SVKM's NMIMS
Mukesh Patel School of Technology Management & Engineering

Final Report on

Search and analyze Customer Complaints, and Feedback received via social media and classifying tweets on the basis of topics, using ML and Natural Language Processing primarily targeting DMRC

Submitted by:

ROHIT BHATIA

Under the guidance of

Ms. KAVYA SHARMA
(Assistant Manager IT DMRC)

And

Prof. RUCHI SHARMA
(Professor, Dept. of Information Technology, MPSTME, Mumbai)

DELHI METRO RAIL CORPORATION Ltd.

दिल्ली मेट्रो रेल कॉरपोरेशन लिमिटेड
Delhi Metro Rail Corporation Limited

A Report submitted in partial fulfilment of the requirements of 5 years integrated MBA (Tech)Program of Mukesh Patel School of Technology Management & Engineering, NMIMS.

SVKM's NMIMS

Mukesh Patel School of Technology Management & Engineering
Vile Parle (W),Mumbai – 400056

TECHNICAL INTERNSHIP REPORT Semester VII – MBA (TECH)

Submitted in Partial Fulfillment of the requirements for Technical Project/Training for VII Semester MBA – (Tech).

Name of the Student: **Rohit Bhatia**.

Roll No. :: **I006**

Batch: **2019-2024**

Academic Year: **2021-2022**.

Name of the Discipline: **MBA-Tech (Information Technology)**

Name and Address of the Company: **Delhi Metro Rail Corporation Ltd.,** Mayur Vihar Phase-1, Ground Floor, Metro Station, New Delhi -110091.

Training Period: **2nd May, 2022 to 25th June, 2022**

THIS IS TO CERTIFY THAT

Mr. Rohit Bhatia

Exam Seat No. **70411019006** has Satisfactorily Completed his Training/Project Work, submittedthe training report and appeared for the Presentation & Viva as required.

External Examiner **Internal Examiner** **Head of Dept.** **Chairperson/Dean**

Date:
Place:
Seal of the University:

ACKNOWLEDGEMENT

I would like to express my gratitude to Delhi Metro Rail Corporation Ltd. For giving me the opportunity to be a part of their prestigious organization as an intern and letting me conduct the project. The internship has been a great learning curve in terms of gaining professional technical knowledge and learning to apply it in real world scenarios through interactions with generous industry professionals.

I am deeply indebted to my industry mentor Ms. Kavya Sharma for all his valuable feedback and support during the project. His inputs and suggestions have been no less than monumental for the successful completion and execution of the project. Despite being occupied with his own duties, he managed to take out the time to listen, guide me to the correct path and always made it a point to constantly check up on my progress.

My sincere gratitude to my academic/faculty mentor Prof. Ruchi Sharma for all her cooperation and guidance and taking out the time to convey all the necessary formalities and information pointers required to meet the internship expectations. I immensely appreciate her being available to clear my doubts and help me get acquainted with the college rules and internship guidelines.

I perceive this opportunity as a huge milestone in my education and career development. I will strive to use the skills and knowledge acquired through this internship in the best way possible, and I will continue to work on improving the same, to attain future career goals and objectives. I look forward to continued cooperation and guidance from all of you in the future.

Sincerely,

Rohit Bhatia

25th June 2022

TABLE OF CONTENTS

ABSTRACT

The Technical Internship Program (TIP) is an important component of education at MPSTME. It is an attempt to bridge the gap between the academic institution and the corporate world. At MPSTME, students undertake 8-week TIP at any Organization during the summer vacation between 6th & 7th Semesters.

The internship is supposed to be a simulation of real work environment, requires that the students undergo the rigor of professional environment both in form and substance.

In the process it provides an opportunity for students, to satisfy their inquisitiveness to know more details, expose them to technical skills, and helps them to acquire social skills by drawing them into communication with outside professionals for continuous interaction. It also enables students to apply their technical knowledge in real world situations and augment their technical skills by dealing with real time problem scenarios.

Students are expected to find solutions to various problems confronted in the project, involving an element of analytical thinking, processing, and decision – making in the scope of the project and in the best interests of the host organization.

The purpose of this report is to document the knowledge gained by the me during the internship period and provide a brief description of the project undertaken and the progress till date.

COMPANY PROFILE

दिल्ली मेट्रो रेल कॉरपोरेशन लिमिटेड
Delhi Metro Rail Corporation Limited

The Delhi Metro has been instrumental in ushering in a new era in the sphere of mass urban transportation in India. The swanky and modern Metro system introduced comfortable, air-conditioned and eco-friendly services for the first time in India and completely revolutionized the mass transportation scenario not only in the National Capital Region but the entire country. Having constructed a massive network of about 390.14 Km with 286 stations (including NOIDA-Greater NOIDA Corridor and Rapid Metro, Gurugram) in record time in Delhi, NCR, the DMRC today stands out as a shining example of how a mammoth technically complex infrastructure project can be completed before time and within budgeted cost by Government agency The Delhi Metro Rail Corporation Limited (DMRC) was registered on 3rd May 1995 under the Companies Act, 1956 with equal equity participation of the Government of the National Capital Territory of Delhi (GNCTD) and the Central Government to implement the dream of construction and operation of a world- class Mass Rapid Transport System (MRTS).

1. Introduction

1.1 Project Description

Sentiment analysis is the systematic recognition, extraction, quantification, and learning of emotional states and personal information using natural language, text analysis, computational linguistics, and biometrics. People frequently use Twitter, one of numerous prominent social media platforms, to convey their thoughts and opinions about a business, a product, or service. Analysis of tweet sentiments is instrumental in detecting whether individuals have a good, negative, or neutral attitude towards a product or a service.

To help DMRC improve its services, I intend to build an algorithm that can find all complaints, feedback, and announcements that can be classified as positive or negative and help in further analysis of feedback to improve service. Further the tweets are arranged according to the topics and helps in segregating the tweets according to the topics. These analysis and results are needed to be shown in a Graphical user interface.

1.2 Technical Knowledge gain

1.2.1 Data Extraction

Data extraction is the process of obtaining information from a single source and transferring it to a new location, whether on-site, on the cloud, or a combination of both. To this goal, a variety of tactics are used, some of whichare difficult and are often carried out manually. Extraction is usually the initial stage in the ETL process of Extraction, Transformation, and Loading, unless data is being extracted only for archiving purposes. This means that data is almost always processed further after initial retrieval in order to make it usable for future study.

There are three steps in the ETL process:

1. Extraction: Data is taken from one or more sources or systems. The extraction locates and identifies relevant data, then prepares it for processing or transformation. Extraction allows many kinds of data to be combined and ultimately mined for business intelligence.
2. Transformation: Once the data has been successfully extracted, it is ready to be refined. During the transformation phase, data is sorted, organized, and cleansed. For example, duplicate entries will be deleted, missing values removed or enriched, and audits will be performed to produce data that is reliable, consistent, and usable.
3. Loading: The transformed, high quality data is then delivered to a single, unified target location for storage and analysis.

1.2.2 Data Extraction using Twitter API

The Tweepy library may be used to retrieve tweets from the Twitter API based on specific themes.

To use the Twitter API, you'll need a Twitter account with developer access. It might take 2–3 hours for the request to be approved. Once you've completed the setup, build an app in which you'll find keys and tokens that will assist us in retrieving data from Twitter. They serve as credentials for logging in.
To begin, import all of the necessary packages and set up the token and key variables. OAuth essentially allows a user to offer another website/service a restricted access authentication token for authorisation to additional resources via an authentication provider with which they have previously successfully authenticated.

1.2.3 Data Cleaning

A tweet can contain a variety of elements, including simple text, mentions, hashtags, links, punctuation, and more. When working on a data science or machine learning project, it's a good idea to delete these items first before continuing to analyse the tweets.

- Lowercasing all the letters
- Removing hashtags and mentions: Hashtags and mentions are common in tweets. There are cases where you want to removethem so you only get the clean content of a tweet without allthese elements. You can remove these hashtags and mentionsusing regex.
- Removing links: In most cases, links aren't required for textprocessing, thus it's best to leave them out.
- Removing punctuations
- Filtering non-alphanumeric characters
- Tokenization: The tweet's words are tokenized. Tokenizationis breaking down a string sequence into parts such as phrases, words, keywords, symbols, and other tokens. Tokenization helps delete undesirable terms from text, suchas the unique symbol linked with the username and hashtagin tweets.
- Stop words: Stop words, which are words that do not changethe meaning of a phrase, are deleted, which reduces the work of categorizing every word of a tweet by lowering the number of words to compare.
- Stemming: Stemming is used to decrease words to harmonizethem across papers and make it easier to group related termsinto categories. Cleaning tweets involves deleting text additions such as URLs, numerals, and special characters, reducing the size of tweets for comparison.

1.2.4 Tweets Classification using TextBlob

TextBlob is a Python (2 and 3) library for processing textual data. It provides a simple API for diving into common natural language processing (NLP) tasks such as part-of-speech tagging, noun phrase extraction, sentiment analysis, classification, translation, and more.
With the use of sentiment analysis, we may learn important details about the environment as well as the general attitude and sentiments of the populace. Data evaluation and categorization in accordance with requirements is the process of sentiment analysis.
TextBlob returns the subjectivity and polarity of a statement. Polarity can range from [-1,1], with -1 denoting a negative attitude and 1 denoting a positive one. To shift a sentence's polarity, use negative language. In TextBlob, semantic labels provide fine-grained analysis. For instance, emojis,

exclamation points, and emoticons. Subjectivity lies inside the [0, 1] number range. Subjectivity is a metric for determining how much of a text is made up of fact and personal opinion. The writing is more subjective than usual, therefore it incorporates personal opinion rather than objective data. Intensity is a further parameter in TextBlob. TextBlob computes subjectivity using the 'intensity'. Whether a word alters the one after it depends on how strong it is. In English, adverbs are employed as modifiers.

The TextBlob's sentiment property delivers a named tuple containing polarity and subjectivity scores after receiving a sentence as input. The subjectivity score spans from 0.0 to 1.0, where 1 represents a subjective assertion, while the polarity score extends from -1.0 to 1.0.

1.2.5 Exploratory Data Analysis/Data Visualization

The graphical display of information and data is known as data visualisation.Data visualisation tools make it easy to observe and comprehend trends, outliers, and patterns in data by employing visual components like charts, graphs, and maps.

Types of Data Visualization

1. Column Chart

It is one of the most widely used data visualisation programmes. A column isa quick and easy way to demonstrate comparisons across distinct data sets.The X-axis of a column chart contains data labels, while the Y-axis displays values or measured metrics. Colors can be used to highlight critical data points in order to track change.

2. Bar graph

Bar graphs are great for comparing more than 10 objects and dealing with more detailed labels that exhibit negative integers.

3. Word cloud

The size of each word represents its frequency or relevance in a word cloud, which is a data visualisation tool for visualizing text data. A word cloud can be used to emphasize important textual data points. Data from social networking websites is frequently analyzed using word clouds.

1.2.6 Model Building

- Naïve bayes

  A Naive Bayes classifier is a probabilistic machine learning model that'sused for classification task. The crux of the classifier is based on the Bayes theorem.

  Types of Naive Bayes Classifier:

  - Multinomial Naive Bayes:

This is mostly used for document classification problem, i.e whether adocument belongs to the category of sports, politics, technology etc. The features/predictors used by the classifier are the frequency of thewords present in the document.

- Bernoulli Naive Bayes:

This is similar to the multinomial naive bayes but the predictors are boolean variables. The parameters that we use to predict the class variable take up only values yes or no, for example if a word occurs in the text or not.

- Gaussian Naive Bayes:

When the predictors take up a continuous value and are not discrete,we assume that these values are sampled from a gaussian distribution.

- SVM

SVM is a supervised machine learning technique that may be used forboth classification and regression. Though we might also argue regression difficulties, categorization is the best fit. The goal of the SVM method is to discover a hyperplane in an N-dimensional space that categorises data points clearly. The hyperplane's size is determined by the number of features. If there are just two input characteristics, the hyperplane is merely a line. When the number of input characteristics reaches three, the hyperplane transforms into a two-dimensional plane. When the number of characteristics exceeds three, it becomes impossible to imagine.

- Logistic regression

Logistic regression estimates the probability of an event occurring, such as voted or didn't vote, based on a given dataset of independentvariables. Since the outcome is a probability, the dependent variable is bounded between 0 and 1. In logistic regression, a logit transformation is applied on the odds—that is, the probability ofsuccess divided by the probability of failure. This is also commonly known as the log odds, or the natural logarithm of odds, and this logistic function is represented by the following formulas:

Logit(pi) = 1/(1+ exp(-pi))

ln(pi/(1-pi)) = Beta_0 + Beta_1*X_1 + … + B_k*K_k

- KNN

K-NN algorithm stores all the available data and classifies a new data point based on the similarity. This means when new data appears then it can be easily classified into a well suite category by using K- NN algorithm. It is also called a lazy learner algorithm because it does not learn from the training set immediately instead it stores the dataset and at the time of classification, it performs an action on the dataset.

- Adaboost

AdaBoost, also known as Adaptive Boosting, is a machine learning method used in an ensemble setting. Decision trees with one level, or Decision trees with only one split, are the most popular method used with AdaBoost. Decision Stumps is another name for these trees.

1.2.7 Topic Modelling

In order to capture how the meaning of words is reliant upon the larger context in which they are employed in natural language, topic modeling is one of a family of text analysis techniques that analyzes "bags" or groupings of words together rather than counting them separately. The identification of clustering within texts has also been done using other approaches, such as cluster analysis and latent semantic analysis, in addition to topic modeling.

LDA is used to classify text in a document to a particular topic. It builds a topic per document model and words per topic model, modeled as Dirichlet distributions.
Each document is modeled as a multinomial distribution of topics and each topic is modeled as a multinomial distribution of words.
LDA assumes that every chunk of text we feed into it will contain words that are somehow related. Therefore choosing the right corpus of data is crucial.
It also assumes documents are produced from a mixture of topics. Those topics then generate words based on their probability distribution.

2. Project Outline

2.1 Proposed Schedule

Week 1: Gather all the background information. Research on the Technology and framework to be used in order to improve the model's performance and aid in the architectural design (**Completed**)

Week 2: Install the software and necessary library to perform the task. **(Completed)**

Week 3: Data collection **(Completed)**

Week 4: Data preprocessing **(Completed)**

Week 5: Model Implementation, Performance Tuning, and Identify Best Fit Model **(Completed)**

Week 6: Visualization **(Completed)**

Week 7: Topic modeling **(Completed)**

Week 8: Final changes and Model finalization **(Completed)**

2.2 Project Requirement Analysis

2.2.1 Functional Requirements

- Using legitimate APIs and ethical techniques, extract data from the organization's social media. Organize and make this information available.
- Download the python 3.9.0 on the system to run the code or google colab can be used.
- Some of the python Libraries are preinstalled and have to be installed.
- The tweepy library should be of version 3.9 or less to make it work.
- Using the Tweets classification, study the data that is positive, negative , neutral.
- Graphs are shown as the python.exe format. Close the graph to make code run to the next part.
- Using the different models , model can be trained to its highest accuracy possible and can be used further in different programs . The data can be analysed and confusion matrix can help draw various conclusion
- Using topic modelling, analyze the data that was extracted. In a comparison manner, the final solution should incorporate the themes that are frequently cited by these groups (may include the use of word clouds).
- Draw conclusions from the study such that typical qualities about the organisation may be derived from the end result.

