

# **SOCIAL MEDIA ANALYTICS**

## **Using Sentiment Analysis**

*Project Report*

*Submitted in partial fulfilment of the requirement for the degree of Bachelor of  
Technology in Computer Science and Engineering*



*By*

**SUCHETAN PATHANIA (161217)**

*Under the supervision of*

**DR JAGPREET SIDHU**

*To*

**Department of Computer Science & Engineering and Information  
Technology**

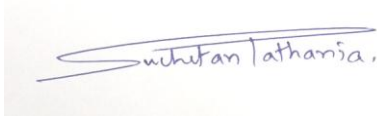
**Jaypee University of Information Technology Waknaghat, Solan-  
173234, Himachal Pradesh**

## Certificate

### Candidate's Declaration

I hereby declare that the work presented in this report entitled “**Social Media Analytics Using Sentiment Analysis**” in partial fulfilment of the requirements for the award of the degree of **Bachelor of Technology in Computer Science and Engineering** submitted in the department of Computer Science & Engineering and Information Technology, Jaypee University of Information Technology Wakanaghat is an authentic record of my own work carried out over a period from July 2019 to May 2020 under the supervision of **Dr Jagpreet Sidhu** (Assistant Professor, Senior Grade, Computer Science & Engineering Department).

The matter embodied in the report has not been submitted for the award of any other degree or diploma.



(Student Signature)

Suchetan Pathania (161217)

This is to certify that above statement made by the candidate is true to the best of my knowledge.



(Supervisor Signature)

Name: Dr Jagpreet Sidhu

Designation: Assistant Professor (Senior Grade)

Department: Computer Science and Engineering

Dated:

## **ACKNOWLEDGEMENT**

I want to take this chance to thank almighty for blessing me with his grace and taking my job to a successful culmination. I owe our profound gratitude to my project supervisor, Dr. Jagpreet Sidhu who took keen interest and guided me all along in my project work titled - “Social Media Analytics Using Sentiment Analysis” till the completion of our project by providing all the necessary information for developing the project. The project development helped me in research and I got to know a lot of new things in our domain. I am really thankful to him.

# TABLE OF CONTENT

<b>Chapter</b>	<b>Page No</b>
Title Page	(i)
Certificate	(ii)
Acknowledgement	(iii)
Table of Content	(iv)
List of Figures	(vi)
List of Tables	(viii)
Abstract	(x)

# TABLE OF CONTENT

<b>Chapter</b>	<b>Page No</b>
<b>1) Introduction</b>	
1.1) Introduction	1
1.2) Problem Statement	2
1.3) Objectives	2
1.4) Methodology	2
1.5) Organisation	3
<b>2) Literature Survey</b>	
2.1) Papers taken for consideration	4
<b>3) System Development</b>	
3.1) Main Classifier	10
3.2) System Requirement	11
3.3) Model Development	11
3.4) Process Flow	12
3.5) Libraries Used	13
<b>4) Performance Analysis</b>	
4.1) Test Plan	16
4.2) Classification	16
<b>5) Conclusion</b>	
5.1) Conclusion	40
5.2) Future Scope	40
5.3) Applications	40

## LIST OF FIGURES

<b>Fig No</b>	<b>Title</b>	<b>Page No</b>
2.1	World Cloud	8
3.1	Model Development for the system	11
3.2	Process Flow	12
4.1	Input for #COVID-19	16
4.2	Tweets collected	17
4.3	Sentiment chart	17
4.4	Input for #togetherwerise	18
4.5	Tweets collected	18
4.6	Sentiment chart	19
4.7	Input for #HelpTheNeedy	19
4.8	Tweets collected	20
4.9	Sentiment chart	20
4.10	Input for Electric Vehicles	21
4.11	Tweets collected	21
4.12	Sentiment chart	22
4.13	Input for Bitcoin	22
4.14	Tweets collected	23
4.15	Sentiment chart	23
4.16	Variation in tweets for one week of May, 19	25
4.17	Variation in tweets	26
4.18	Variation in tweets	27

<b>Fig No</b>	<b>Title</b>	<b>Page No</b>
4.19	Analysis	28
4.20	Variation in tweets	29
4.21	Variation in tweets	31
4.22	Variation in tweets	32
4.23	Variation in tweets	33
4.24	Variation in tweets	35
4.25	Variation in tweets	36
4.26	Variation in tweets	37
4.27	Variation in tweets	38

## LIST OF TABLES

<b>Table No</b>	<b>Title</b>	<b>Page No</b>
4.1	Analysis of tweets for seven days in May, 19	25
4.2	Mathematical Analysis	25
4.3	Variations of tweets for one week	26
4.4	Mathematical Analysis	26
4.5	Variations of tweets for one week	27
4.6	Mathematical Analysis	27
4.7	Variations of tweets for one week	28
4.8	Mathematical Analysis	28
4.9	Variations of tweets for one week	29
4.10	Mathematical Analysis	29
4.11	Comparison of average support for politicians	30
4.12	Variations of tweets for one week	31
4.13	Mathematical Analysis	31
4.14	Variations in tweets for a week	32
4.15	Mathematical Analysis	32
4.16	Variations of tweets for one week	33
4.17	Mathematical Analysis	33
4.18	Comparison of support for various technologies	34
4.19	Variations of tweets for one week	35
4.20	Mathematical Analysis	35



<b>Table No</b>	<b>Title</b>	<b>Page No</b>
4.21	Variations of tweets for one week	36
4.22	Mathematical Analysis	36
4.23	Variations of tweets for one week	37
4.24	Mathematical Analysis	37
4.25	Variations of tweets for one week	38
4.26	Mathematical Analysis	38
4.27	Comparison of sentiments for issues related to pandemic	39

## **ABSTRACT**

Twitter is a micro-blogging platform where users express their views via tweets, with a growing user base. With our developer account we have integrated the Twitter API into our project so we could access the data from Twitter. In this project, we performed Sentiment Analysis on the tweets fetched and divided the opinion of the users into positive, neutral or negative sentiments. This was achieved by first collecting and cleaning the data, and then classifying our data with the use of Naïve Bayes classifier. Further, we studied the data, analysed the importance of predicting outcomes based on the public sentiments in various real life situations and recorded our observations. For future we propose to improve the accuracy of our model and making better domain related analysis.

# CHAPTER 1: INTRODUCTION

## 1.1 Introduction:

Social networking sites have become very popular. Millions of users conveniently express their views and opinions on a wide array of topics. These users share a lot of information about trending topics and current issues. Thus, these platforms generate a large amount of data. Because of a free format of messages and an easy access to micro blogging platforms there is a shift from traditional blogs to these services.

Twitter is one such micro-blogging service which has a growing user base. The users express their views on the topics of interest through messages called tweets. These opinions on products, events or other users are useful in real world applications. By performing sentiment analysis on their response we can get an idea of how users perceive these issues from their point of view. Sentiment analysis involves “Natural Language Processing (NLP)” and data mining. In recent tasks, sentiments like “strongly positive”, “weakly positive”, “strongly negative” and “weakly negative” are also being considered.

Sentiment analysis is a research area in text mining. Researchers have been trying to find many ways for analysing tweets based on their understanding of the concept. They are using various machine learning algorithms to classify tweets into positive and negative classes based on the sentiments, such as Naïve Bayes Classifier, Support Vector Machine, etc.

The main goal is to connect on Twitter and search for the tweets that contain a particular keyword and then evaluate the polarity of the tweet.

Sentiment analysis on the tweets collected is useful for knowing the public sentiments. But, due to misspellings and slangs in the tweets; Twitter Sentiment Analysis is different from general sentiment analysis. However, by identifying the specific domains and performing sentiment analysis on them we can find the effect of domain information in classification of the sentiments. This is quite useful for researchers who are trying to research a topic or issue and businesses who are trying to market a new product or understanding the opinion on existing product or service.

## **1.2 Problem Statement:**

Sentiment analysis is introduced to tell us that the information is correct or wrong in each scenario using the social media tags. As we know, great amount of data is available to us from the platforms like Twitter. Extracting and analysing this data can be useful for many purposes like:

- 1.2.1 Politics:** In politics the political parties can use this sentiment analysis to plan their campaigns, know the popular public opinion or we can even use it to predict the election outcomes.
- 1.2.2 Business:** Businesses can benefit from this by developing their marketing strategies according to customer response towards liking of a particular product or business decision.
- 1.2.3 Public Actions:** Sentiment analysis can be used to monitor social developments or determining the general public sentiments after any event.

