- The classifier can be further improved by trying to extract more features from the tweets, trying different kinds of features, tuning the Hyperparameters (In ML, parameters used to control the learning process), and also by making it work on various Indian languages.
- In terms of architecture, presence of unclear or scarce datasets and lack of labelled data can pose a barrier to the advancements in the area of sentiment analysis. The project, in its initial stage of development, while data pre-processing has been able to handle this to a certain extent.

## References


[1] A Kaushik , A Kaushik , S Naithani, "A Study on Sentiment Analysis: Methods and Tools," International Journal of Science and Research (IJSR), vol. 4, no. 12, pp. 287–292, Dec. 2015.

[2] Alsaeedi M Khan, "A Study on Sentiment Analysis Techniques of Twitter Data," International Journal of Advanced Computer Science and Applications, vol. 10, no. 2, pp. 361–374, 2019.

[3] Deshwal and S. K. Sharma, "Twitter Sentiment Analysis using Various Classification Algorithms," in 2016 5th International Conference on Reliability, Infocom Technologies and Optimization, 2016, pp. 251–257.

[4] Go, R. Bhayani, and L. Huang, "Twitter Sentiment Classification using Distant Supervision," cs.stanford.edu, 2009. .

[5] Pak and P. Paroubek, "Twitter as a Corpus for Sentiment Analysis and Opinion Mining," in Proceedings of the International Conference on Language Resources and Evaluation, 2010, pp. 1320–1326.

[6] Sarlan, C. Nadam, and S. Basri, "Twitter Sentiment Analysis," in 2014 International Conference on Information Technology and Multimedia, 2014, pp. 212–216.

[7] D. Krishna Madhuri, "A Machine Learning based Framework for Sentiment Classification: Indian Railways Case Study," International Journal of Innovative Technology and Exploring Engineering, vol. 8, no. 4, pp. 441–445, 2019.

[8] E Kouloumpis , T Wilson , J Moore, "Twitter Sentiment Analysis: The Good the Bad and the OMG!," in Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media, 2011, pp. 538–541.

[9] G. Forman, "An Extensive Empirical Study of Feature Selection Metrics for Text Classification," Journal of Machine Learning Research, vol. 3, pp. 1290–1305, 2003.

[10] G. X. Z Jianqiang, "Comparison Research on Text Pre-processing Methods on Twitter Sentiment Analysis," IEEE Access, vol. 5, pp. 2870–2879, Feb. 2017.

[11] H. Saif, Y. He, and H. Alani, "Semantic Sentiment Analysis of Twitter," in International Semantic Web Conference, 2012, pp. 508–524.

[12] L Zhang , R Ghosh , M Dekhil , M Hsu , B Liu, "Combining Lexicon-based and Learning-based Methods for Twitter Sentiment Analysis," HP Laboratories, vol. 89, Jan. 2011.

[13] M. M. M Desai, "Techniques for Sentiment Analysis of Twitter Data: A Comprehensive Survey," in International Conference on Computing, Communication and Automation, 2016, pp. 149–154.

[14] P Tyagi , , Dr RC Tripathi, "A Review towards the Sentiment Analysis Techniques for the Analysis of Twitter Data," in Proceedings of 2nd International Conference on Advanced Computing and Software Engineering (ICACSE) 2019, 2019, p. 5.

[15] R Pandarachalil . S Sendhilkumar , , G. S. Mahalakshmi, "Twitter Sentiment Analysis for Large-Scale Data: An Unsupervised Approach," Cognitive Computation, vol. 7, no. 2, Apr. 2014.

[16] S Kiritchenko , W E. Hipson , R J. Coplan , S M. Mohammad, "SOLO: A Corpus of Tweets for Examining the State of Being Alone," in Proceedings of the 12th Language Resources and Evaluation Conference, 2020, pp. 1567–1577.

[17] Vishal A. Kharde , , S Sonawane, "Sentiment Analysis of Twitter Data: A Survey of Techniques," International Journal of Computer Applications, vol. 139, no. 11, pp. 5–15, Apr. 2016.

[18] M. Ahmad, S. Aftab, S. S. Muhammad, and S. Ahmad, "Machine Learning Techniques for Sentiment Analysis: A Review," INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY SCIENCES AND ENGINEERING, vol. 8, no. 3, Apr. 2017.

[19] M. Ebrahimi, A. H. Yazdavar, and A. Sheth, "Challenges of Sentiment Analysis for Dynamic Events," EEE Intelligent Systems, vol. 32, no. 5, pp. 70–75, Sep. 2017.

[20] M. H. Abd El-Jawad, R. Hodhod, and Y. M. K. Omar, "Sentiment Analysis of Social Media Networks Using Machine Learning," in 14th International Computer Engineering Conference (ICENCO), 2018.

[21] M. M. M Desai, "Techniques for Sentiment Analysis of Twitter Data: A Comprehensive Survey," in International Conference on Computing, Communication and Automation, 2016, pp. 149–154.