2.2.2 Non-Functional Requirements

- Conduct research on given use case and identify possible inoperability from a technical standpoint. If any, find viable solutions to these. If there are no solutions to the issues, look for changes in the use case (or ask for clarification).
- Clean the data before beginning analysis. Lowercase all words, remove digits, punctuation marks, stop words etc.
- Time needed to extract the tweets is according to no of tweets .
- Perform lemmatization and vectorize the data for feeding into the models.
- Attempt all possible models and justify the use of the final model with relevant reasoning (comparisons).

2.3 Use Case

The user needs to enter the account whose analysis need to be done and the date from which the tweets are requires. This program can be used by different companies to find their weak points or their strong points for a certain period of time or to check the performance of theirs in a span of time .

Use Case 1: Social media monitoring:

Because they are uninvited, social media posts frequently include some of the most frank reviews of your goods, services, and companies. With the use of sentiment analysis tools, you can quickly sift through all of that data to examine both the feelings of specific people and the general public as a whole across all social media platforms .Sentiment analysis is able to read beyond simple definition to recognise sarcasm, understand popular chat acronyms (lol, rofl, etc.), and correct for grammatical and spelling errors that are frequently made.

Use Case 2: Customer support

Customer support management presents many challenges due to the sheer number of requests, varied topics, and diverse branches within a company – not to mention the urgency of any given request. Sentiment analysis with natural language understanding (NLU) reads regular human language for meaning, emotion, tone, and more, to understand customer requests, just as a person would. You can automatically process customer support tickets, online chats, phone calls, and emails by sentiment to prioritize any urgent issues.

Use Case 3: Brand monitoring and reputation management

One of the most often used uses of sentiment analysis in business is brand monitoring. Online negative reviews may snowball, and the longer you leave them, the worse off you'll be. You will receive fast notification when a brand is mentioned negatively thanks to sentiment analysis technologies.

Use Case 4: Listen to voice of the customer (VoC)

All of your consumer input from the web, surveys, chats, call centers, and emails should be combined and evaluated. You may organize and categorize this data using sentiment analysis to spot trends and find recurrent issues.

Use Case 5: Product Analysis

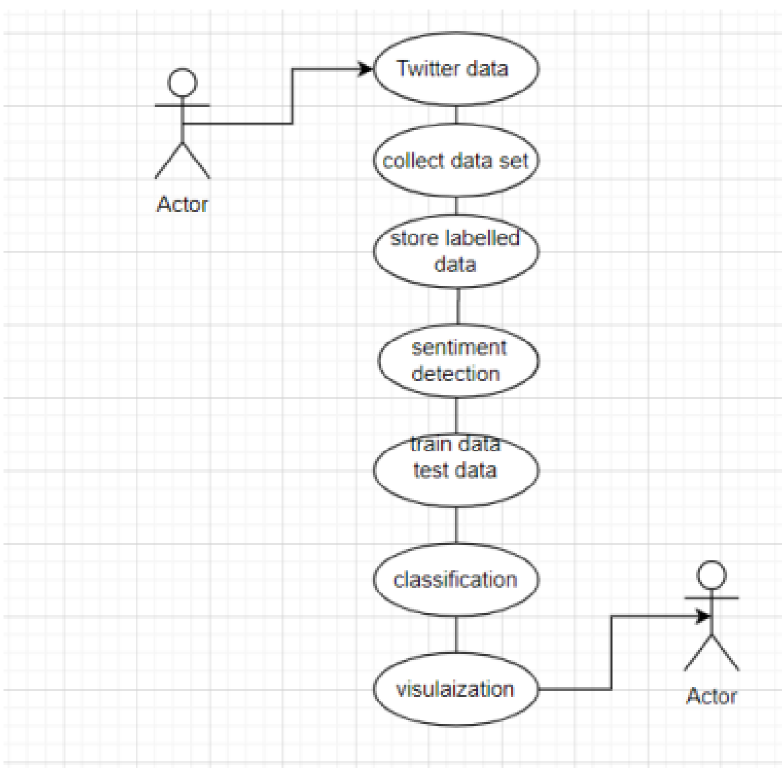
Analyze years of feedback you may not have noticed to learn what the public is saying about a new product as soon as it is released. To discover exactly the information you want, you may conduct a keyword search for a specific product feature (interface, UX, functionality).

Learn how a product is seen by your target market, what has to be changed about it, and what would satisfy your most valued clients. with sentiment analysis in each.

Use Case 6: Market and Competitor Research

For market and competition research, use sentiment analysis. Find out who among your rivals is being mentioned favorably and how your marketing initiatives stack up.

Examine the uplifting language that your rivals are using to address their clients and incorporate some of it into your own brand message and style manual.



3. Implementation and Design Aspect

3.1 Approach

- Tweets are collected by taking the input from the Twitter API in the form of hashtags, and the number of tweets that will be considered is to be restricted between 5 and 5000.
- The following steps were followed for data pre-processing: Lowercasing, removing punctuations, Removing URLs, Removing Emojis/Emoticons. Removing words with numbers, Removing stopwords, Lemmatization, Vectorization
- After that using Text Blob library, polarity and subjectivity is calculated. Based on the score received from them, Tweets are assigned as neutral, positive, or negative
- To make it seem more easier, Label encoding is done. 0: negative tweet, 1: neutral tweet, 2: positive tweet
- Positive tweets are stored in data_positive.csv, and negative Tweets are stored in data_negative.csv
- Now data visualization is done to be on the tweets. Word clouds are made of all the tweets, positive and negative tweets. Graphs of tweets like the word length, no of tweets, average word length are made.
- The next step includes the model creation based on no of tweets
- Several models like KNN, SVC, Random Forest, Naïve bayes and Adaboost classifier are used
- The classification report of all the models are shown and each accuracy is determined and stored in a dictionary for further use
- The maximum accuracy model is now used to find if the the input sentence is giving positive or neutral or negative reaction . in our case maximum accuracy is being given by Random Forest Classifier
- Now the tweets are to be segregated topic wise for which the concept of topic modeliing is applied .
- A copy of all the tweets is now stored in a separate file . the most occurred words in the tweets are stored in a dictionary and data preprocessing is performed again
- Using the Linear Discriminant Analysis(LDA) technique, modeling is performed on the tweets and one with max score is assigned as an topic to the tweet .
- To make LDA perform better , the TF-idf vectorization is performed and model is trained to see better results .

3.2 Screen shots

Welcome page

Now the user needs to enter the information off the Twitter account whose complaints turn tweets needs to be extracted after that a valid date has to be entered which is from which date till present you need to see the tweets start you can also instruct how many tweets the user want.

```
\    /\    /  _          _    _        _
 \  /  \  /  |_  |    |   | |  |\/| |_
  \/    \/   |_  |_   |_  |_|  |  | |_

 ********************Enter information*************************
Enter accounts separated by comma @officialDMRC,#delhimetro,#DMRC,@ArvindMedi,@DCP_DelhiMetro
Enter a date in YYYY-MM-DD format2020-01-01
how many tweets u want  3000
```

Now the cleaning of tweets is done, and subjectivity and polarity of the tweet is found using the text blob module. now according to the subjectivity and polarity of the tweet ,analysis of the tweet is done that is if it is positive negative or neutral .

```
\    /\    /  _          _    _        _
 \  /  \  /  |_  |    |   | |  |\/| |_
  \/    \/   |_  |_   |_  |_|  |  | |_

 ********************Enter information*************************
Enter accounts separated by comma @officialDMRC,#delhimetro,#DMRC,@ArvindMedi,@DCP_DelhiMetro
Enter a date in YYYY-MM-DD format2020-01-01
how many tweets u want  3000
Total Tweets fetched: 2754
****    Cleaning in Process  *****
************Classifying Tweets**************
**********************************************************
**********************************************************
***     1 .....Show Tweets                           ****
***     2 .....Show graphs                           ****
***     3 .....Model creation                        ****
***     4 .....Topic Modelling                       ****
***     5 .....Tweets sentiments using model created ****
***     6 ..... Exit                                 ****
**********************************************************
**********************************************************
Enter a choice      ::::
```

This is the user interface of the project. there are 6 options the first option includes to show the tweets next includes to show the graph 3rd option includes the model creation the 4th option includes a topic modelling add the 5th includes the input from the user add classifying it based on model created

Selecting choice 1 to see the tweets

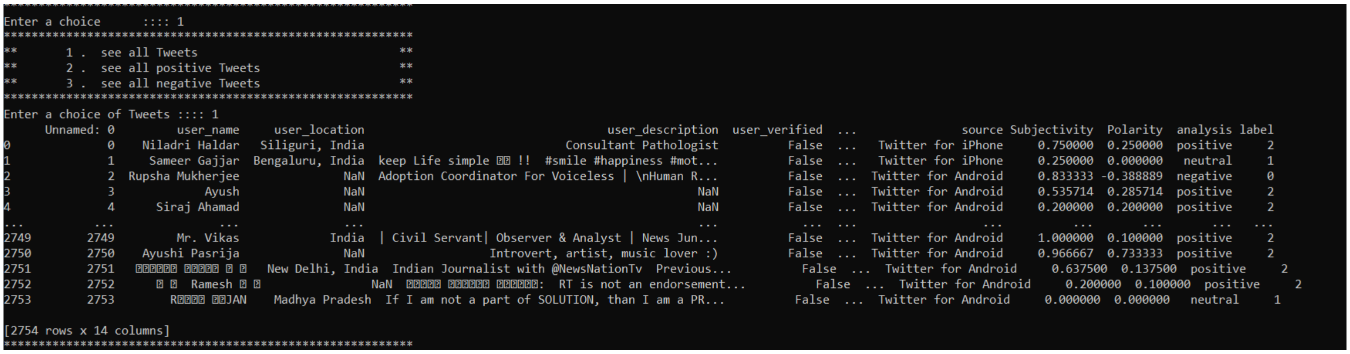
```
\    /\    /  _          _   _        _
 \  /  \  /  |_  |     |    | |  |\/| |_
  \/    \/   |_  |_    |_   |_|  |  | |_

 ********************Enter information*************************
Enter accounts separated by comma @officialDMRC,#delhimetro,#DMRC,@ArvindMedi,@DCP_DelhiMetro
Enter a date in YYYY-MM-DD format2020-01-01
how many tweets u want  3000
Total Tweets fetched: 2754
****    Cleaning in Process  *****
*************Classifying Tweets**************
***********************************************************
***********************************************************
***     1 .....Show Tweets                               ****
***     2 .....Show graphs                               ****
***     3 .....Model creation                            ****
***     4 .....Topic Modelling                           ****
***     5 .....Tweets sentiments using model created     ****
***     6 ..... Exit                                     ****
***********************************************************
***********************************************************
Enter a choice      :::: 1
```

There is an option to see all the tweets ,positive tweets, negative tweets

Show tweets

```
Command Prompt - python  v5_final_project_with_input.py
\    /\    /  _          _   _        _
 \  /  \  /  |_  |     |    | |  |\/| |_
  \/    \/   |_  |_    |_   |_|  |  | |_

 ********************Enter information*************************
Enter accounts separated by comma @officialDMRC,#delhimetro,#DMRC,@ArvindMedi,@DCP_DelhiMetro
Enter a date in YYYY-MM-DD format2020-01-01
how many tweets u want  3000
Total Tweets fetched: 2754
****    Cleaning in Process  *****
*************Classifying Tweets**************
***********************************************************
***********************************************************
***     1 .....Show Tweets                               ****
***     2 .....Show graphs                               ****
***     3 .....Model creation                            ****
***     4 .....Topic Modelling                           ****
***     5 .....Tweets sentiments using model created     ****
***     6 ..... Exit                                     ****
***********************************************************
***********************************************************
Enter a choice      :::: 1
***********************************************************
**       1 .  see all Tweets                               **
**       2 .  see all positive Tweets                      **
**       3 .  see all negative Tweets                      **
***********************************************************
Enter a choice of Tweets ::::
```

1. All Tweets



```
************************************************************
Enter a choice      :::: 1
************************************************************
**       1 .  see all Tweets                              **
**       2 .  see all positive Tweets                     **
**       3 .  see all negative Tweets                     **
************************************************************
Enter a choice of Tweets :::: 1
      Unnamed: 0         user_name     user_location                                 user_description  user_verified  ...              source Subjectivity  Polarity  analysis label
0              0    Niladri Haldar   Siliguri, India                             Consultant Pathologist           False  ...   Twitter for iPhone     0.750000  0.250000  positive     2
1              1     Sameer Gajjar  Bengaluru, India  keep Life simple ?? !!  #smile #happiness #mot...           False  ...   Twitter for iPhone     0.250000  0.000000   neutral     1
2              2  Rupsha Mukherjee               NaN  Adoption Coordinator For Voiceless | \nHuman R...           False  ...  Twitter for Android     0.833333 -0.388889  negative     0
3              3             Ayush               NaN                                                NaN           False  ...  Twitter for Android     0.535714  0.285714  positive     2
4              4      Siraj Ahamad               NaN                                                NaN           False  ...  Twitter for Android     0.200000  0.200000  positive     2
...          ...               ...               ...                                                ...             ...  ...                  ...          ...       ...       ...   ...
2749        2749         Mr. Vikas             India  | Civil Servant| Observer & Analyst | News Jun...           False  ...  Twitter for Android     1.000000  0.100000  positive     2
2750        2750    Ayushi Pasrija               NaN                     Introvert, artist, music lover :)           False  ...  Twitter for Android     0.966667  0.733333  positive     2
2751        2751   ?????? ????? ? ?   New Delhi, India  Indian Journalist with @NewsNationTv  Previous...           False  ...  Twitter for Android     0.637500  0.137500  positive     2
2752        2752       ? ?  Ramesh ? ?               NaN  ????? ?????? ??????:  RT is not an endorsement...           False  ...  Twitter for Android     0.200000  0.100000  positive     2
2753        2753        R???? ??JAN    Madhya Pradesh  If I am not a part of SOLUTION, than I am a PR...           False  ...  Twitter for Android     0.000000  0.000000   neutral     1

[2754 rows x 14 columns]
************************************************************
```