## **1.3 Objectives:**

To automate our analysis based on the reviews or comments in the social media by the people on an unknown tweet stream. Social media provides great opportunities for businesses to connect to consumers easily.

## **1.4 Methodology:**

- 1) The data is collected from Twitter using the Twitter API using the keywords.
- 2) The extracted raw data is pre-processed using Natural Language Toolkit techniques
- 3) The sentiments of the tweets are evaluated based on feature selection of score words.
- 4) Naïve Bayes classifier is used for training and testing the features and also evaluating the sentimental polarity.
- 5) The proposed system is implemented using Python
- 6) Data visualisation is performed using the Matplotlib library to show the plot

## **1.5 Organisation:**

The project work is divided into five main chapters:

### Chapter 1: Introduction

This chapter describes the growth of social-media platforms and its important impact on our lives. This also introduces us the concept of social media analytics via Twitter Sentiment Analysis.

### Chapter 2: Literature Review

This chapter discusses the literature used from platforms such as ACM, IEEE, Google Scholar, Springer and other repositories. These papers were useful to understand various methods used in existing sentiment analysis.

### Chapter 3: System Development

The chapter focuses on the hardware and software requirements. Moreover the chapter further details the methodology used.

### Chapter 4: Performance Analysis

This chapter highlights performance of our model by extensively covering the analysis in various domains and documents our results.

### Chapter 5: Conclusion

This chapter summaries the work and reviews the diversity of domains and various applications of our analysis.

## CHAPTER 2: LITERATURE SURVEY

We considered the following literature-

**i) Opinion Mining and Sentiment Analysis[1]:**

We start by classification of a paragraph to subjective and objective sentences. We apply text categorization techniques to only the subjective portion in the data and also, removing the objective part. This subjective part is extracted following it, we create separate group of words into classes and apply the formula to individual and associated sentiment class and calculate the score for each word. Then we chose the class with optimal partitioning cost to be the base class.

**ii) Opinion Analysis based on Lexical Clue sets: [2]**

Twitter sentiment analysis is problematic due to more contextual information. Therefore, the domain is essential. First, we extract sentiments into different domains. We identify a sentiment clue as a primary subject in expression like event or product. Then put this in a clue-set which comprises most-likely feature-words for that word. The algorithm then identifies the sentiment topics from training data. Finally, the clues that are associated to a sentiment topic are arranged to its clue-set. The classifier gets updated for identifying other domains from the dataset.

**iii) Sentiment-Based Meta-level Features for SA: [3]**

This paper addresses the problem of automatically learning to classify the sentiment of messages by the data from meta-level features. A new word is checked for similarities with pre-classified dataset. Then the most similar content match is taken to be its true sentiment word. It is derived using K-nearest neighbour from the sentiment distribution. The document polarity of the neighbours is obtained by unsupervised lexical methods.

**iv) SA on Political Campaigns with the help of Facebook: [4]**

This paper studies the sentiment of users for campaigns run by political parties. It analyzes the data gathered from Facebook posts.

$$\textit{Sentiment Index} = \ln \left[ \frac{1 + \textit{TOTAL}^{\textit{POSITIVE}}}{1 + \textit{TOTAL}^{\textit{NEGATIVE}}} \right]$$

The paper reveals the perception of voters and the sentiments that are developed during a course of time. It also describes how popular opinion on social media in favor may also result in loss in some cases.

**v) Evaluating SA Using Advanced linguistic-features: [5]**

It addresses how common method of “Bag-of-Words” is not efficient as it does not accounts how the meaning of a word changes with context. It introduces us to a new schema by analyzing the target-of-opinions and modifies the coverage of the word with contextual-domains called “SentiML”. It uses advanced linguistic-features and allows more coverage than a sentiment-dictionary with faster linkage of the wordset.

**vi) Evaluation of Features on Sentimental-Analysis:[11]**

First step is the pre-processing of data, which involves the initial conversion to lowercase of all the textual data. Then the words are converted into their original form. This text is filtered by removing the stop words. Further, this processed data is partitioned into training and testing data. Next follows the process of removal of suffixes called Stemming. Further statistical methods are used for feature selection.

**vii) SA using Lexicon Approach :[6]**

This paper discusses two sentiment classification techniques – Machine learning approach and Lexicon based approach. Lexicon based approach is further classified into Dictionary based approach and Corpus based approach. Lexicon approach involves calculating orientation for a document from semantic inclination of words or phrases. Dictionary based approach collects seeds (words) with known orientation: +ve or –ve manually. The algorithm then searches online available dictionary for synonyms or antonyms and expands the set.

This approach then matches for pattern to calculate sentiment orientation.

**viii) SA using Text Mining and Hybrid Classification Approach on Tweets: [7]**

This begins with concept of “Opinion Mining”, which further proposes to use the info and predict the sentiments of public to an issue. This describes to first save the tweets to a database. The data is then pre-processed for filtering the special characters, short words, etc. The suffixes are removed by Stemming and then converted into tokens. A TDM is created which is passed for classification by classifiers. The paper also analyses for a hybrid of KNN and Naïve Baye’s classifiers.

**ix) SA of Flipkart reviews using Naïve Bayes and Decision Tree algorithm:[8]**

Naïve Bayes approach for classifying is based on probabilistic classifying where Bayes Theorem is used and theory of total probability is used. The paper claims it’s suitability for large datasets. Just like text mining it starts with tokenization then removal of stop words followed by text transformation. Then features are selected for the parts of document that contribute for positive and negative words. These parts are joined in such a way such that the probability is maximum of the resulting sentence existing in either terms of +ve or –ve.



**x) SA using Neuro-Fuzzy and HMM(s) of Textual data:[9]**

This work describes how NeuroFuzzy Models applied to dialog analysis could determine user intent with reasonable accuracy. Hidden Markov models (HMM) achieved comparable accuracy, but with a different pattern of errors. The data from different sources is combined into a corpus. The combined text is put into a collection (array) of words. Then this array of words is sorted. Stemming is not done so as to keep words different and preserve their context.

**xi) Predicting Ratings emphasized on Social Sentiments From Textual Reviews:[10]**

This paper categorizes into types of lists. One is the word list which puts the words into positive and negative polarity. The second list is Topics list which puts topics as a root sentiment into a list. This is done to avoid the info overloading problem.

Firstly it uses a social sentiment measurement approach and calculates the sentiments for each user on a particular product. Secondly, we also consider the inter-personal sentiment influence on a user.

Finally characterize the user preferences by fusing the sentiment-similarity, interpersonal sentiment influence and the item reputation.

**xii) Twitter Part-Of-Speech Tagging Using Pre-classification HMM:[12]**

This talks of Hidden Markov models (HMM), which are highly used in Natural Language Processing (NLP). HMM considers data present in as unobserved states. It uses the Naïve Bayes model to classify the dataset into objective or subjective data. In this paper, we propose a novel treatment of HMM model to use the result of sentimental analysis in syntactic level task, i.e. part-of-speech (POS) tagging. The subjectivity degree of the testing sentence is used as a combination factor to choose an appropriate value

**xiii) Sentiment Analysis from Binary to Multi-Class Classification :[13]**

Binary classification of data means to classify data into either positive or negative based on polarity. The paper proposes a pattern-based approach that gets deep into the classification of tweets. This approach has proven to reach greater accuracy in classifying the sentiments into positive or negative polarity. The score is calculated by the net score for positive wt and negative wt.

$$\rho(t) = \frac{PW(t) - NW(t)}{PW(t) + NW(t)}$$

**xiv) Bag-of-Words Model : [14]**

Bag-of-Words (BoW) is a classical model used in text mining and natural language processing (NLP). It is a vector-based representation used to effectively present documents with the use of numerical-vectors. This paper presents a new method called Fuzzy Bag-of-Words model to overcome the shortcomings of BoW like extreme sparsity, high dimensionality and capture high level semantics behind textual data.