[22] M. R. Huq, A. Ali, and A. Rahman, "Sentiment Analysis on Twitter Data using KNN and SVM," International Journal of Advanced Computer Science and Applications, vol. 8, no. 6, pp. 19–25, 2017.

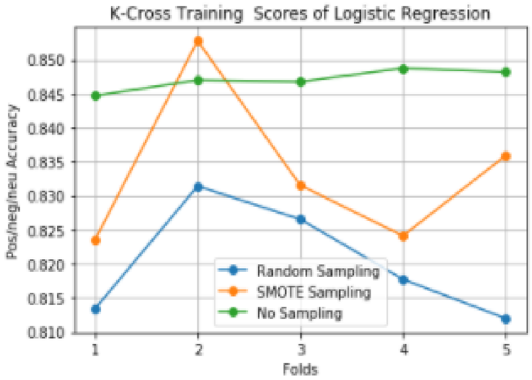
[23] M. Trupthi, S. Pabboju, and G. Narasimha, "SENTIMENT ANALYSIS ON TWITTER USING STREAMING API," in IEEE 7th International Advance Computing Conference, 2017.

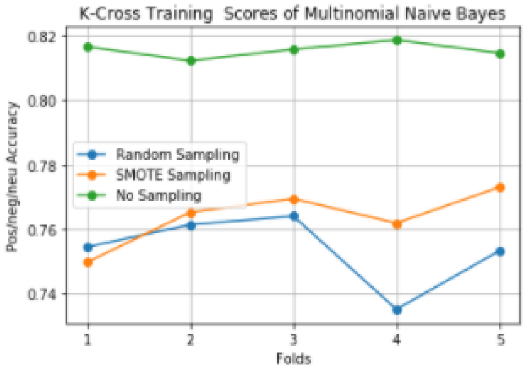
[24] M. Z. Asghar, F. M. Kundi, S. Ahmad, A. Khan, and F. Khan, "T-SAF: Twitter sentiment analysis framework using a hybrid classification scheme," Expert Systems, vol. 35, no. 1, 2018.

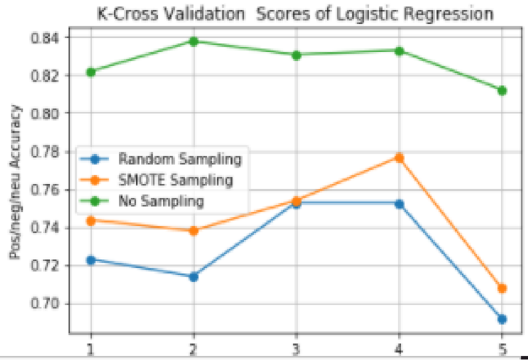
[25] N. Pagar and B. S. Satpue, "Survey Paper on Hybrid Approach for Twitter Sentiment Analysis using Supervised Machine Learning Algorithms," International Journal of Engineering Research & Technology (IJERT), vol. 9, no. 1, pp. 513–515, Jan. 2020.