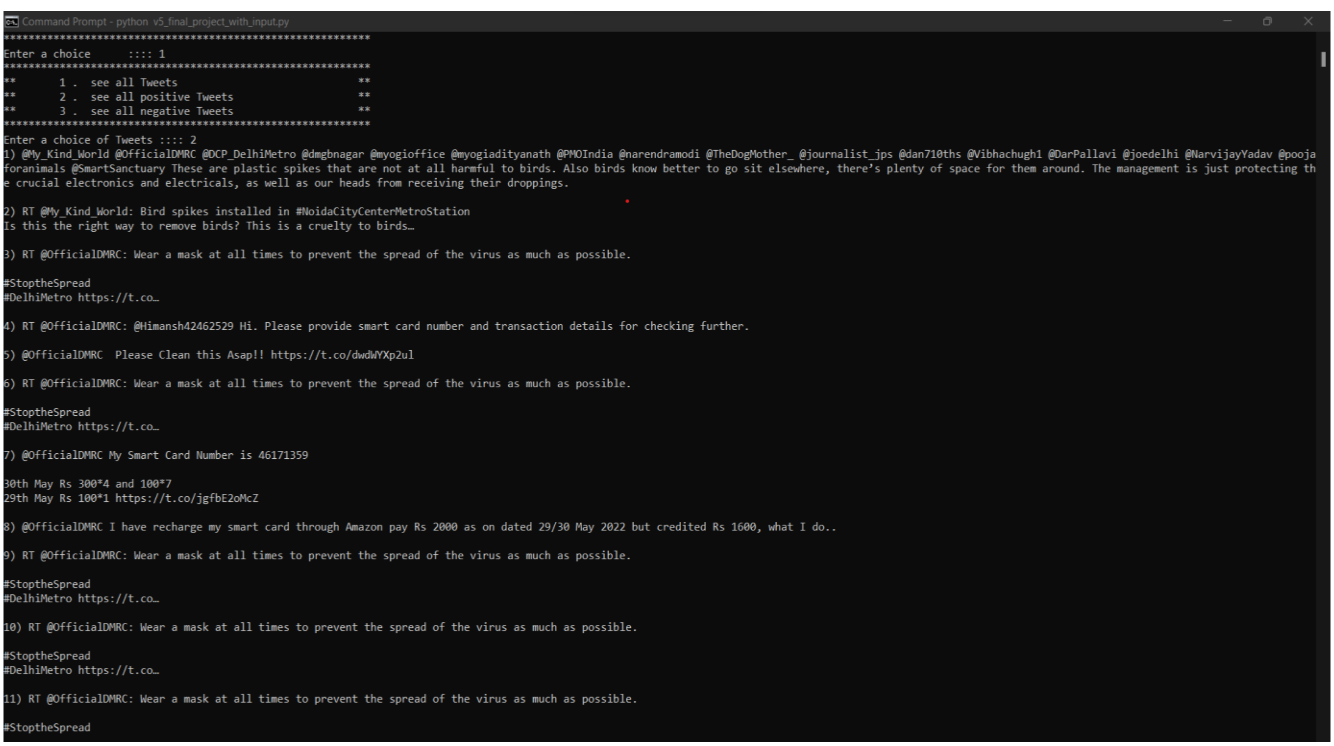
Converted to csv (data.csv)

	A	B	C	D	E	F	G	H
1		user_name	user_locat	user_desci	user_verifi	date	text	original
2	0	Niladri Hal	Siliguri, Ind	Consultant	FALSE	########	@My_Kinc	@My_Kind_World @OfficialDMRC @DCP_DelhiMetro @dmgbnagar @myogioffice @myogiadityanath @PMOIndia @narendramodi @TheDogMother_ @journalist_jps @dan710
3	1	Sameer Ga	Bengaluru,	keep Life s	FALSE	########	@concern	@concern13161846 @hebbaltrafficps @blrcitytraffic Requesting the authorities here plz take some guidance from Central Govt. to give some relief to Bengaluru citizens, the city
4	2	Rupsha Mukherjee		Adoption	FALSE	########	@OfficialD	@OfficialDMRC Thanks for the revert. Instead of spikes pls use a different method to prevent. Birds are very fragile if they gets hurt it's very tough to treat them. Hope we will see
5	3	Ayush			FALSE	########	RT	RT @My_Kind_World: Bird spikes installed in #NoidaCityCenterMetroStation
6	4	Siraj Ahamad			FALSE	########	RT	RT @OfficialDMRC: Wear a mask at all times to prevent the spread of the virus as much as possible.
7	5	Robert		I am difere	FALSE	########	@My_Kinc	@My_Kind_World @OfficialDMRC @DCP_DelhiMetro @dmgbnagar @myogioffice @myogiadityanath @PMOIndia @narendramodi @TheDogMother_ @journalist_jps @dan710
8	6	Rajeev Mit	Navi Mum	à¤¬à¤¦à¤	FALSE	########	RT @kgahl	RT @kgahlot: There have been recurring incidences of breakdowns on various lines of Delhi Metro this month causing inconvenience to passengâ€¦
9	7	SinghBhhupinder		Realistic	FALSE	########	@PriyaKu7	@PriyaKu73754749 @OfficialDMRC Come towards Uttam Nagar West metro station and half the road under metro station there is occupied by the e-rickshaws who travel on th
10	8	Syed			FALSE	########	@My_Kinc	@My_Kind_World @OfficialDMRC @DCP_DelhiMetro @dmgbnagar @myogioffice @myogiadityanath @PMOIndia @narendramodi @TheDogMother_ @journalist_jps @dan710
11	9	Himanshu Ranjan			FALSE	########	RT @Offic	RT @OfficialDMRC: @Himansh42462529 Hi. Please provide smart card number and transaction details for checking further.
12	10	Sipplin		I've Nothir	FALSE	########	@OfficialD	@OfficialDMRC Please Clean this Asap!! https://t.co/dwdWYXp2ul
13	11	sanjeev rawat		Man on mi	FALSE	########	@My_Kinc	@My_Kind_World @OfficialDMRC @DCP_DelhiMetro @dmgbnagar @myogioffice @myogiadityanath @PMOIndia @narendramodi @TheDogMother_ @journalist_jps @dan710
14	12	Narendra.j	New Delhi, Delhi		FALSE	########	@OfficialD	@OfficialDMRC Please provide sufficient salary to your staff along with proper training so that they do not cheat the innocent travelers. This happened at botanical garden metro
15	13	Himanshu Ranjan			FALSE	########	@OfficialD	@OfficialDMRC Please check transaction details, https://t.co/tY2pZiMqsU
16	14	Swati Singh			FALSE	########	RT	RT @OfficialDMRC: Wear a mask at all times to prevent the spread of the virus as much as possible.
17	15	Himanshu Ranjan			FALSE	########	@Official	@OfficialDMRC My Smart Card Number is 46171359
18	16	Sagar			FALSE	########	RT @Jony4	RT @Jony49251436: @kgahlot @OfficialDMRC When will the DTC contract staff do parmanent you... @LtGovDelhi @ArvindKejriwal @msisodia @dtchq_â€¦
19	17	SHIV PARAKASH DWIVED			FALSE	########	@LtGovD	@LtGovDelhi @ArvindKejriwal
20	18	Himanshu Ranjan			FALSE	########	@OfficialD	@OfficialDMRC I have recharge my smart card through Amazon pay Rs 2000 as on dated 29/30 May 2022 but credited Rs 1600, what I do..
21	19	Devender	New Delhi	Kya likhu a	FALSE	########	RT	RT @OfficialDMRC: Wear a mask at all times to prevent the spread of the virus as much as possible.
22	20	Shilpa Gup	Jammu, Ja	Former Th	FALSE	########	@My_Kinc	@My_Kind_World @OfficialDMRC @DCP_DelhiMetro @dmgbnagar @myogioffice @myogiadityanath @PMOIndia @narendramodi @TheDogMother_ @journalist_jps @dan710
23	21	Dinesh Bhatt			FALSE	########	@OfficialD	@OfficialDMRC Why uttam ngr east metro station is dirty compair to others stations.
24	22	Vipin Baro	New Delhi	Senior Bus	FALSE	########	@ChitraAh	@ChitraAhanthem @OfficialDMRC Proper ventilation also matters. There is no reason to wear mask in a crowded metro when ACâ€™s are not working.
25	23	ROHIT		hello	FALSE	########	@kgahlot	@kgahlot @OfficialDMRC When will the DTC contract staff do parmanent you... @LtGovDelhi @ArvindKejriwal @msisodia @dtchq_delhi

data

Ready Accessibility: Unavailable

2. Positive tweets



Command Prompt - python v5_final_project_with_input.py

```
************************************************************
Enter a choice      :::: 1
************************************************************
**       1 .  see all Tweets                              **
**       2 .  see all positive Tweets                     **
**       3 .  see all negative Tweets                     **
************************************************************
Enter a choice of Tweets :::: 2
1) @My_Kind_World @OfficialDMRC @DCP_DelhiMetro @dmgbnagar @myogioffice @myogiadityanath @PMOIndia @narendramodi @TheDogMother_ @journalist_jps @dan710ths @Vibhachugh1 @DarPallavi @joedelhi @NarvijayYadav @pooja
foranimals @SmartSanctuary These are plastic spikes that are not at all harmful to birds. Also birds know better to go sit elsewhere, there's plenty of space for them around. The management is just protecting th
e crucial electronics and electricals, as well as our heads from receiving their droppings.

2) RT @My_Kind_World: Bird spikes installed in #NoidaCityCenterMetroStation
Is this the right way to remove birds? This is a cruelty to birds…

3) RT @OfficialDMRC: Wear a mask at all times to prevent the spread of the virus as much as possible.

#StoptheSpread
#DelhiMetro https://t.co…

4) RT @OfficialDMRC: @Himansh42462529 Hi. Please provide smart card number and transaction details for checking further.

5) @OfficialDMRC  Please Clean this Asap!! https://t.co/dwdWYXp2ul

6) RT @OfficialDMRC: Wear a mask at all times to prevent the spread of the virus as much as possible.

#StoptheSpread
#DelhiMetro https://t.co…

7) @OfficialDMRC My Smart Card Number is 46171359

30th May Rs 300*4 and 100*7
29th May Rs 100*1 https://t.co/jgfbE2oMcZ

8) @OfficialDMRC I have recharge my smart card through Amazon pay Rs 2000 as on dated 29/30 May 2022 but credited Rs 1600, what I do..

9) RT @OfficialDMRC: Wear a mask at all times to prevent the spread of the virus as much as possible.

#StoptheSpread
#DelhiMetro https://t.co…

10) RT @OfficialDMRC: Wear a mask at all times to prevent the spread of the virus as much as possible.

#StoptheSpread
#DelhiMetro https://t.co…

11) RT @OfficialDMRC: Wear a mask at all times to prevent the spread of the virus as much as possible.

#StoptheSpread
```

Converted to csv

1		original
2	0	@My_Kind_World @OfficialDMRC @DCP_DelhiMetro @dmgbnagar @myogioffice @myogiadityanath @PMOIndia @narendramodi @TheDogMother_ @journalist_jps @dan710ths @Vibhachugh1 @DarPallavi @joedelhi @NarvijayYadav @
3	3	RT @My_Kind_World: Bird spikes installed in #NoidaCityCenterMetroStation
4	4	RT @OfficialDMRC: Wear a mask at all times to prevent the spread of the virus as much as possible.
5	9	RT @OfficialDMRC: @Himansh42462529 Hi. Please provide smart card number and transaction details for checking further.
6	10	@OfficialDMRC Please Clean this Asap!! https://t.co/dwdWYXp2ul
7	14	RT @OfficialDMRC: Wear a mask at all times to prevent the spread of the virus as much as possible.
8	15	@OfficialDMRC My Smart Card Number is 46171359
9	18	@OfficialDMRC I have recharge my smart card through Amazon pay Rs 2000 as on dated 29/30 May 2022 but credited Rs 1600, what I do..
10	19	RT @OfficialDMRC: Wear a mask at all times to prevent the spread of the virus as much as possible.
11	25	RT @OfficialDMRC: Wear a mask at all times to prevent the spread of the virus as much as possible.
12	30	RT @OfficialDMRC: Wear a mask at all times to prevent the spread of the virus as much as possible.
13	31	RT @OfficialDMRC: Wear a mask at all times to prevent the spread of the virus as much as possible.
14	33	RT @OfficialDMRC: Wear a mask at all times to prevent the spread of the virus as much as possible.
15	34	@OfficialDMRC Coach No M3A08 some kind of sound is continuously coming while metro is moving please look immediately
16	35	@ChitraAhanthem @OfficialDMRC Yes That Happens Frequently!
17	37	RT @OfficialDMRC: Wear a mask at all times to prevent the spread of the virus as much as possible.
18	43	RT @OfficialDMRC: Wear a mask at all times to prevent the spread of the virus as much as possible.
19	47	@OfficialDMRC long lines daily in the hot sunny morning at dwarka metro station this is only due to only one x ray machine need atleast one more machine to go smoothly. https://t.co/ObXmm7HNib
20	51	@RoshanRaj08 @aarnavg17 @OfficialDMRC @AmirSai43048804 @OfficialDMRC DMRC has started operating trains
21	52	@OfficialDMRC
22	68	RT @Nm3530: @CMODelhi @OfficialDMRC @cpsavesoil @ArvindKejriwal @delhi Good initiative, by display in Delhi metro will spread the awarenesâ€¦
23	70	@CMODelhi @OfficialDMRC @cpsavesoil @ArvindKejriwal @delhi Good initiative, by display in Delhi metro will spread the awareness for good reach with more impact . #SaveSoil #SaveSoilMovement #pollution #Gurgaon https://t.co/0Ge
24	85	RT @My_Kind_World: Bird spikes installed in #NoidaCityCenterMetroStation
25	88	@OfficialDMRC alot of metro staff dnt wear mask/keep it down but 2day they knew that v have a conspiracy of fine going on everyone's mask was up according to dis u shud make an announcement 4 the public too that 2day we will fine U
26	90	@OfficialDMRC
27	94	Why does this blue line suck?/.... Why is it hot inside metro... why isn't the ac working...damn!ðŸ˜©
28	108	@OfficialDMRC Coach no M7D033