**Fig 2.1:** Word Cloud

Fuzzy BoW adopts a fuzzy mapping based on semantic correlation among words quantified by similarity.

**xv) SA of Product Reviews Using ML Approaches: [15]**

The paper discusses the collection of useful information from the feedback customers provide on the e-com websites. It focuses on the sentiment analysis by the application of NLP. The customer reviews are useful to know the perception of a product or service and the important info that can be extracted to understand the potential of a product. It also talks about Machine Learning methods that can be applied on this data as it is unstructured and how it can be processed and classified.

**xvi) Survey on Sentiment Analysis using Twitter Dataset: [16]**

This paper discusses comparison of various techniques of sentiment analysis. It shows types of sentiment analyses and methods to extract the sentiment polarity from the Twitter dataset of tweets. It also presents the challenges and opportunities of different approaches and also highlights the time taken for execution.

## CHAPTER 3: SYSTEM DEVELOPMENT

### 3.1 Main Classifier

In Machine learning we use a classifier to differentiate between different entities based on some parameters.

#### 3.1.1 Naïve Bayesian Classifier

The Naive Bayesian classifier is a commonly used classifier in machine learning models that is based on Naïve Bayes probabilistic classification. It uses Bayes theorem to find the probability of occurring an event given that some other event has occurred. It classifies a feature given the classification of some other independent feature. It is quite fast in examining the whole input dataset. Further, presence of few bad data-points, or noise, do not change the results by much. The Naive Bayesian classification system can be calculated as

$$P(a|b) = P(a)*P(b|a)/P(a)$$

Here, we consider ‘a’ to be a specific-class with P(a) its probability and ‘b’ is taken to be the text we have to classify with P(b) its probability. P(b|a) gives us the probability of text (b) appearing in the given class a. Text b can be a phrase or sentence while the class may be positive or negative. Finally we have to choose such a value of a so as to maximize the P(a|b).

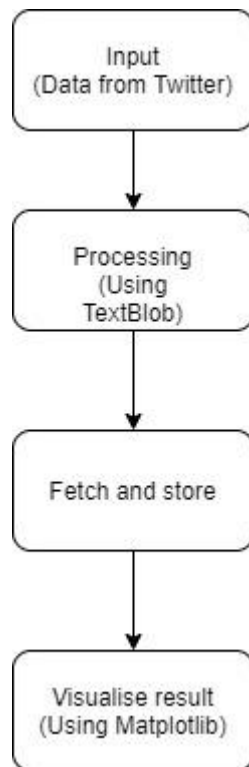
The sub-types of Naïve Bayes classifier-

1. Multinomial Naïve Bayes-  
Mainly used for document classification purposes. To check if it lies in sports, politics or technology category.
2. Bernoulli Naïve Bayes-  
It uses Boolean variables for predicting the class variables.
3. Gaussian Naïve Bayes-  
It uses continuous values for the predictors. The values are sampled using Gaussian distribution.

### 3.2 System Requirements

- i) Hardware used:
  - Data Bus: 64-bit
  - Processor(CPU): Intel Core-i5(7<sup>th</sup> Gen)
  - 8 GB RAM
- ii) Softwares used:
  - OS: Windows 10
  - Python programming language 3.7
- iii) Twitter Developer Account to access Twitter API

### 3.3 Model Development



**Fig 3.1:** Model Development for the system

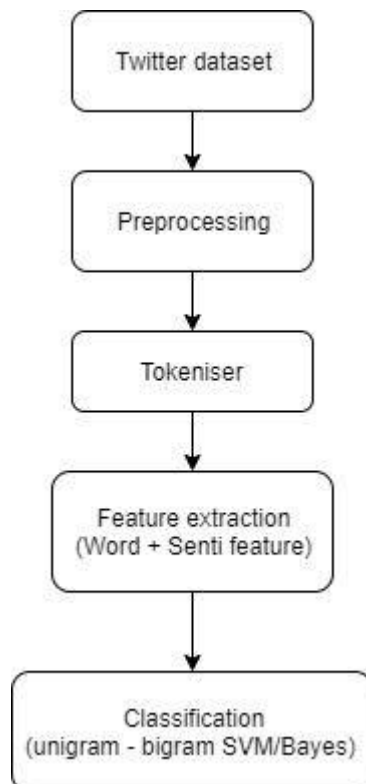
**3.3.1 Input of data:** The model takes the tweets as the input from a live Twitter stream or tweets made on the given date.

**3.3.2 Processing the data:** The Textblob library is an open source library. It includes functions like TextBlob particularly for natural language processing work, which is used to make either positive, neutral or negative characterization

**3.3.3 Fetching and Storing:** We fetch the data using the credentials from Twitter dev account and then store the tweets fetched to be able to filter them and execute our program. Then these filtered tweets are stored to a csv file.

**3.3.4 Visualization of results:** We use the Matplotlib library to plot the results of the executed program. The pie charts give us a visual representation of our analysis.

### 3.4 Process Flow



**Fig 3.2:** Process flow

### 3.5 Libraries Implemented:

1. Tweepy
2. Textblob
3. Matplotlib
4. Regular Expression (re)

**4.5.1 Tweepy:** Tweepy provides access to the Twitter RESTful API methods. Each method can accept various parameters and returns responses.[17]

- **Classes implemented:**

- i) API: Methods in this class are used to enable interaction with Twitter after authentication.
- ii) OAuthHandler: It takes a Customer key and a Secret key as parameters to request a token from Twitter.
- iii) Cursor: This class will pass parameters to methods when cursor request is made.

- **Methods implemented:**

- i) `auth.set_access_token`: it uses the token to redirect to Twitter. Else the user must manually supply it.
- ii) Cursor: This method is used to provide parameters like the term we need to search for, the language it should search, etc.
- iii) `items` : This method is used to supply the number of search items as parameter we need to search.

**4.5.2 Textblob:** It is a popular library to perform natural language processing operations on Twitter tweets-

- **Classes Implemented:**

- i) Sentiment: The textblob.sentiments module contains two sentiment analysis implementations.:[18]

- a. PatternAnalyzer – based on patterns

- b. NaiveBayesAnalyzer – based on NLTK

- **Methods Implemented:**

- i) Polarity: It returns the polarity score as a float within the range [-1, +1].

**4.5.3 Matplotlib:** Matplotlib is used to create comprehensive and interactive visualisations in Python.

- **Classes Utilized:**

- i) pyplot: It is a state-based interface to matplotlib, mainly used for interactive plots and programmatic plot generation.[19]

- **Methods Utilized:**

- i) pie: Plots a pie chart

- ii) legend: Places a legend on the axes.

- iii) title: Sets a title for the axis

- iv) axis: gets axis properties

- v) tight\_layout: adjusts subplot parameters to give padding



#### **4.5.4 Regular Expressions**

A regular expression (RE) specifies a set of strings that matches it. The functions in this module allow us to check if a particular string matches a given regular expression.[20]

##### **Methods Implemented:**

- i) `sub`: It returns the string obtained by replacing the leftmost non-overlapping occurrences of patterns in a given string by the given replacement.

## CHAPTER 4: PERFORMANCE ANALYSIS

### 4.1 Test Planning:

#### 4.1.1 Dataset for new keywords:

We input a new keyword and store the tweets obtained manually.

#### 4.1.2 Varying number of tweets :

The keyword/hashtag is kept to be same, we will analyse for different number of tweets.

#### 4.1.3 Fixing number of tweets :

We keep the number of tweets fixed but perform analysis on different keywords.

### 4.2 Classification:

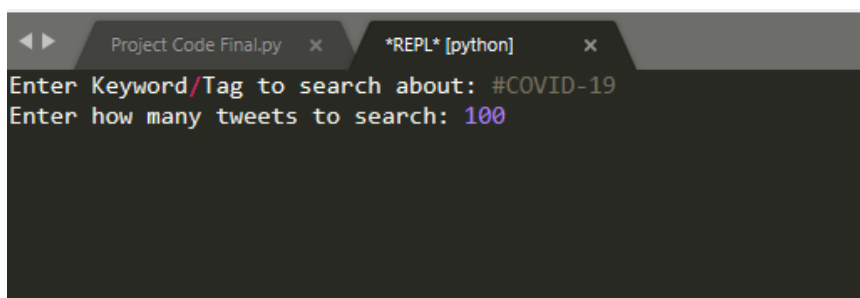
Results can be analysed as follows:

- Variable search-set
- Continuous search-set

#### 4.2.1 Variable search-set

As per our plan, we perform execution of our program, for various different combinations of keywords/ hashtags and the number of tweets. We analyse our results by plotting the output for different trending words. The pie chart depicts the distribution of the sentiments.