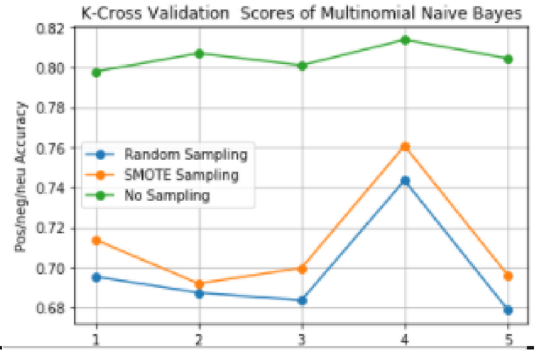
# Appendix

1. https://ieee-dataport.org/open-access/coronavirus-covid-19-tweets-dataset

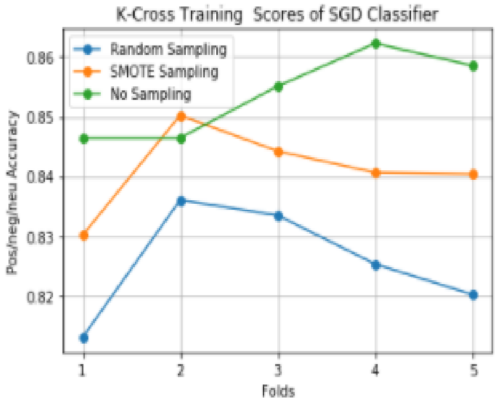


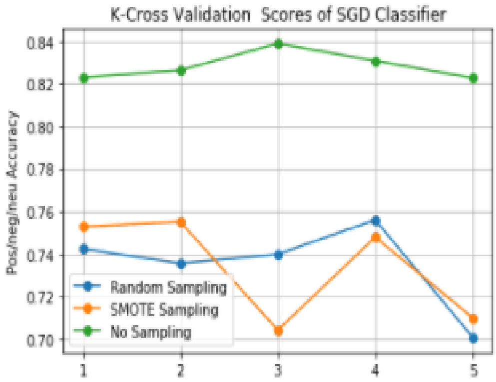



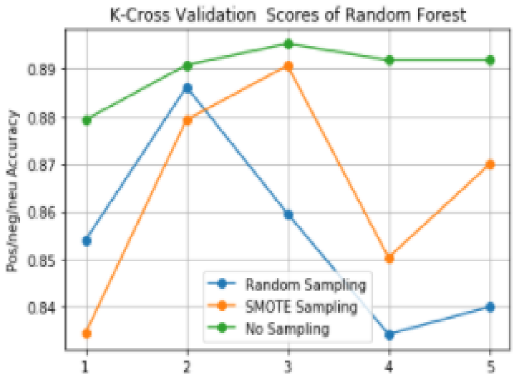



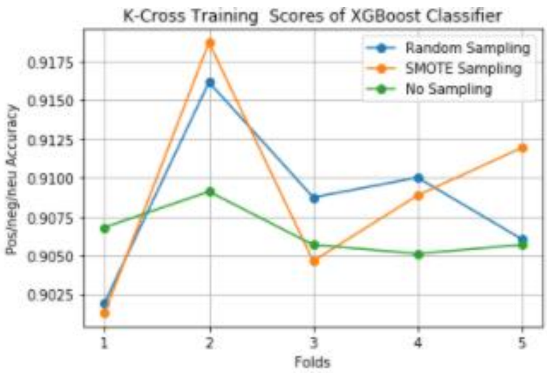

2.

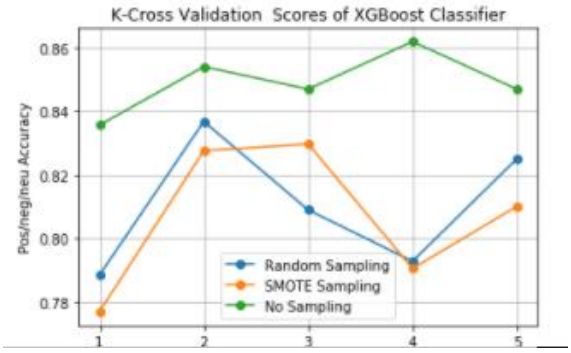






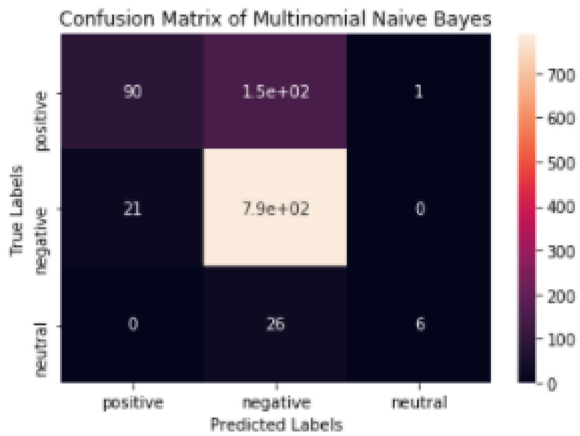


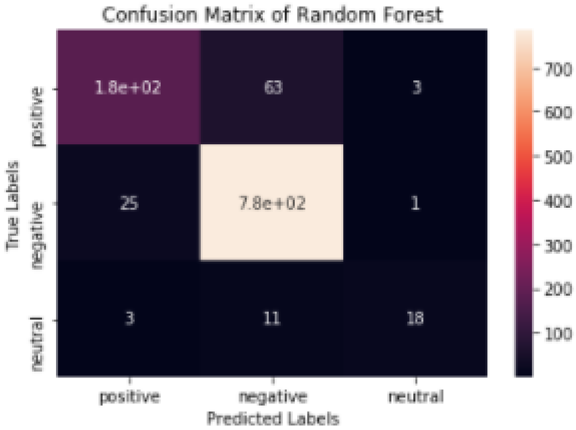

```
Training Accuracy :  84.79
Testing Accuracy  :  83.99
```


3.

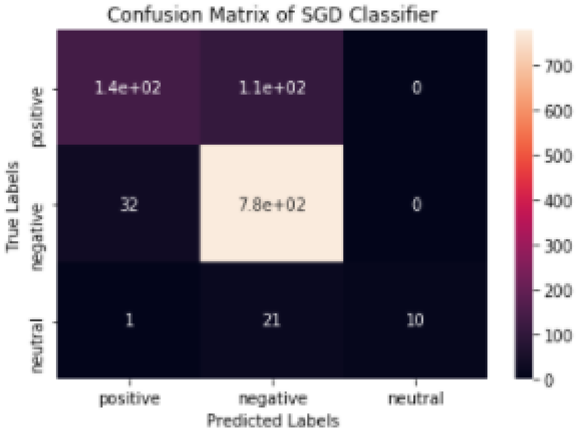
```
Training Accuracy :  81.96
Testing Accuracy  :  81.42
```



```
Training Accuracy :  97.82
Testing Accuracy  :  90.25
```




```
Training Accuracy :  85.8
Testing Accuracy  :  85.1
```




```
Training Accuracy :  90.55
Testing Accuracy  :  85.19
```