data_positive

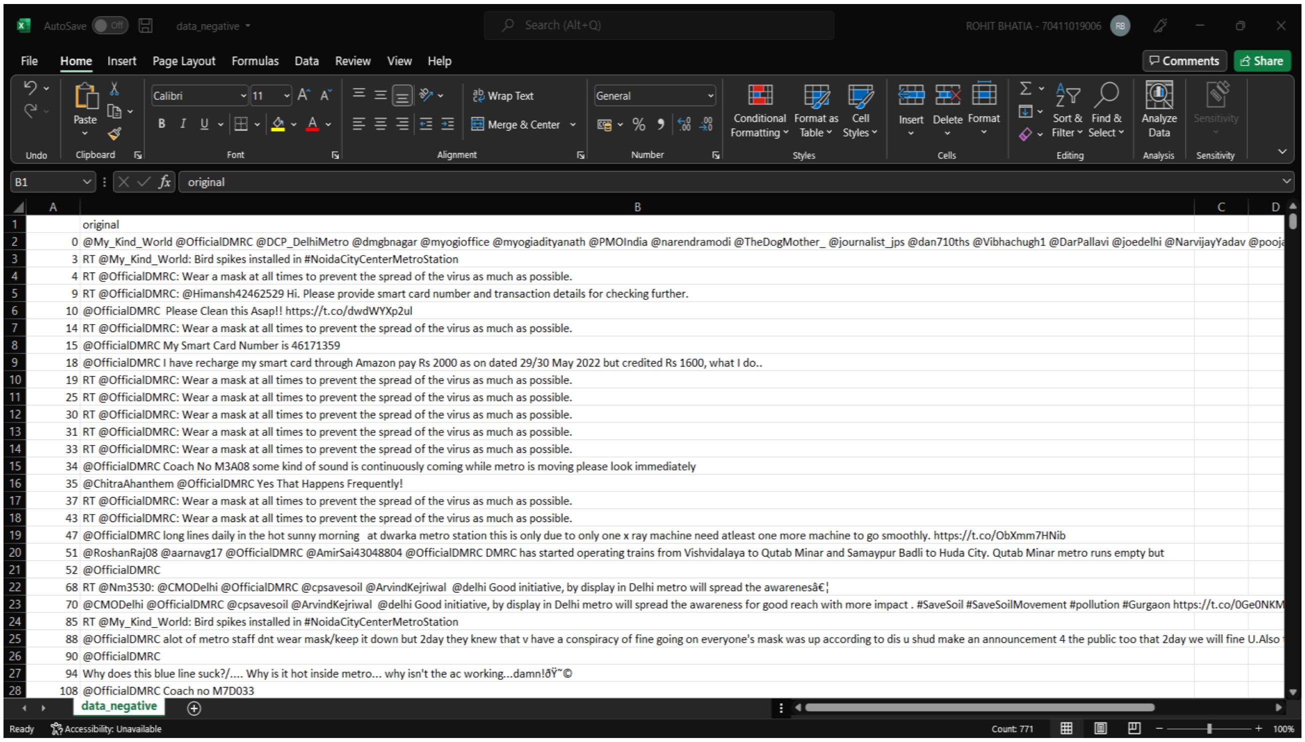
Ready Accessibility: Unavailable

3. Negative tweets

Select Command Prompt - python v5_final_project_with_input.py

```
************************************************************
************************************************************
Enter a choice      :::: 1
************************************************************
**       1 .  see all Tweets                              **
**       2 .  see all positive Tweets                     **
**       3 .  see all negative Tweets                     **
************************************************************
Enter a choice of Tweets :::: 3
1) @OfficialDMRC Thanks for the revert. Instead of spikes pls use a different method to prevent. Birds are very fragile if they gets hurt it's very tough to treat them. Hope we will see a change soon. Thank you.


2) @PriyaKu73754749 @OfficialDMRC Come towards Uttam Nagar West metro station and half the road under metro station there is occupied by the e-rickshaws who travel on the wrong side and impact the flow of traffi
c throughout the day

3) @LtGovDelhi @ArvindKejriwal
Please save the Delhi metro passengers from this fake mask Challan. The corruption has entered even in DELHI METRO also. Please save #delhites
Mask challan ke nam pe Police ki loot band karo @OfficialDMRC  you must take care your customers @AmitShah https://t.co/6jYg2In8oJ

4) @My_Kind_World @OfficialDMRC @DCP_DelhiMetro @dmgbnagar @myogioffice @myogiadityanath @PMOIndia @narendramodi @TheDogMother_ @journalist_jps @dan710ths @Vibhachugh1 @DarPallavi @joedelhi @NarvijayYadav @pooja
foranimals @SmartSanctuary Shame on noida city center metro authority.

5) @OfficialDMRC what's wrong with voilet line metro? Metro is moving at a speed of 10kmph and sudden stops and causing delay and even ac is not working?what a poor transit system.

6) @OfficialDMRC I had done wrong recharge on 22nd May 2022, in which 9 digits were done, so no-one can't topup . But after the month i did not get refund till the date now. Plz ressolve my problem https://t.co/
K3UStqMq6d

7) @RonakRathi23 @OfficialDMRC Bad manners to take anybody's pic

8) @OfficialDMRC we face this issue every day at Dwarka metro station.  Every one standing a long queue, because of single scanned machine.Please look this. https://t.co/dC0MJx0ZXS

9) This is Central Secretariat metro station Gate No.4, the escalators are under repair for a month now @OfficialDMRC Such inconvenience for elderly and females with small children. @DCP_DelhiMetro @ArvindKejriw
al https://t.co/IW5t1e8nu5

10) @OfficialDMRC Why the yellow line running slow for last many days. Yesterday you tweeted that normal services have resumed. But today again it's slow from Arjangarh onwards. Can somebody get this sorted as i
t is resulting into loss of time for commuters. @ndtv @timesofindia

11) @kgahlot @OfficialDMRC @TransportDelhi @ArvindKejriwal @TransportDelhi is worst department, they have lost my RC documents,unko pta hi nhi kuch bhi ,

12) @OfficialDMRC Hire a half-way decent PR agency. These tweets are absurd.

13) I hate this Violet Line - Blue Line bug on @googlemaps  @OfficialDMRC https://t.co/14q8kgHGUz

14) @OfficialDMRC  Hii  small request to dmrc and all station manager. To change ur contract staff late night timing . One of ur contract staff was injured in accident  falling asleep no was helping him on road
.soo plz change there shift timing or hire a contract staff 10 pm.

15) @OfficialDMRC service on yellow line is very slow

16) @ManchNajafgarh @OfficialDMRC @msisodia Complaint pertains to DMRC @OfficialDMRC
```

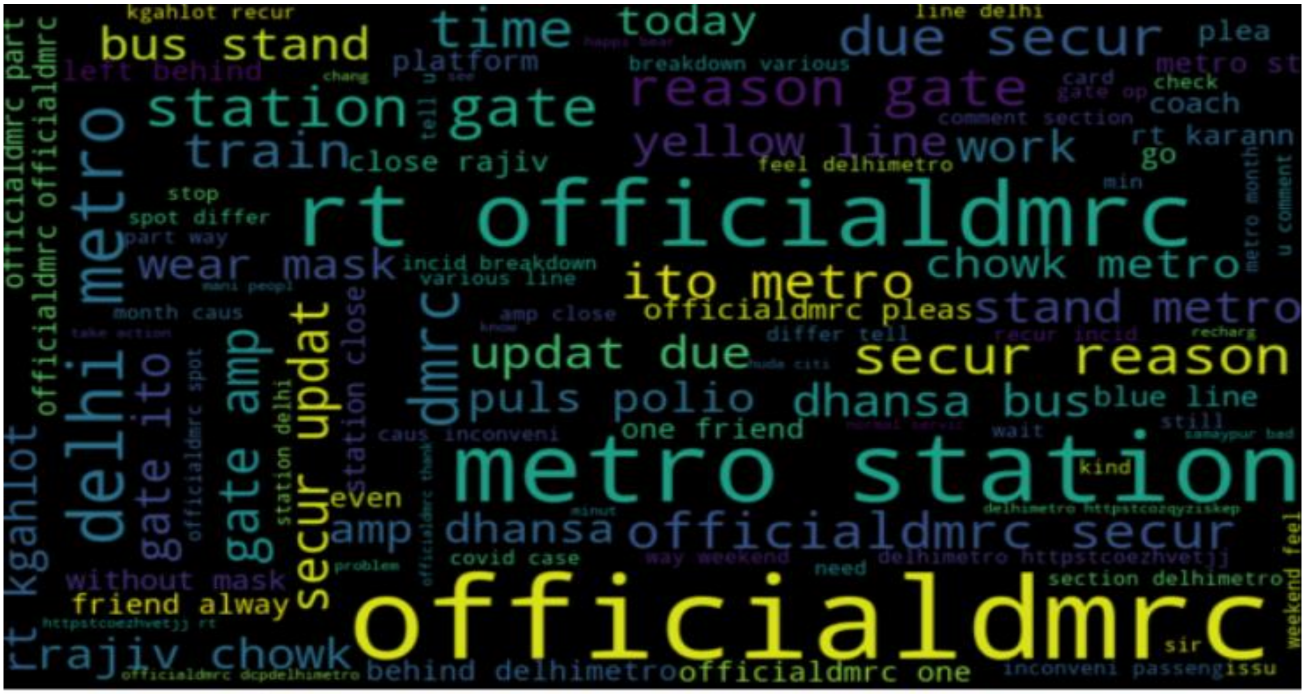
Converted to csv


data_negative
original
0 @My_Kind_World @OfficialDMRC @DCP_DelhiMetro @dmgbnagar @myogioffice @myogiadityanath @PMOIndia @narendramodi @TheDogMother_ @journalist_jps @dan710ths @Vibhachugh1 @DarPallavi @joedelhi @NarvijayYadav @pooj
3 RT @My_Kind_World: Bird spikes installed in #NoidaCityCenterMetroStation
4 RT @OfficialDMRC: Wear a mask at all times to prevent the spread of the virus as much as possible.
9 RT @OfficialDMRC: @Himansh42462529 Hi. Please provide smart card number and transaction details for checking further.
10 @OfficialDMRC Please Clean this Asap!! https://t.co/dwdWYXp2ul
14 RT @OfficialDMRC: Wear a mask at all times to prevent the spread of the virus as much as possible.
15 @OfficialDMRC My Smart Card Number is 46171359
18 @OfficialDMRC I have recharge my smart card through Amazon pay Rs 2000 as on dated 29/30 May 2022 but credited Rs 1600, what I do..
19 RT @OfficialDMRC: Wear a mask at all times to prevent the spread of the virus as much as possible.
25 RT @OfficialDMRC: Wear a mask at all times to prevent the spread of the virus as much as possible.
30 RT @OfficialDMRC: Wear a mask at all times to prevent the spread of the virus as much as possible.
31 RT @OfficialDMRC: Wear a mask at all times to prevent the spread of the virus as much as possible.
33 RT @OfficialDMRC: Wear a mask at all times to prevent the spread of the virus as much as possible.
34 @OfficialDMRC Coach No M3A08 some kind of sound is continuously coming while metro is moving please look immediately
35 @ChitraAhanthem @OfficialDMRC Yes That Happens Frequently!
37 RT @OfficialDMRC: Wear a mask at all times to prevent the spread of the virus as much as possible.
43 RT @OfficialDMRC: Wear a mask at all times to prevent the spread of the virus as much as possible.
47 @OfficialDMRC long lines daily in the hot sunny morning at dwarka metro station this is only due to only one x ray machine need atleast one more machine to go smoothly. https://t.co/ObXmm7HNib
51 @RoshanRaj08 @aarnavg17 @OfficialDMRC @AmirSai43048804 @OfficialDMRC DMRC has started operating trains from Vishvidalaya to Qutab Minar and Samaypur Badli to Huda City. Qutab Minar metro runs empty but
52 @OfficialDMRC
68 RT @Nm3530: @CMODelhi @OfficialDMRC @cpsavesoil @ArvindKejriwal @delhi Good initiative, by display in Delhi metro will spread the awarenesâ€¦
70 @CMODelhi @OfficialDMRC @cpsavesoil @ArvindKejriwal @delhi Good initiative, by display in Delhi metro will spread the awareness for good reach with more impact . #SaveSoil #SaveSoilMovement #pollution #Gurgaon https://t.co/0Ge0NKM
85 RT @My_Kind_World: Bird spikes installed in #NoidaCityCenterMetroStation
88 @OfficialDMRC alot of metro staff dnt wear mask/keep it down but 2day they knew that v have a conspiracy of fine going on everyone's mask was up according to dis u shud make an announcement 4 the public too that 2day we will fine U.Also
90 @OfficialDMRC
94 Why does this blue line suck?/.... Why is it hot inside metro... why isn't the ac working...damn!ðŸ˜©
108 @OfficialDMRC Coach no M7D033

Graphs

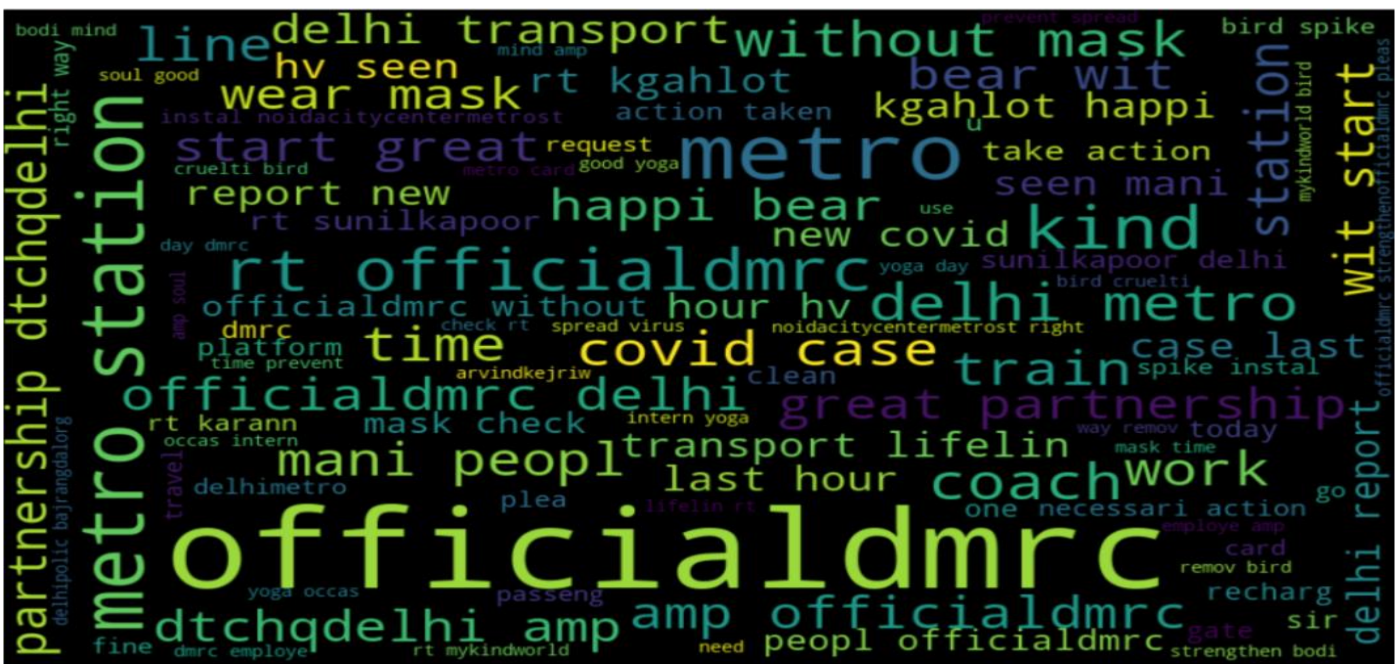
```
Command Prompt - python  v5_final_project_with_input.py
626) @delhi_search @MCD_Delhi @DCP_DelhiMetro @delhimetroapp @addl_

************************************************************
************************************************************
***      1 .....Show Tweets                                ****
***      2 .....Show graphs                                ****
***      3 .....Model creation                             ****
***      4 .....Topic Modelling                            ****
***      5 .....Tweets sentiments using model created      ****
***      6 ..... Exit                                      ****
************************************************************
************************************************************
Enter a choice       :::: 2
 __________________________________________________
|                                                  |
|     1 .  print wordcloud******                   |
|     2 .  print positive wordlcoud******          |
|     3 .  print  negative  wordcloud******        |
|     4 .  print  Subjectivity and polarity******  |
|     5 .  print  show graph of no tweets ******   |
|     6 .  print  no of tweets in a word ******    |
|     7 .  print  average word length              |
|     8 .  print common words                      |
|     9 .  Exit                                    |
|__________________________________________________|
-------------------------------------------------------
enter a choice of tweets
```