##### 4.2.1.1 Case 1



```
Project Code Final.py x *REPL* [python] x
Enter Keyword/Tag to search about: #COVID-19
Enter how many tweets to search: 100
```

**Fig 4.1:** Input for #COVID-19

```

Project Code Final.py x *REPL* [python] x
Let us appreciate all of our janasainiks for doing a tremendous role during COVID-19 PANDEMIC...
Tweet's Polarity: Sentiment(polarity=0.3333333333333333, subjectivity=1.0)
Tweet : RT @LloydMathias: Media consumption by generation over the #Covid_19 period nicely represented by @VisualCap .
GenZ: online video+streaming...
Tweet's Polarity: Sentiment(polarity=0.6, subjectivity=1.0)
Tweet : RT @COVIDNewsByMIB: #IndiaFightsCorona:

Secretary, @MoHFW_INDIA held a high level review meeting today with 11 #COVID19 high case load mun...
Tweet's Polarity: Sentiment(polarity=0.16, subjectivity=0.5399999999999999)
Tweet : 50% success chance expected from @UniofOxford's #COVID19 #vaccine trial https://t.co/7YvDXhDH4A
Tweet's Polarity: Sentiment(polarity=0.09999999999999999, subjectivity=0.2)
Tweet : RT @TheEconomist: China's ambassador to Australia recently warned the country it was treading a "dangerous" path https://t.co/
St0W1Cft0p
Tweet's Polarity: Sentiment(polarity=-0.3, subjectivity=0.575)
How people are reacting on #COVID-19 by analyzing 100 tweets.

General Report: 0.07906861471861469
Weakly Positive

Detailed Report:
9.00% people thought it was positive
39.00% people thought it was weakly positive
3.00% people thought it was strongly positive
7.00% people thought it was negative
7.00% people thought it was weakly negative
0.00% people thought it was strongly negative
35.00% people thought it was neutral

```

Fig 4.2: Tweets collected

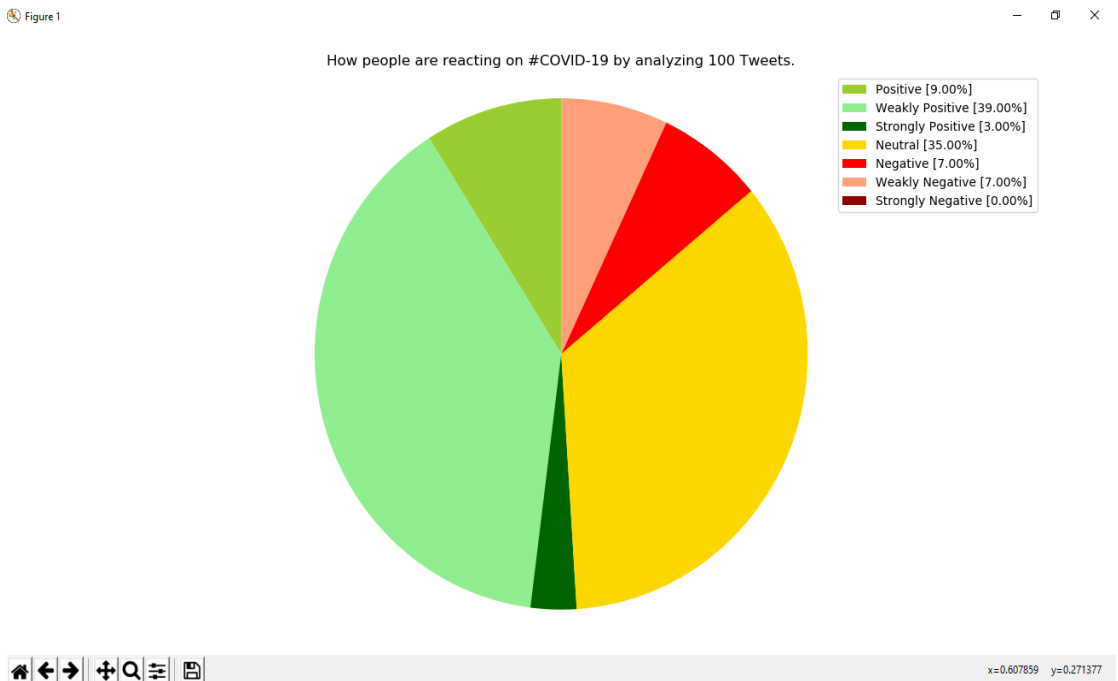


Fig 4.3: Sentiment chart

#### 4.2.1.2 Case 2

```
Project Code Final.py x *REPL* [python] x
Enter Keyword/Tag to search about: #togetherwerise
Enter how many tweets to search: 100
```

Fig 4.4 Input for #togetherwerise

```
Project Code Final.py x *REPL* [python] x
Tweet : RT @MahindraRise: At a time when the unavailability of food has become an issue for many, Kantibhai Amliyar, a farmer from Gujarat is makin...
Tweet's Polarity: Sentiment(polarity=0.5, subjectivity=0.5)
Tweet : @ElizabethHurley Individualism, a big #problem of our #society
https://t.co/p9CU59f4U #Societe #Sociedad_ https://t.co/4WMMUJrgiIg
Tweet's Polarity: Sentiment(polarity=0.0, subjectivity=0.1)
Tweet : RT @MahindraRise: At a time when the unavailability of food has become an issue for many, Kantibhai Amliyar, a farmer from Gujarat is makin...
Tweet's Polarity: Sentiment(polarity=0.5, subjectivity=0.5)
Tweet : RT @Dhara_MSD: An advise from the community leaders.
@dhruvashastril12 : When you don't know how to do a task, say yes to yourself and comp...
Tweet's Polarity: Sentiment(polarity=0.0, subjectivity=0.0)
Tweet : RT @Dhara_MSD: You can say what qualities perfect leader needs to have from @dhruvashastril12 's talk.
Being authentic, Deep listening, Cre...
Tweet's Polarity: Sentiment(polarity=0.5, subjectivity=0.7166666666666667)
Tweet : RT @Dhara_MSD: @dhruvashastril12 sharing her learning on making and being the part of one of the successfully community. Sharing the
importa...
Tweet's Polarity: Sentiment(polarity=0.75, subjectivity=0.95)
How people are reacting on #togetherwerise by analyzing 100 tweets.
General Report: 0.31690568181818174
Positive
Detailed Report:
57.00% people thought it was positive
8.00% people thought it was weakly positive
6.00% people thought it was strongly positive
1.00% people thought it was negative
3.00% people thought it was weakly negative
0.00% people thought it was strongly negative
25.00% people thought it was neutral
```

Fig 4.5 Tweets collected

How people are reacting on #togetherwiser by analyzing 100 Tweets.

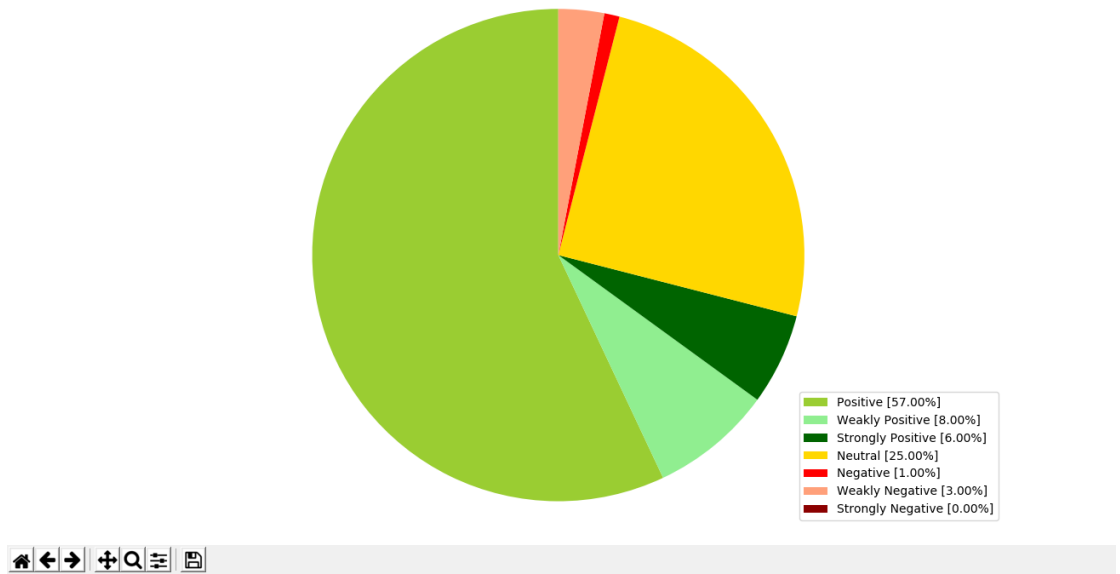


Fig 4.6 Sentiment chart

### 4.2.1.3 Case 3

```
Project Code Final.py × *REPL* [python] ×
Enter Keyword/Tag to search about: #HelpTheNeedy
Enter how many tweets to search: 100
```

Fig 4.7: Input for #HelpTheNeedy

```

Project Code Final.py x *REPL* [python] x
Tweet : RT @AditiAroraInsa7: @derasachasauda volunteers try to do best to #helpTheNeedy people by providing food and ration. https://t.co/
oRAgvsRDv0
Tweet's Polarity: Sentiment(polarity=1.0, subjectivity=0.3)
Tweet : RT @anjuinsan557: #HelpTheNeedy - The summers are here and temperature are soaring. To the help passersby, water dispensing
services can b...
Tweet's Polarity: Sentiment(polarity=0.0, subjectivity=0.0)
Tweet : RT @Sarbjee92727722: Whenever we get needy in this lock-down time, So we should #HelpTheNeedy with the permission of the
government, such a...
Tweet's Polarity: Sentiment(polarity=0.0, subjectivity=0.5)
Tweet : RT @SumitInsan98: Corona crisis has made life tough for many needy people,who face food shortage or are sick.Let us try to do our
best to #...
Tweet's Polarity: Sentiment(polarity=0.3703703703703704, subjectivity=0.5444444444444445)
Tweet : RT @Deep46422630: #HelpTheNeedy
If you help poor people, God will help you! Please help poor people! Giving a helping hand to a needy one i...
Tweet's Polarity: Sentiment(polarity=-0.5, subjectivity=0.6)
How people are reacting on #HelpTheNeedy by analyzing 100 tweets.

General Report: 0.09731225574194326
Weakly Positive

Detailed Report:
7.00% people thought it was positive
27.00% people thought it was weakly positive
6.00% people thought it was strongly positive
2.00% people thought it was negative
8.00% people thought it was weakly negative
0.00% people thought it was strongly negative
50.00% people thought it was neutral

```

Fig 4.8: Tweets collected

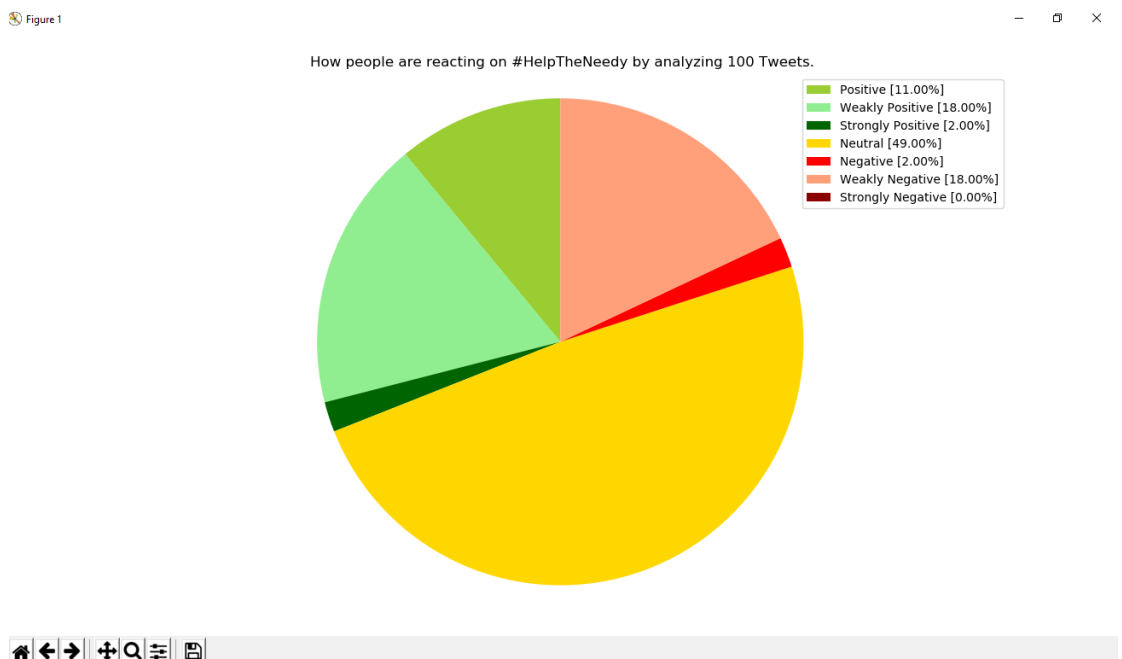


Fig 4.9: Sentiment chart

#### 4.2.1.4 Case 4

```
Project Code Final.py x *REPL* [python] x
Enter Keyword/Tag to search about: Electric Vehicles
Enter how many tweets to search: 100
```

Fig 4.10: Input for Electric Vehicles

```
Project Code Final.py x *REPL* [python] x
Tweet's Polarity: Sentiment(polarity=0.0, subjectivity=0.0)
Tweet : RT @devinamehra: When you read articles like this, remember:
- Tesla does not spend a penny on advertising
- Many companies make electric...
Tweet's Polarity: Sentiment(polarity=0.5, subjectivity=0.5)
Tweet : Battered oil demand faces threat from electric vehicles. "Oil and gas companies have got no ability to defeat elect... https://t.co/
ZttC2YgIx1
Tweet's Polarity: Sentiment(polarity=0.0, subjectivity=0.0)
Tweet : RT @Reuters: ICYMI: Tesla's 'million-mile battery' has the potential to bring the cost of electric vehicles in line with gasoline
models, a...
Tweet's Polarity: Sentiment(polarity=0.0, subjectivity=1.0)
Tweet : Its absolute failure of @CESCLimited to plan and prepare for the disaster. They refused to invest for preparation a... https://t.co/
zVsEi9dvNv
Tweet's Polarity: Sentiment(polarity=-0.05833333333333335, subjectivity=0.6)
How people are reacting on Electric Vehicles by analyzing 100 tweets.
General Report: 0.10572244047619055
Weakly Positive
Detailed Report:
3.00% people thought it was positive
64.00% people thought it was weakly positive
2.00% people thought it was strongly positive
0.00% people thought it was negative
4.00% people thought it was weakly negative
1.00% people thought it was strongly negative
26.00% people thought it was neutral
```

Fig 4.11: Tweets collected

How people are reacting on Electric Vehicles by analyzing 100 Tweets.