Wordcloud



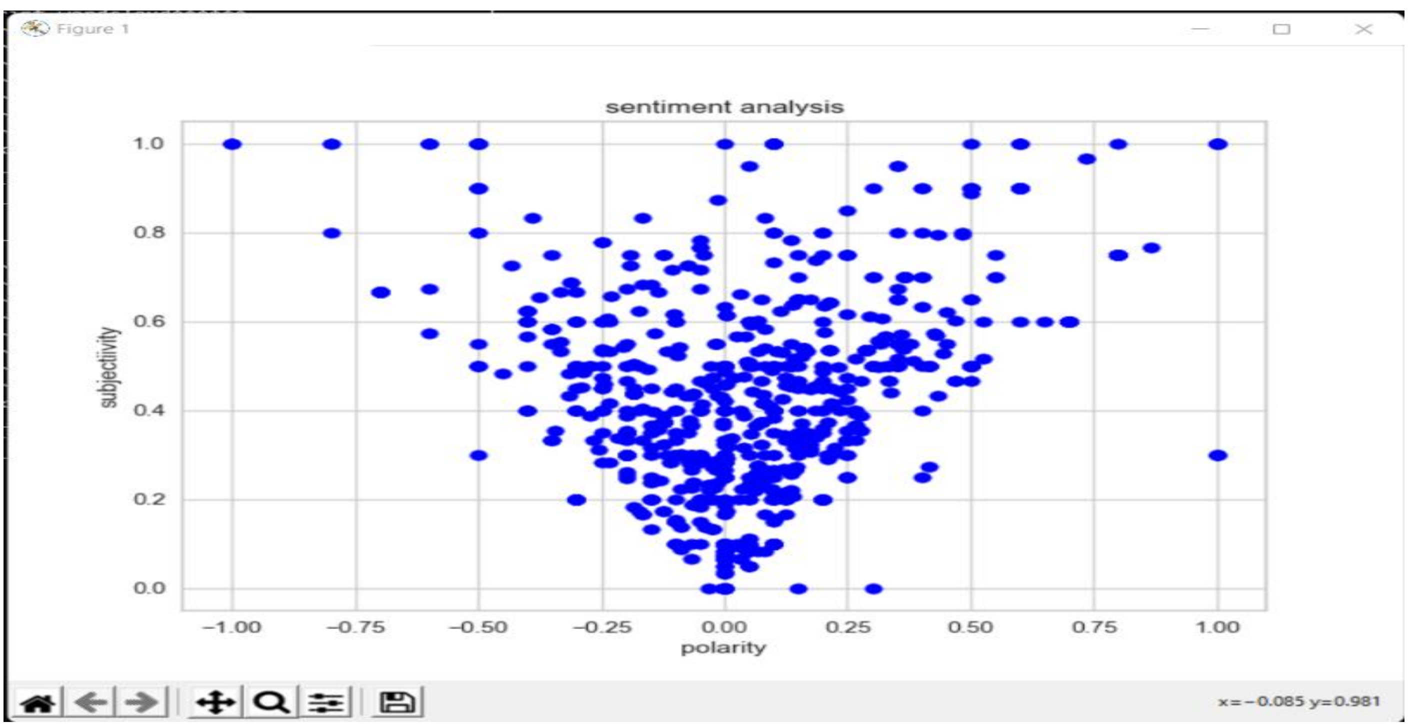
Positive Word Cloud

```
--------------------------------------------------------------------
enter a choice of tweets 1
 __________________________________________________________________
|                                                                  |
|     1 .  print wordcloud******                                   |
|     2 .  print positive wordlcoud******                          |
|     3 .  print  negative  wordcloud******                        |
|     4 .  print  Subjectivity and polarity******                  |
|     5 .  print  show graph of no tweets ******                   |
|     6 .  print  no of tweets in a word ******                    |
|     7 .  print  average word length                              |
|     8 .  print common words                                      |
|     9 .  Exit                                                    |
|__________________________________________________________________|
--------------------------------------------------------------------
enter a choice of tweets 2
28.0
```



Negative Word Cloud

Subjectivity and polarity Graph



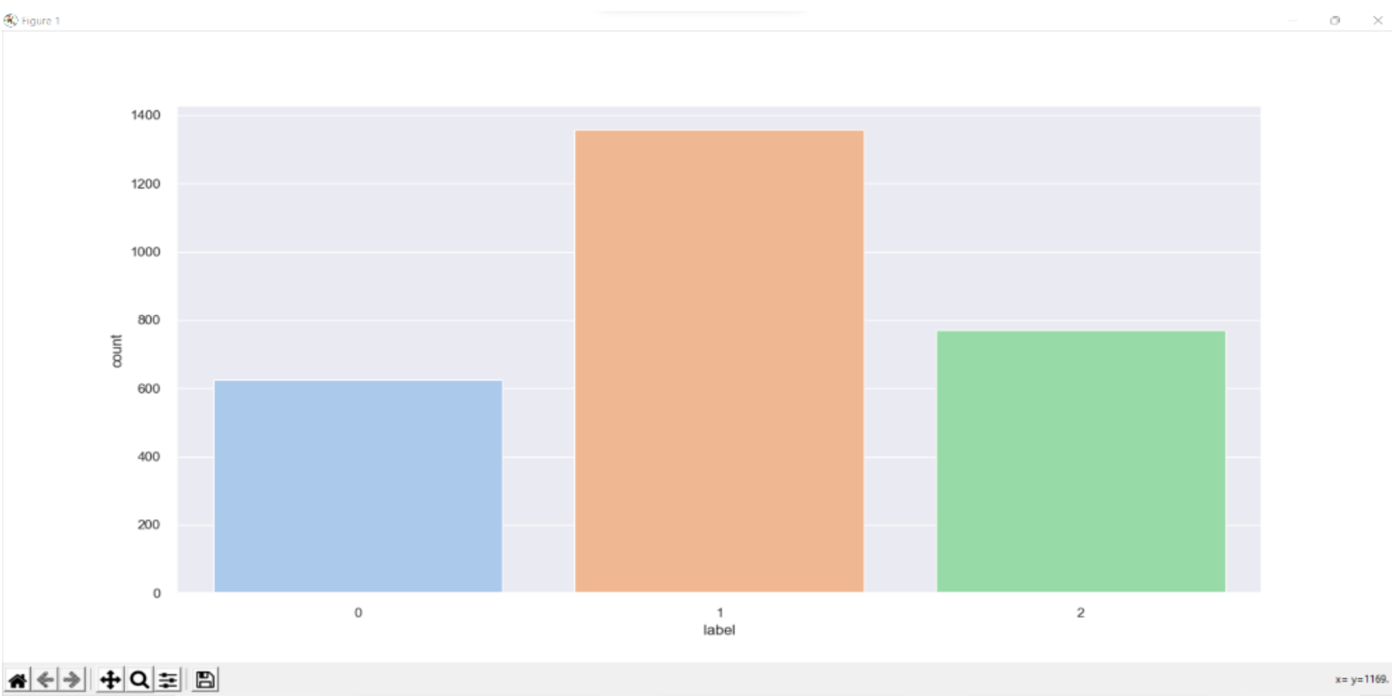
```
Command Prompt - python v5_final_project.py
--------------------------------------------------------
enter a choice of tweets 4
 _____________________________________________________
|                                                     |
|     1 .  print wordcloud******                      |
|     2 .  print positive wordlcoud******             |
|     3 .  print  negative  wordcloud******           |
|     4 .  print  Subjectivity and polarity******     |
|     5 .  print  show graph of no tweets ******      |
|     6 .  print  no of tweets in a word ******       |
|     7 .  print  average word length                 |
|     8 .  print common words                         |
|     9 .  Exit                                       |
|_____________________________________________________|
--------------------------------------------------------
enter a choice of tweets 5
   label  text
1      1  1358
2      2   770
0      0   626
```

Show no of tweets that are positive, negative, neutral

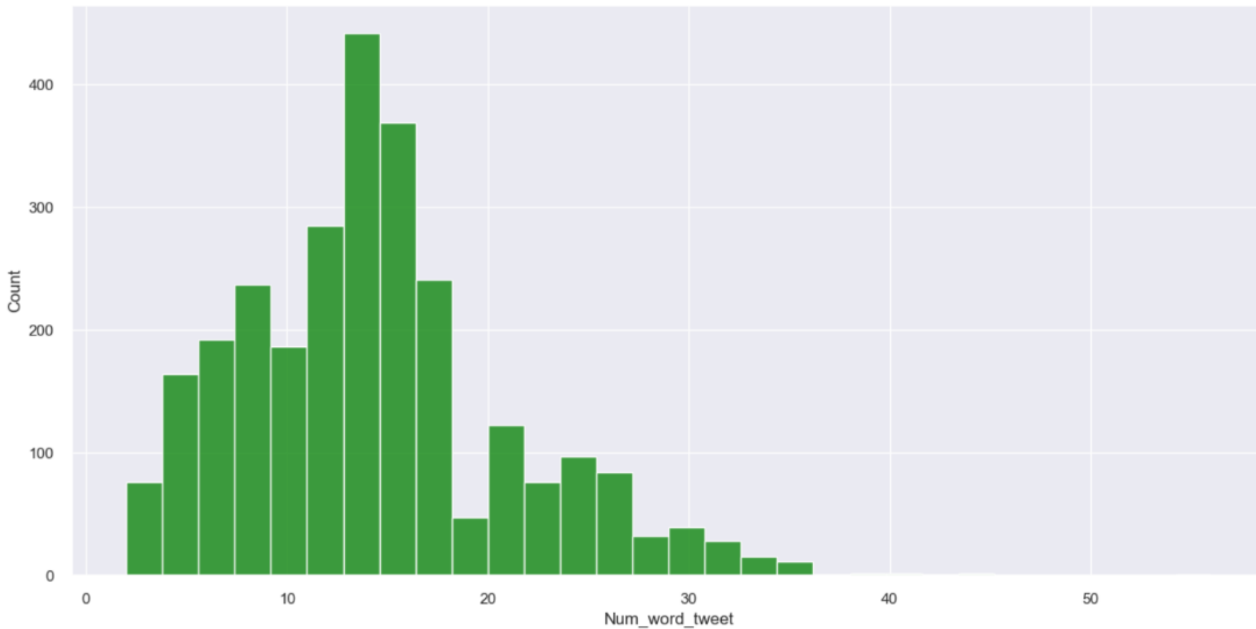
0: negative

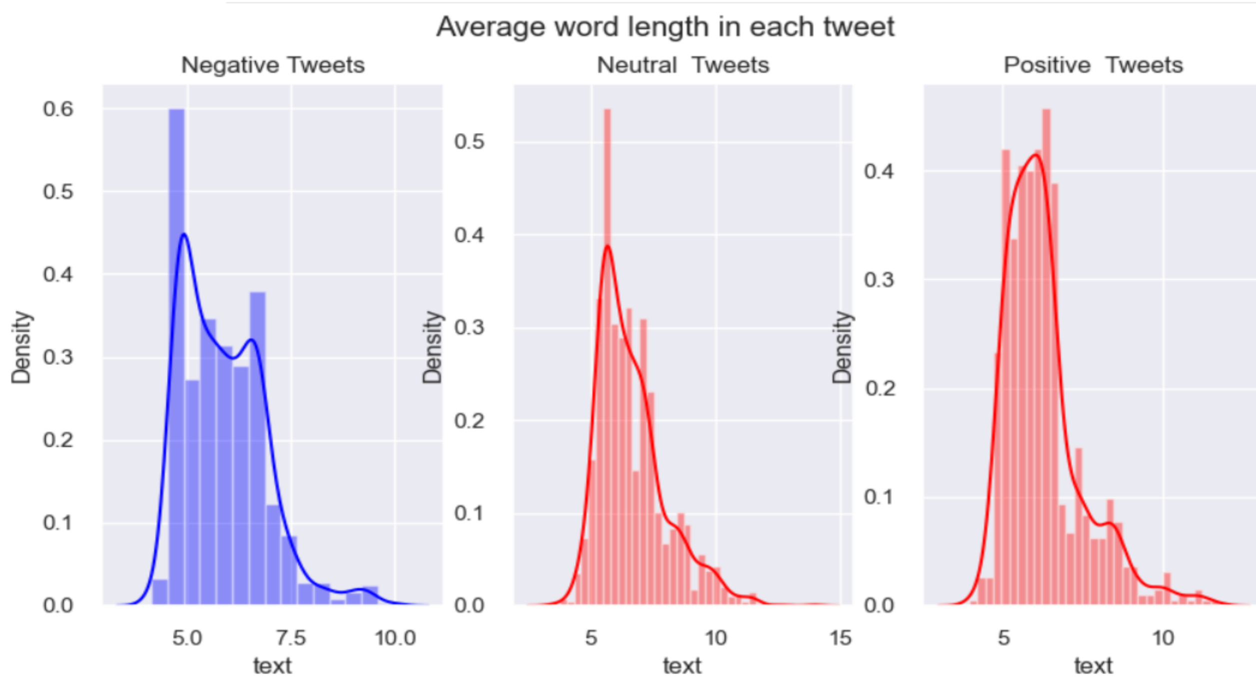
1: neutral

2: positive



Count of Number of words in each tweet



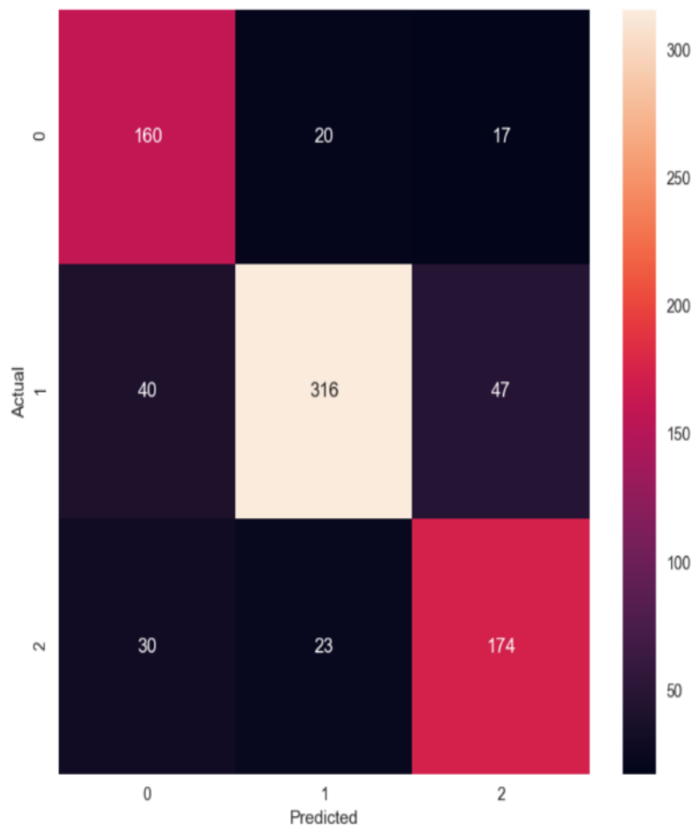


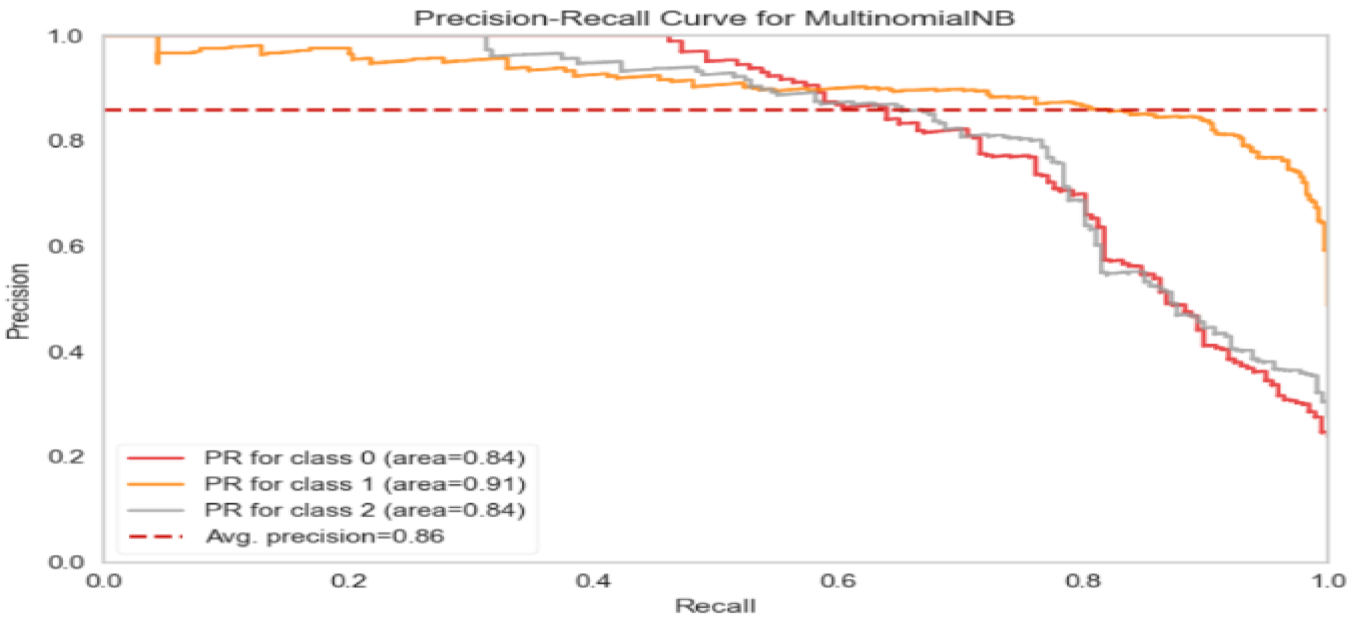
Model Creation

```
**********************************************************
**********************************************************
***     1 .....Show Tweets                           ****
***     2 .....Show graphs                           ****
***     3 .....Model creation                        ****
***     4 .....Topic Modelling                       ****
***     5 .....Tweets sentiments using model created ****
***     6 ..... Exit                                 ****
**********************************************************
**********************************************************
Enter a choice      :::: 3
 _________________________________________________
|    1.  Multonmial Naive bayes                   |
|    2.  SVC                                      |
|    3.  KNN                                      |
|    4.  RAndom forest                            |
|    5.  Aadaboost                                |
|    6.  show accuracy of all                     |
|    7.  Boosting of Random forest                |
|    8.  Boosting of Multinomial naive bayes      |
|    9.  Exit                                     |
|_________________________________________________|
enter a choice
```

Mulnomial Naives bayes

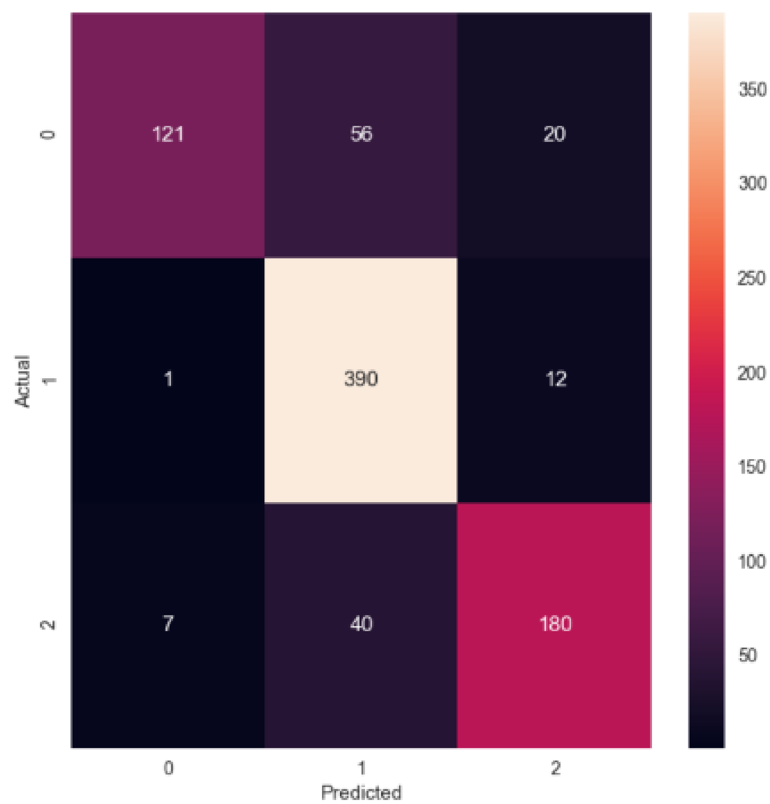
```
Enter a choice      :::: 3
 __________________________________________________
|     1.  Multonmial Naive bayes                  |
|     2.  SVC                                     |
|     3.  KNN                                     |
|     4.  RAndom forest                           |
|     5.  Aadaboost                               |
|     6.  show accuracy of all                    |
|     7.  Boosting of Random forest               |
|     8.  Boosting of Multinomial naive bayes     |
|     9.  Exit                                    |
|_________________________________________________|
enter a choice 1
____________Applying algorithms____________
X Test
              precision    recall  f1-score   support

           0       0.70      0.81      0.75       197
           1       0.88      0.78      0.83       403
           2       0.73      0.77      0.75       227

    accuracy                           0.79       827
   macro avg       0.77      0.79      0.78       827
weighted avg       0.80      0.79      0.79       827

X Train
              precision    recall  f1-score   support

           0       0.83      0.95      0.89       429
           1       0.97      0.89      0.93       955
           2       0.89      0.92      0.91       543

    accuracy                           0.91      1927
   macro avg       0.90      0.92      0.91      1927
weighted avg       0.92      0.91      0.92      1927
```

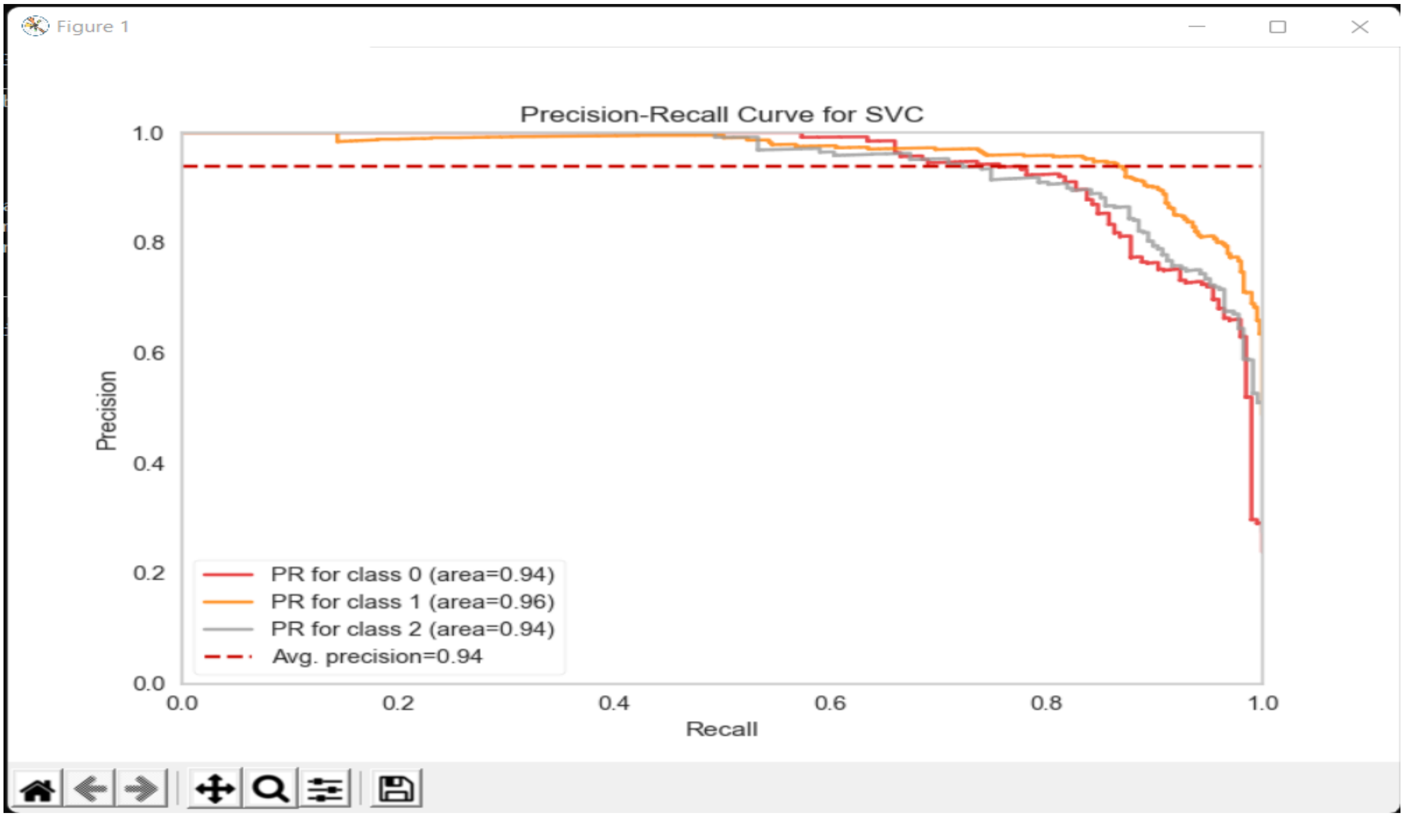




SVC

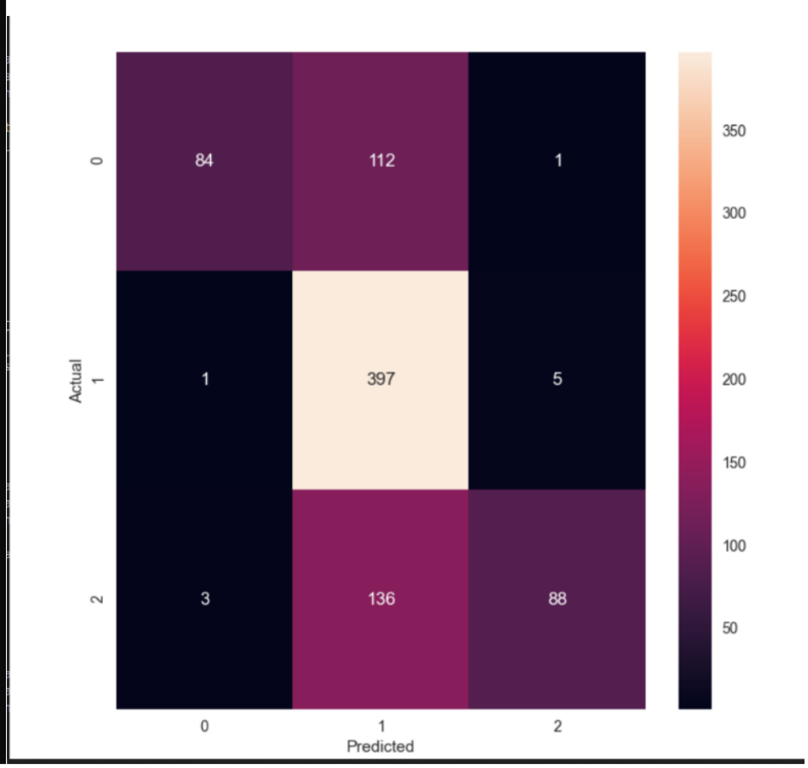
```
 ________________________________________________
|     1.  Multonmial Naive bayes                 |
|     2.  SVC                                    |
|     3.  KNN                                    |
|     4.  RAndom forest                          |
|     5.  Aadaboost                              |
|     6.  show accuracy of all                   |
|     7.  Boosting of Random forest              |
|     8.  Boosting of Multinomial naive bayes    |
|     9.  Exit                                   |
|________________________________________________|
enter a choice 2
____________Applying algorithms____________
X Test
              precision    recall  f1-score   support

           0       0.94      0.61      0.74       197
           1       0.80      0.97      0.88       403
           2       0.85      0.79      0.82       227

    accuracy                           0.84       827
   macro avg       0.86      0.79      0.81       827
weighted avg       0.85      0.84      0.83       827

X Train
              precision    recall  f1-score   support

           0       1.00      0.93      0.96       429
           1       0.94      1.00      0.97       955
           2       1.00      0.95      0.97       543

    accuracy                           0.97      1927
   macro avg       0.98      0.96      0.97      1927
weighted avg       0.97      0.97      0.97      1927
```