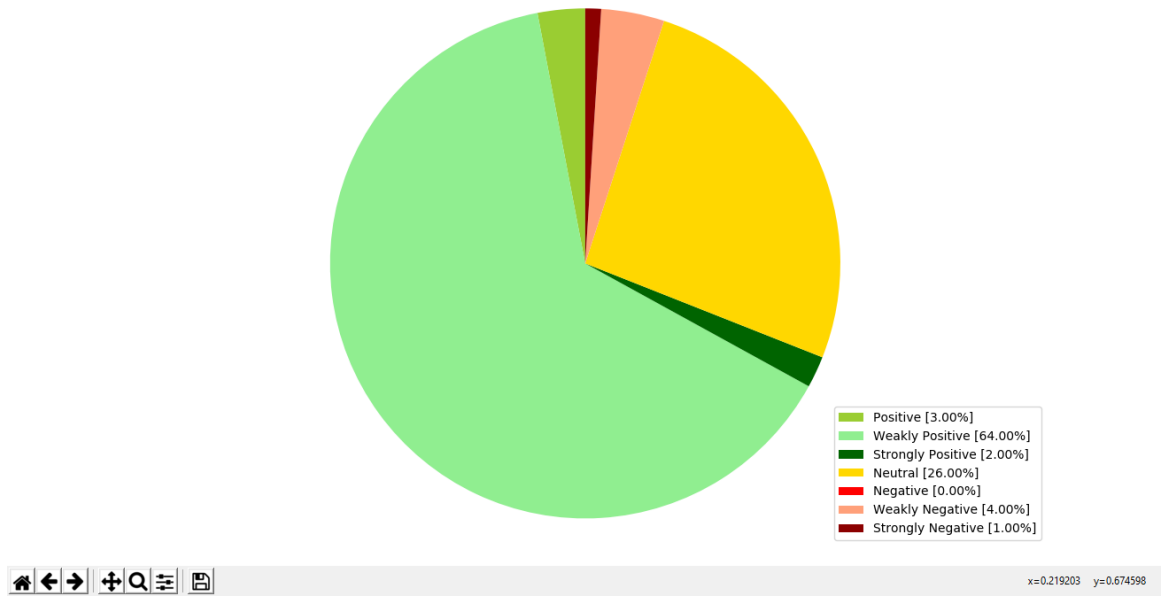


Fig 4.12: Sentiment chart

#### 4.2.1.5 Case 5

```
Project Code Final.py × *REPL* [python] ×
Enter Keyword/Tag to search about: Bitcoin
Enter how many tweets to search: 100
```

Fig 4.13: Input for Bitcoin



```

Project Code Finalzy x *REPL* [python] x
Tweet's Polarity: Sentiment(polarity=0.0, subjectivity=0.0)
Tweet : @ScottTRXWarrior @bitcoin_lyz @SydneyIfergan Nots ON https://t.co/0sYFSryiB5
Tweet's Polarity: Sentiment(polarity=0.0, subjectivity=0.0)
Tweet : Teste de vacina Covid-19 de Oxford tem 50% de chance de suce..
investingcom - News - Noticias - Bitcoin - CryptoCu... https://t.co/TDa18xvNn5
Tweet's Polarity: Sentiment(polarity=0.0, subjectivity=0.0)
Tweet : Trading idea: Bitcoin fractal (19) "Bullrun" REMASTERED https://t.co/2kNh1BFNYm https://t.co/KWxBY15HRX
Tweet's Polarity: Sentiment(polarity=0.0, subjectivity=0.0)
Tweet : Katie Price poses for a sweet snap with 'guard dog' Blade
DailyMailUK - News - Noticias - Bitcoin - CryptoCurrency... https://t.co/167uo51tmx
Tweet's Polarity: Sentiment(polarity=0.35, subjectivity=0.65)
How people are reacting on Bitcoin by analyzing 100 tweets.

General Report: 0.11458997339466091
Weakly Positive

Detailed Report:
14.00% people thought it was positive
25.00% people thought it was weakly positive
4.00% people thought it was strongly positive
2.00% people thought it was negative
7.00% people thought it was weakly negative
0.00% people thought it was strongly negative
48.00% people thought it was neutral

```

Fig 4.14: Tweets collected

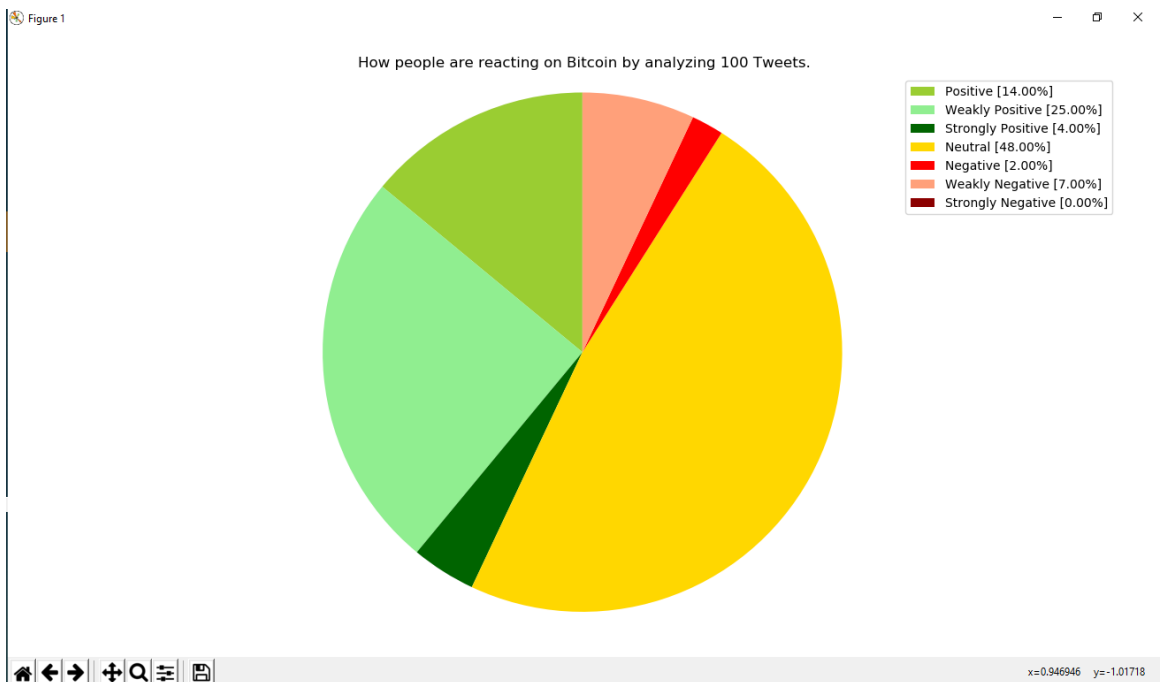


Fig 4.15: Sentiment chart

#### **4.2.2 Continuous search-set**

We analyze our system for different domains continuously for one week. The observations made are done at different courses of time for different domains. The following fields have been considered for observation-

##### **4.2.2.1 Politics**

- 1.1 #2019election**
- 1.2 #Modi1Year**
- 1.3 Arvind Kejriwal**
- 1.4 Rahul Gandhi**
- 1.5 Narendra Modi**

##### **4.2.2.2 Technology**

- 2.1 5G**
- 2.2 Bitcoin**
- 2.3 Electric Vehicle**

##### **4.2.2.3 COVID-19 Pandemic**

- 3.1 #COVID-19**
- 3.2 #HelpTheNeedy**
- 3.3 #Migrant**
- 3.4 #IndiaFightsCorona**

## 1. Politics:

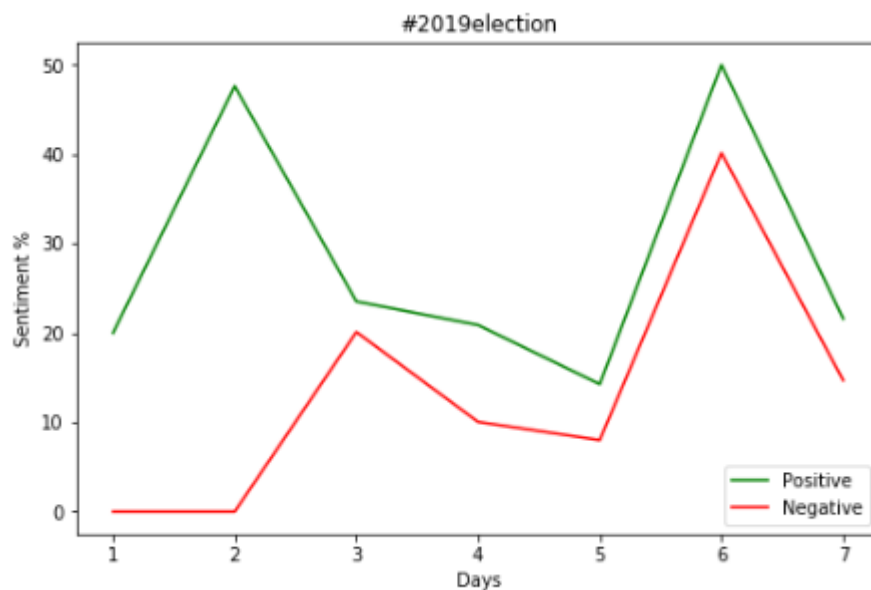
### 1.1 #2019election

**Table 4.1** Analysis of tweets for seven days in May, 19

Day	Positive	Negative
1	20.08	0
2	47.64	0
3	23.58	20.10
4	20.88	10.04
5	14.26	8
6	50.0	40.14
7	21.6	14.7

**Table 4.2** Mathematical Analysis

	Positive	Negative
<b>Average</b>	24.445	7.33
<b>Std Dev</b>	14.12	9.12
<b>Max</b>	50.0	40.14
<b>Min</b>	14.26	0



**Fig 4.16:** Variation in tweets for one week of May, 19.

**Observation:** The topic of politics is always an issue of debate. And during the elections it becomes the favourite discussion. These were the observations made for the month of May, 19 for a week. It shows a positive support for the election.