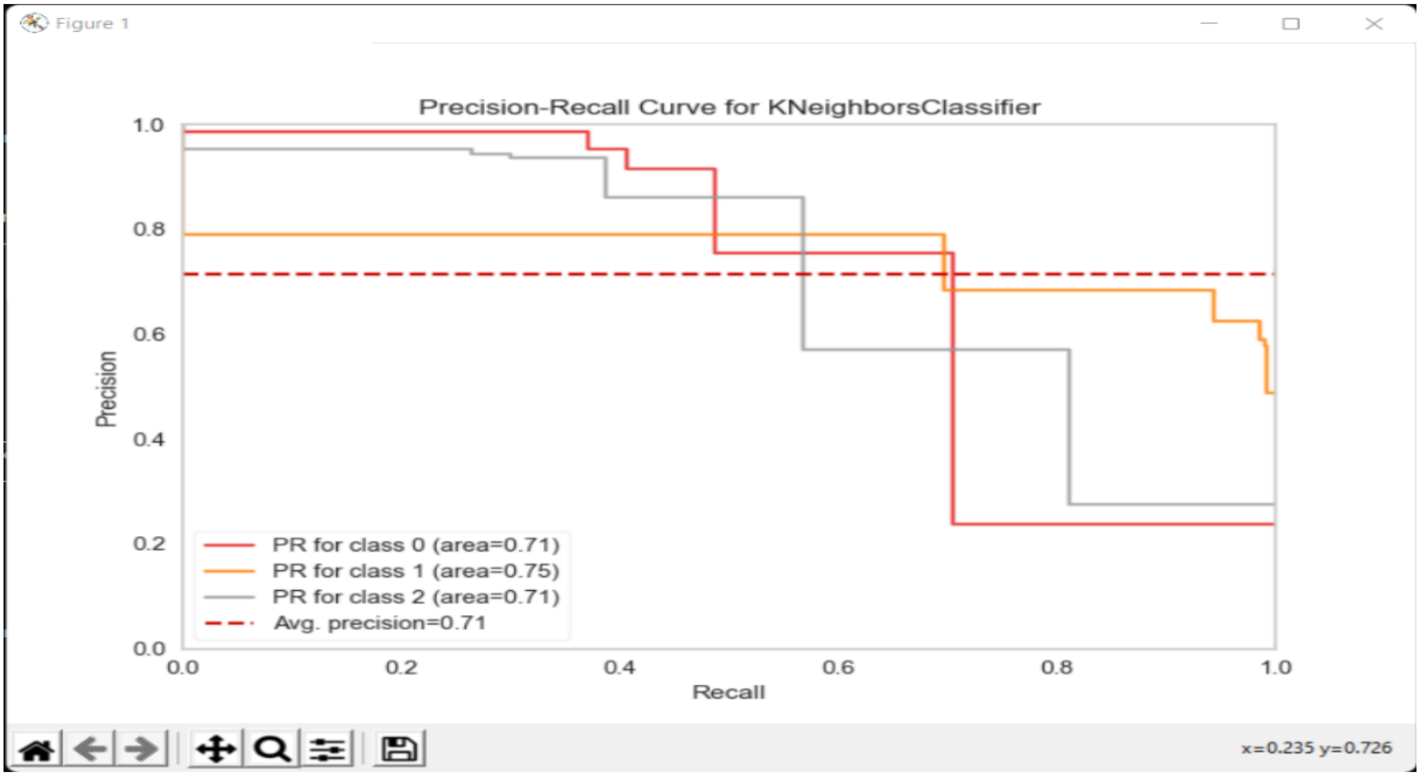




KNN

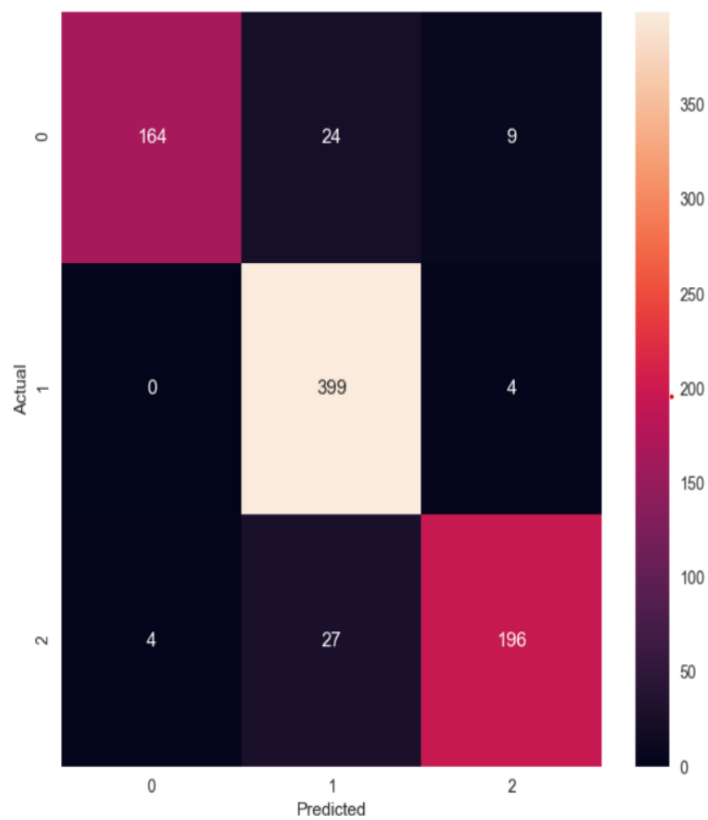
```
{MultinomialNB(): 0.7859733978234583, SVC(): 0.8355501813
 _____________________________________________________________
|     1.  Multonmial Naive bayes                              |
|     2.  SVC                                                 |
|     3.  KNN                                                 |
|     4.  RAndom forest                                       |
|     5.  Aadaboost                                           |
|     6.  show accuracy of all                                |
|     7.  Boosting of Random forest                           |
|     8.  Boosting of Multinomial naive bayes                 |
|     9.  Exit                                                |
|_____________________________________________________________|
enter a choice 3
____________Applying algorithms____________
X Test
              precision    recall  f1-score   support

           0       0.95      0.43      0.59       197
           1       0.62      0.99      0.76       403
           2       0.94      0.39      0.55       227

    accuracy                           0.69       827
   macro avg       0.84      0.60      0.63       827
weighted avg       0.78      0.69      0.66       827

X Train
              precision    recall  f1-score   support

           0       0.97      0.53      0.69       429
           1       0.66      0.98      0.79       955
           2       0.93      0.47      0.62       543

    accuracy                           0.74      1927
   macro avg       0.85      0.66      0.70      1927
weighted avg       0.80      0.74      0.72      1927
```

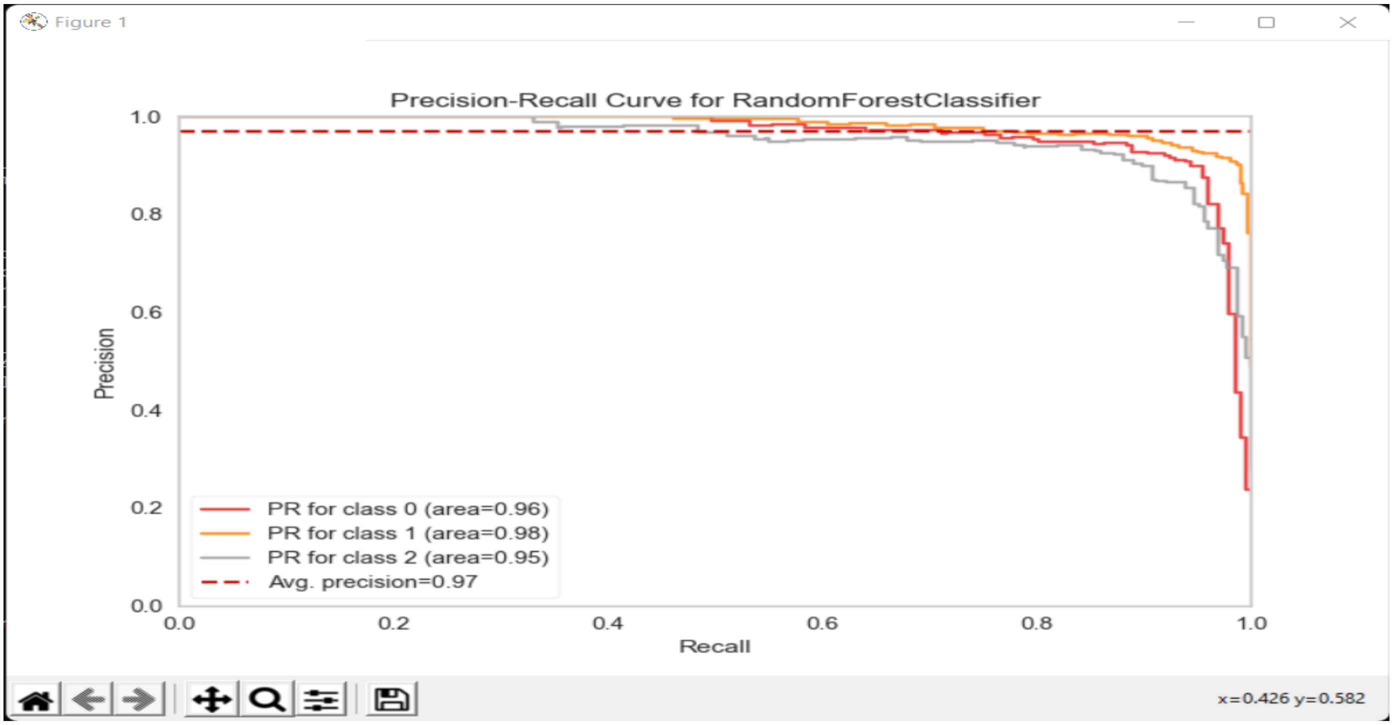




Random Forest Classification

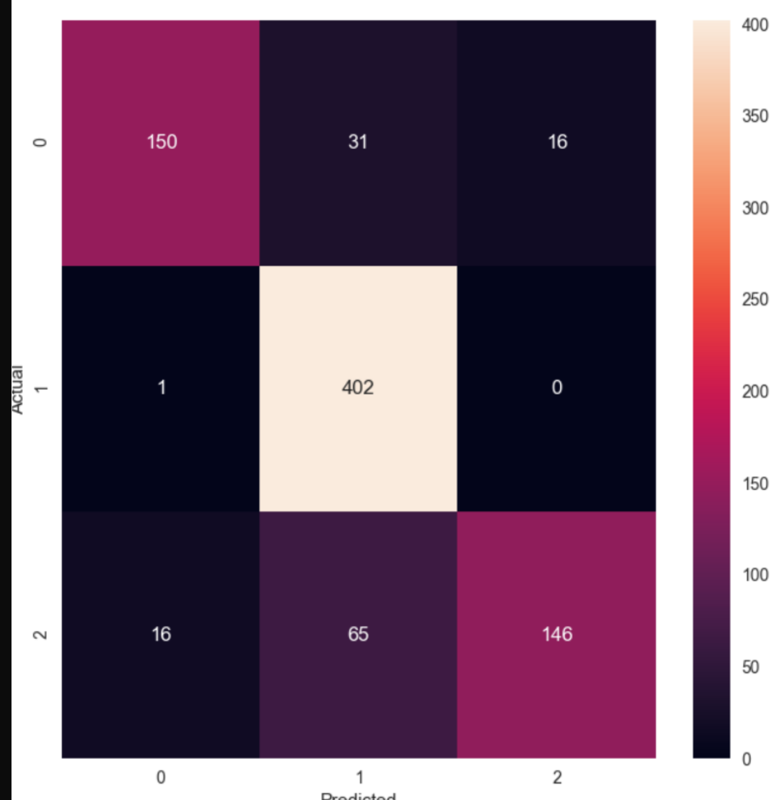
```
 ________________________________________________
|    1.  Multonmial Naive bayes                  |
|    2.  SVC                                     |
|    3.  KNN                                     |
|    4.  RAndom forest                           |
|    5.  Aadaboost                               |
|    6.  show accuracy of all                    |
|    7.  Boosting of Random forest               |
|    8.  Boosting of Multinomial naive bayes     |
|    9.  Exit                                    |
|________________________________________________|
enter a choice 4
___________Applying algorithms____________
X Test
              precision    recall  f1-score   support

           0       0.98      0.83      0.90       197
           1       0.89      0.99      0.94       403
           2       0.94      0.86      0.90       227

    accuracy                           0.92       827
   macro avg       0.93      0.90      0.91       827
weighted avg       0.92      0.92      0.92       827

X Train
              precision    recall  f1-score   support

           0       1.00      1.00      1.00       429
           1       1.00      1.00      1.00       955
           2       1.00      1.00      1.00       543

    accuracy                           1.00      1927
   macro avg       1.00      1.00      1.00      1927
weighted avg       1.00      1.00      1.00      1927
```

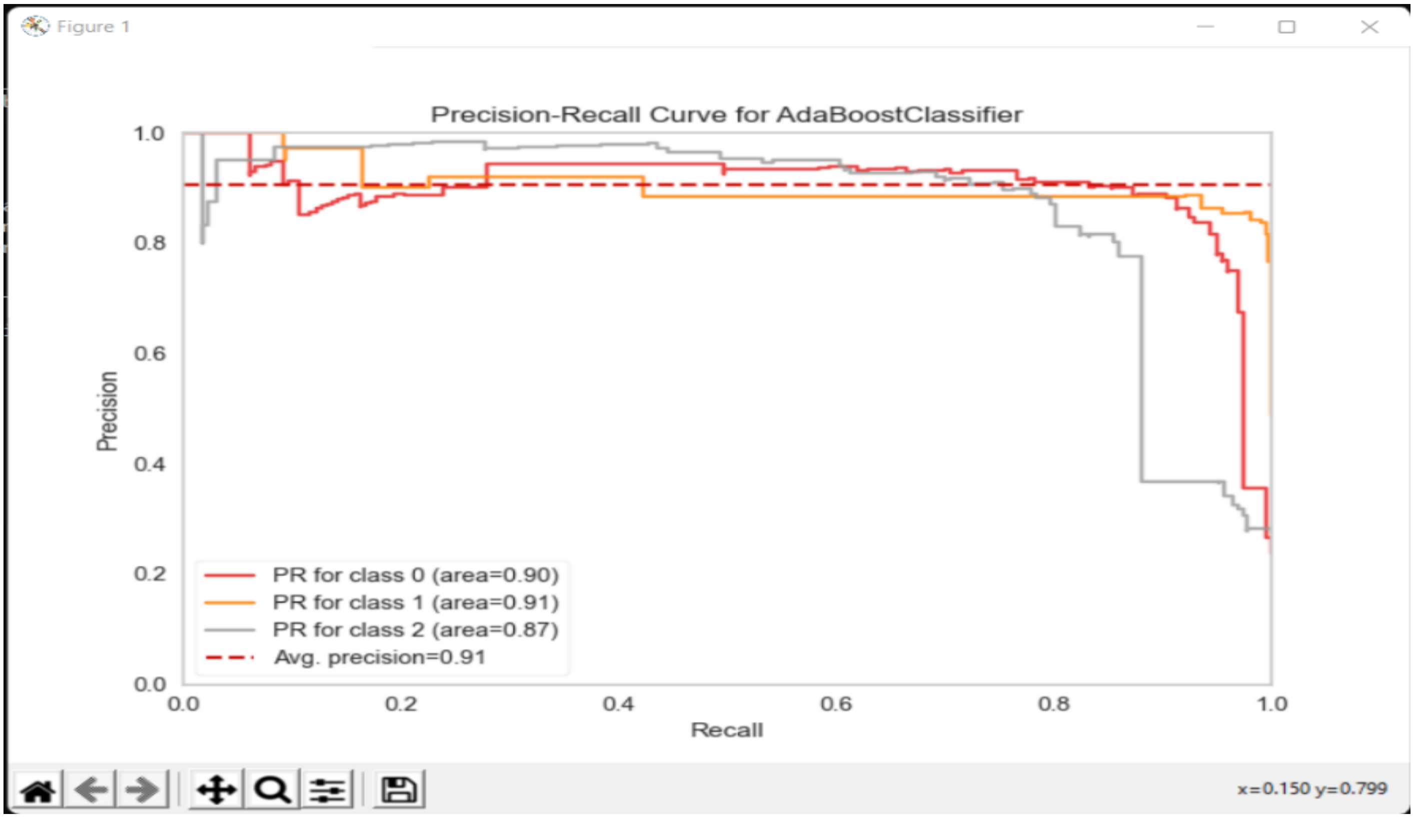




Adaboost Classifier

```
 ______________________________________________________
|    1.  Multonmial Naive bayes                        |
|    2.  SVC                                           |
|    3.  KNN                                           |
|    4.  RAndom forest                                 |
|    5.  Aadaboost                                     |
|    6.  show accuracy of all                          |
|    7.  Boosting of Random forest                     |
|    8.  Boosting of Multinomial naive bayes           |
|    9.  Exit                                          |
|______________________________________________________|
enter a choice 5
____________Applying algorithms____________
X Test
              precision    recall  f1-score   support

           0       0.90      0.76      0.82       197
           1       0.81      1.00      0.89       403
           2       0.90      0.64      0.75       227

    accuracy                           0.84       827
   macro avg       0.87      0.80      0.82       827
weighted avg       0.85      0.84      0.84       827

X Train
              precision    recall  f1-score   support

           0       0.94      0.73      0.82       429
           1       0.81      0.99      0.89       955
           2       0.90      0.70      0.79       543

    accuracy                           0.85      1927
   macro avg       0.88      0.81      0.84      1927
weighted avg       0.87      0.85      0.85      1927
```





Accuracy of all Models

```
 ______________________________________________
|     1.  Multonmial Naive bayes               |
|     2.  SVC                                  |
|     3.  KNN                                  |
|     4.  RAndom forest                        |
|     5.  Aadaboost                            |
|     6.  show accuracy of all                 |
|     7.  Boosting of Random forest            |
|     8.  Boosting of Multinomial naive bayes  |
|     9.  Exit                                 |
|______________________________________________|
enter a choice 6
___________Applying algorithms_____________
{MultinomialNB(): 0.7859733978234583, SVC(): 0.8355501813784765, KNeighborsClassifier(): 0.6880290205562273, RandomForestClassifier(): 0.9177750906892382, AdaBoostClassifier(): 0.8440145102781137}
```

Topic Classification

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1		Unnamed:	user_name	user_locat	user_descı	user_verifi	date	text	original	hashtags	source	Subjectivit	Polarity	analysis	label	Topic mod	Topic modeling lDA Tuning
2	0	0	Niladri Hal	Siliguri, Ind	Consultant	FALSE	########	mykindwo	@My_Kind_World @(		Twitter foı	0.75	0.25	positive	2	0.064*""me	0.036*""metro" + 0.034*""station" + 0.031*""gate" + 0.028*""close" + 0.026*""car
3	1	1	Sameer Ga	Bengaluru,	keep Life s	FALSE	########	concern he	@concern13161846 (		Twitter foı	0.25	0	neutral	1	0.071*""me	0.019*""aarnavg" + 0.019*""line" + 0.016*""metro" + 0.016*""video" + 0.016*""ma
4	2	2	Rupsha Mukherjee		Adoption	FALSE	########	officialdm	@OfficialDMRC Than		Twitter foı	0.833333	-0.38889	negative	0	0.077*""htt	0.059*""gate" + 0.053*""secur" + 0.031*""reason" + 0.030*""metro" + 0.029*""upc
5	3	3	Ayush			FALSE	########	rt mykindw	RT	['NoidaCit	Twitter foı	0.535714	0.285714	positive	2	0.077*""htt	0.059*""gate" + 0.053*""secur" + 0.031*""reason" + 0.030*""metro" + 0.029*""upc
6	4	4	Siraj Ahamad			FALSE	########	rt officiald	RT	['StoptheS	Twitter foı	0.2	0.2	positive	2	0.077*""htt	0.059*""gate" + 0.053*""secur" + 0.031*""reason" + 0.030*""metro" + 0.029*""upc
7	5	5	Robert		I am difere	FALSE	########	mykindwo	@My_Kind_World @(		Twitter We	0	0	neutral	1	0.064*""me	0.036*""metro" + 0.034*""station" + 0.031*""gate" + 0.028*""close" + 0.026*""car
8	6	6	Rajeev Mit	Navi Mum	à¤¬à¤¦à¤	FALSE	########	rt kgahlot	RT @kgahlot: There h		Twitter foı	0.5	0	neutral	1	0.106*""me	0.031*""servic" + 0.026*""line" + 0.020*""dmrc" + 0.019*""delay" + 0.017*""yellow
9	7	7	SinghBhhupinder		Realistic	FALSE	########	priyaku of	@PriyaKu73754749 @		Twitter foı	0.533333	-0.33333	negative	0	0.081*""me	0.034*""http" + 0.034*""delhimetro" + 0.026*""leav" + 0.026*""zhvetjj" + 0.025*""
10	8	8	Syed			FALSE	########	mykindwo	@My_Kind_World @(		Twitter foı	0	0	neutral	1	0.064*""me	0.036*""metro" + 0.034*""station" + 0.031*""gate" + 0.028*""close" + 0.026*""car
11	9	9	Himanshu Ranjan			FALSE	########	rt officiald	RT @OfficialDMRC: @		Twitter We	0.642857	0.214286	positive	2	0.077*""htt	0.059*""gate" + 0.053*""secur" + 0.031*""reason" + 0.030*""metro" + 0.029*""upc
12	10	10	Sipplin		I've Nothir	FALSE	########	officialdm	@OfficialDMRC Plea:		Twitter foı	0.7	0.366667	positive	2	0.043*""htt	0.025*""miss" + 0.023*""work" + 0.023*""metro" + 0.020*""station" + 0.019*""gat
13	11	11	sanjeev rawat		Man on mi	FALSE	########	mykindwo	@My_Kind_World @(		Twitter foı	0	0	neutral	1	0.064*""me	0.036*""metro" + 0.034*""station" + 0.031*""gate" + 0.028*""close" + 0.026*""car
14	12	12	Narendra.j	New Delhi, Delhi		FALSE	########	officialdm	@OfficialDMRC Pleas		Twitter foı	0.1	0	neutral	1	0.081*""me	0.034*""http" + 0.034*""delhimetro" + 0.026*""leav" + 0.026*""zhvetjj" + 0.025*""
15	13	13	Himanshu Ranjan			FALSE	########	officialdm	@OfficialDMRC Pleas		Twitter We	0	0	neutral	1	0.077*""htt	0.059*""gate" + 0.053*""secur" + 0.031*""reason" + 0.030*""metro" + 0.029*""upc
16	14	14	Swati Singh			FALSE	########	rt officiald	RT	['StoptheS	Twitter foı	0.2	0.2	positive	2	0.077*""htt	0.059*""gate" + 0.053*""secur" + 0.031*""reason" + 0.030*""metro" + 0.029*""upc
17	15	15	Himanshu Ranjan			FALSE	########	officialdm	@Official		Twitter We	0.642857	0.214286	positive	2	0.077*""htt	0.059*""gate" + 0.053*""secur" + 0.031*""reason" + 0.030*""metro" + 0.029*""upc
18	16	16	Sagar			FALSE	########	rt joni kgal	RT @Jony49251436: (		Twitter foı	0	0	neutral	1	0.106*""me	0.031*""servic" + 0.026*""line" + 0.020*""dmrc" + 0.019*""delay" + 0.017*""yellow
19	17	17	SHIV PARAKASH DWIVED			FALSE	########	ltgovdelhi	@LtGovD		Twitter foı	1	-0.5	negative	0	0.106*""me	0.031*""servic" + 0.026*""line" + 0.020*""dmrc" + 0.019*""delay" + 0.017*""yellow
20	18	18	Himanshu Ranjan			FALSE	########	officialdm	@OfficialDMRC I hav		Twitter We	0.642857	0.214286	positive	2	0.077*""htt	0.059*""gate" + 0.053*""secur" + 0.031*""reason" + 0.030*""metro" + 0.029*""upc
21	19	19	Devender I	New Delhi	Kya likhu a	FALSE	########	rt officiald	RT	['StoptheS	Twitter foı	0.2	0.2	positive	2	0.077*""htt	0.059*""gate" + 0.053*""secur" + 0.031*""reason" + 0.030*""metro" + 0.029*""upc
22	20	20	Shilpa Gup	Jammu, Ja	Former Th	FALSE	########	mykindwo	@My_Kind_World @(		Twitter foı	0.1	-0.1	negative	0	0.064*""me	0.036*""metro" + 0.034*""station" + 0.031*""gate" + 0.028*""close" + 0.026*""car
23	21	21	Dinesh Bhatt			FALSE	########	officialdm	@OfficialDMRC Why		Twitter foı	0	0	neutral	1	0.071*""me	0.019*""aarnavg" + 0.019*""line" + 0.016*""metro" + 0.016*""video" + 0.016*""ma
24	22	22	Vipin Baro	New Delhi	Senior Bus	FALSE	########	chitraahan	@ChitraAhanthem @		Twitter foı	0.1	0	neutral	1	0.106*""me	0.031*""servic" + 0.026*""line" + 0.020*""dmrc" + 0.019*""delay" + 0.017*""yellow
25	23	23	ROHIT		hello	FALSE	########	kgahlot of	@kgahlot @OfficialD		Twitter foı	0	0	neutral	1	0.064*""me	0.036*""metro" + 0.034*""station" + 0.031*""gate" + 0.028*""close" + 0.026*""car
26	24	24	Ranbir Gul	Badli, Jhajj	Ranbir Gul	FALSE	########	rt kgahlot	RT @kgahlot: There h		Twitter We	0.5	0	neutral	1	0.106*""me	0.031*""servic" + 0.026*""line" + 0.020*""dmrc" + 0.019*""delay" + 0.017*""yellow
27	25	25	Mukul Ana	Delhi	Nation Firs	FALSE	########	rt officiald	RT	['StoptheS	Twitter foı	0.2	0.2	positive	2	0.077*""htt	0.059*""gate" + 0.053*""secur" + 0.031*""reason" + 0.030*""metro" + 0.029*""upc
28	26	26	manish kuı	india	I AM LIVIN	FALSE	########	pleas start	Please		Twitter foı	0	0	neutral	1	0.106*""me	0.031*""servic" + 0.026*""line" + 0.020*""dmrc" + 0.019*""delay" + 0.017*""yellow

topic_modelling

Ready Accessibility: Unavailable

Sentence as an input and classifying on the basis of model created

```
Command Prompt - python  v5_final_project.py
***      1 .....Show Tweets                                   ****
***      2 .....Show graphs                                   ****
***      3 .....Model creation                                ****
***      4 .....Topic Modelling                               ****
***      5 .....Tweets sentiments using model created         ****
***      6 .....  Exit                                        ****
**********************************************************
**********************************************************
Enter a choice       :::: 5
WARNING: !! Train the model first
RandomForestClassifier()
the accuracy is
0.9129383313180169
*****************************************
 Enter a stringtrain is  late
*****************************************
*******************OUTPUT**********
[1]
**********************************************************
**********************************************************
***      1 .....Show Tweets                                   ****
***      2 .....Show graphs                                   ****
***      3 .....Model creation                                ****
***      4 .....Topic Modelling                               ****
***      5 .....Tweets sentiments using model created         ****
***      6 .....  Exit                                        ****
**********************************************************
**********************************************************
Enter a choice       :::: 5
WARNING: !! Train the model first
RandomForestClassifier()
the accuracy is
0.9044740024183797
*****************************************
 Enter a stringtrain is not working
*****************************************
*******************OUTPUT**********
[1]
**********************************************************
**********************************************************
***      1 .....Show Tweets                                   ****
***      2 .....Show graphs                                   ****
***      3 .....Model creation                                ****
***      4 .....Topic Modelling                               ****
***      5 .....Tweets sentiments using model created         ****
***      6 .....  Exit                                        ****
**********************************************************
**********************************************************
Enter a choice       ::::
```

4. Glossary

ML: Machine learning

NLP: Natural Language Processing

API : Application programming interface

KNN: k-nearest neighbors

SVM: Support Vector Machine

Adaboost: Adaptive Boosting

SVC: Linear Support Vector Classifier

5. References

• https://towardsdatascience.com/https-medium-com-vishalmordehumanizing-customer-complaints-using-nlp-algorithms-64a820cef373

• https://rpubs.com/nr2670/week09_tasection

• https://www.kaggle.com/code/danielbeltsazar/twitter-sentiment-speechanalysis-and-modelling

• V. Prakruthi, D. Sindhu and D. S. Anupama Kumar, "Real Time Sentiment

Analysis Of Twitter Posts," 2018 3rd International Conference on

Computational Systems and Information Technology for Sustainable

Solutions (CSITSS), 2018, pp. 29-34, doi: 10.1109/CSITSS.2018.8768774.

➢ https://ieeexplore.ieee.org/document/8768774

• L. Mandloi and R. Patel, "Twitter Sentiments Analysis Using Machine

Learninig Methods," 2020 International Conference for Emerging

Technology (INCET), 2020, pp. 1-5, doi:

10.1109/INCET49848.2020.9154183.

➢ https://ieeexplore.ieee.org/document/9154183

• https://www.pluralsight.com/guides/building-a-twitter-sentiment-analysisin-python

• https://www.geeksforgeeks.org/twitter-sentiment-analysis-using-python/

• https://www.analyticsvidhya.com/blog/2021/06/twitter-sentimentanalysis-a-nlp-use-case-for-beginners/

•https://www.analyticsvidhya.com/blog/2021/06/twitter-sentimentanalysis-a-nlp-use-case-forbeginners/#:~:text=In%20the%20problem%20statement%20we,Logistic%2

0Regression