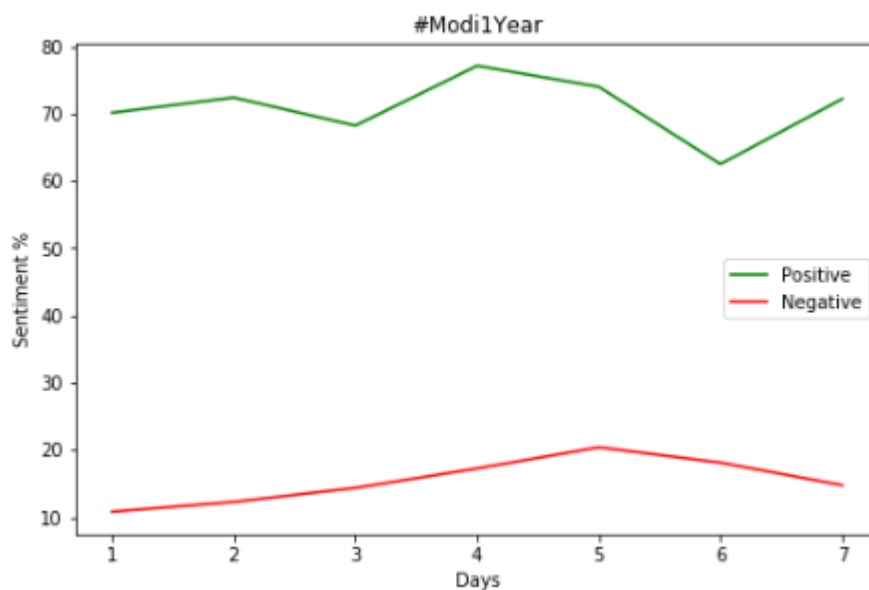
## 1.2 #Modi1Year

**Table 4.3** Variations of tweets for one week

Day	Positive	Negative
1	70.14	10.82
2	72.40	12.28
3	68.23	14.36
4	77.12	17.25
5	74.00	20.40
6	62.50	18.12
7	72.20	14.77

**Table 4.4** Mathematical Analysis

	Positive	Negative
<b>Average</b>	70.94	15.42
<b>Std Dev</b>	4.66	3.36
<b>Max</b>	77.12	20.40
<b>Min</b>	70.14	10.82



**Fig 4.17:** Variation in tweets

**Observation:** The current democratically elected government is quite popular even after a year of coming into power. The public sentiments collected show a maximum positive of 77.12% support and 10.82% opposition.

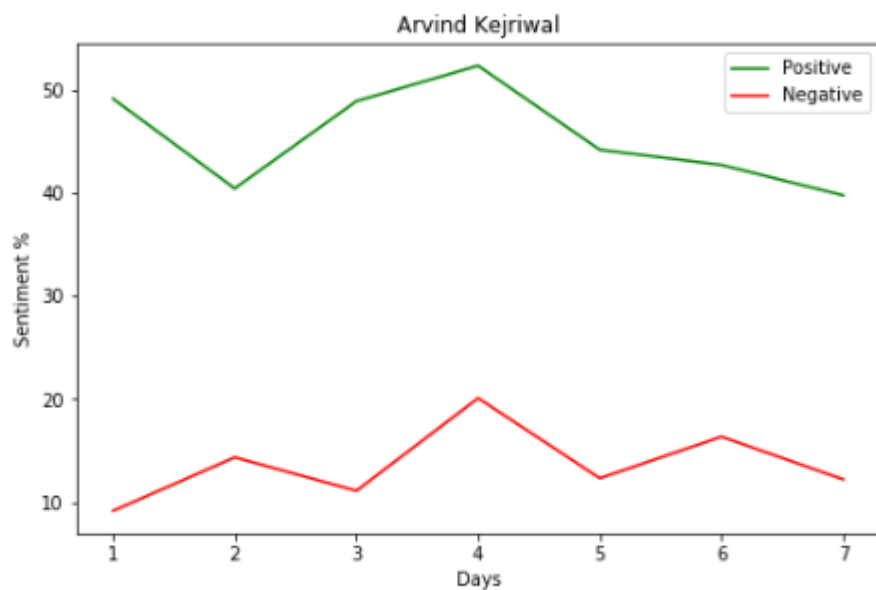
### 1.3 Arvind Kejriwal

**Table 4.5** Variations of tweets for one week

Day	Positive	Negative
1	49.12	9.17
2	40.42	14.36
3	48.88	11.12
4	52.32	20.10
5	44.17	12.33
6	42.70	16.36
7	39.75	12.22

**Table 4.6** Mathematical Analysis

	Positive	Negative
<b>Average</b>	45.34	13.66
<b>Std Dev</b>	4.81	3.64
<b>Max</b>	52.32	20.10
<b>Min</b>	39.75	9.17



**Fig 4.18:** Variation in tweets

**Observations:** Arvind Kejriwal who got re-elected this year again after serving a term as the CM of Delhi. He is a popular figure in his electorate. The public sentiments collected from tweets show overall a positive sentiment.

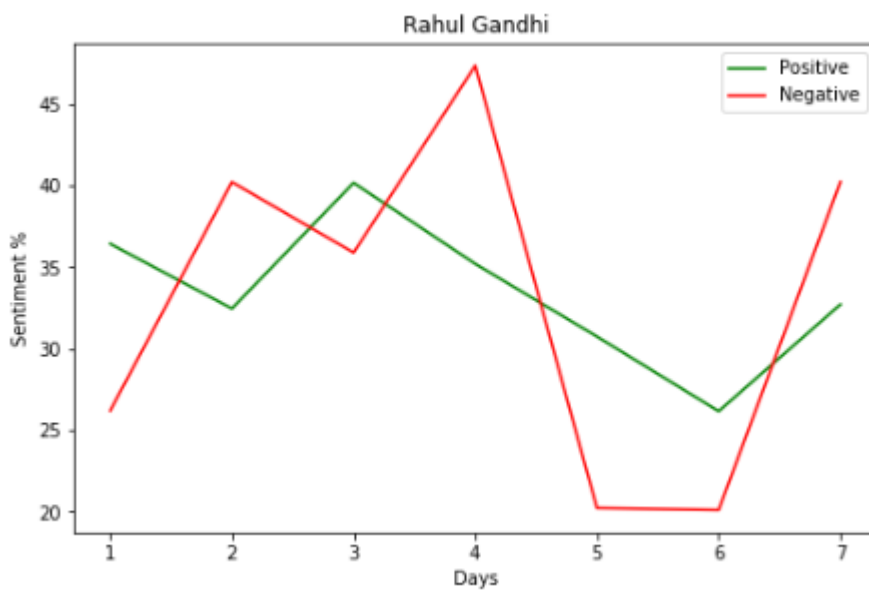
## 1.4 Rahul Gandhi

**Table 4.7** Variations of tweets for one week

Day	Positive	Negative
1	36.43	26.20
2	32.46	40.22
3	40.17	35.89
4	35.20	47.36
5	30.74	20.22
6	26.16	20.10
7	32.71	40.23

**Table 4.8** Mathematical Analysis

	Positive	Negative
<b>Average</b>	33.41	32.88
<b>Std Dev</b>	4.46	10.76
<b>Max</b>	40.17	47.36
<b>Min</b>	26.16	20.10



**Fig 4.19:** Analysis

**Observation:** Rahul Gandhi is the face of Congress, the grand old party of India. He serves as the Member of Parliament and was the former President of INC. He got a series of mixed reactions for his various public statements. With an average support of 33.41% shows he is not very popular among the masses.

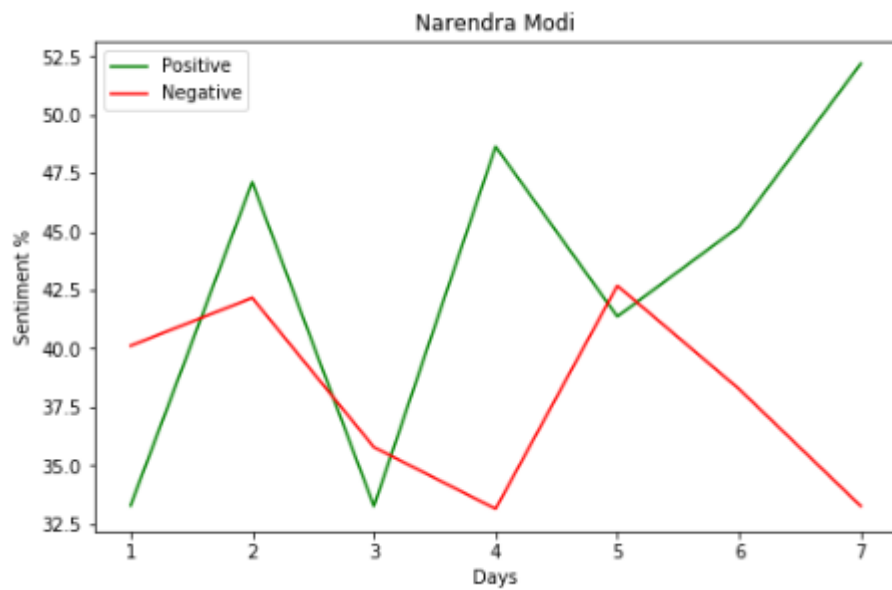
## 1.5 Narendra Modi

**Table 4.9** Variations of tweets for one week

Day	Positive	Negative
1	33.28	40.12
2	47.12	42.17
3	33.26	35.78
4	48.64	33.14
5	41.37	42.69
6	45.22	38.27
7	52.19	33.26

**Table 4.10** Mathematical Analysis

	Positive	Negative
<b>Average</b>	43.01	37.19
<b>Std Dev</b>	7.42	3.97
<b>Max</b>	52.19	42.69
<b>Min</b>	33.26	33.14



**Fig 4.20:** Variation in tweets

**Observation:** Namo being the PM of India for consecutive term is a popular public leader. These observations were made during one of the important parliamentary decisions of his government. There were mixed reactions, ultimately the tweets in support increased over a period of days.

**Inference:**

We observed the tweets collected for a week, we are able to see the order of average sentiments observed over a week as:

**Table 4.11** Comparison of average support for politicians

Arvind Kejriwal	45.34%
Narendra Modi	43.01%
Rahul Gandhi	33.41%

This data was collected over the period of elections in Delhi, and the data we analysed shows a favourable support for Arvind Kejriwal. Eventually, Mr Kejriwal won the electoral mandate with a popular support and formed the government in the state assembly.

This shows that our data was successful in predicting the outcome to a great extent.



## 2. Technology:

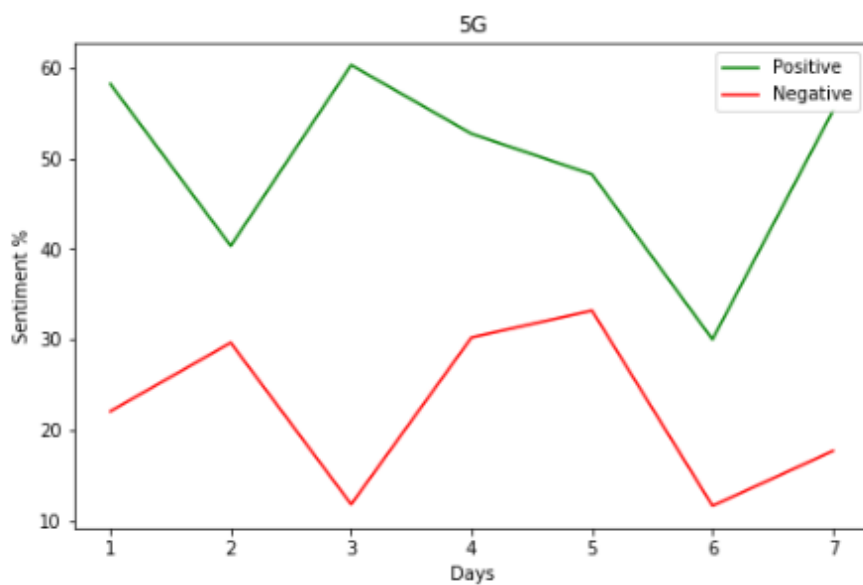
### 2.1 5G

**Table 4.12** Variations of tweets for one week

Day	Positive	Negative
1	58.19	22.06
2	40.32	29.63
3	60.28	11.79
4	52.69	30.22
5	48.21	33.17
6	29.99	11.62
7	55.22	17.68

**Table 4.13** Mathematical Analysis

	Positive	Negative
<b>Average</b>	49.27	22.31
<b>Std Dev</b>	10.81	8.95
<b>Max</b>	60.28	33.17
<b>Min</b>	29.99	11.62



**Fig 4.21:** Variation in tweets

**Observation:** As we can see, the tweets opposing were less than the tweets made in its support. Maximum generated tweets involve a maximum of about 60% in its support, whereas, the tweets in opposition were about 11% in opposition on that day.

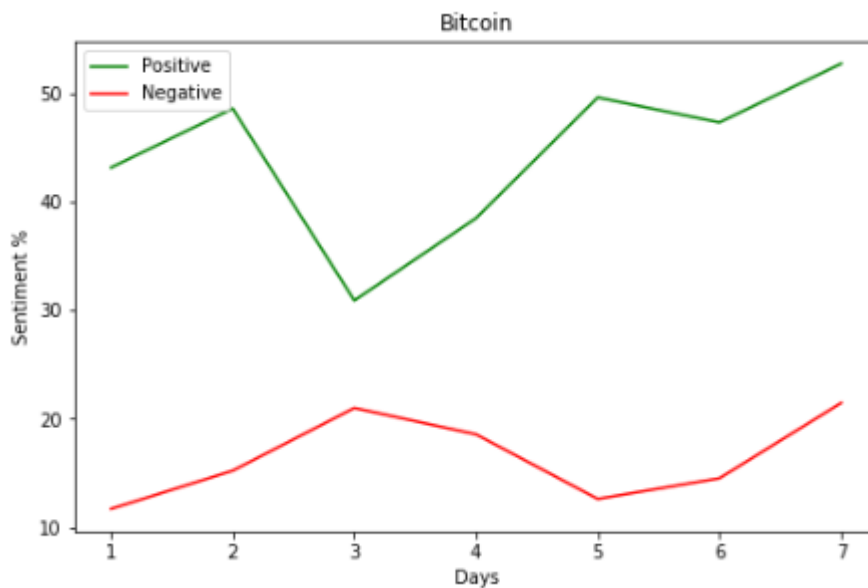
## 2.2 Bitcoin

**Table 4.14** Variations in tweets for a week

Day	Positive	Negative
1	43.12	11.68
2	48.56	15.20
3	30.88	20.98
4	38.46	18.55
5	49.58	12.60
6	47.28	14.50
7	52.69	21.44

**Table 4.15** Mathematical Analysis

	Positive	Negative
<b>Average</b>	44.37	16.42
<b>Std Dev</b>	7.54	3.93
<b>Max</b>	52.69	21.44
<b>Min</b>	30.88	11.68



**Fig 4.22:** Variation in tweets

**Observation:** A type of cryptocurrency, Bitcoin has gathered popularity in past few years as the currency of the future. It is quite a hype, and it can be observed by looking at the support for it in the graph.

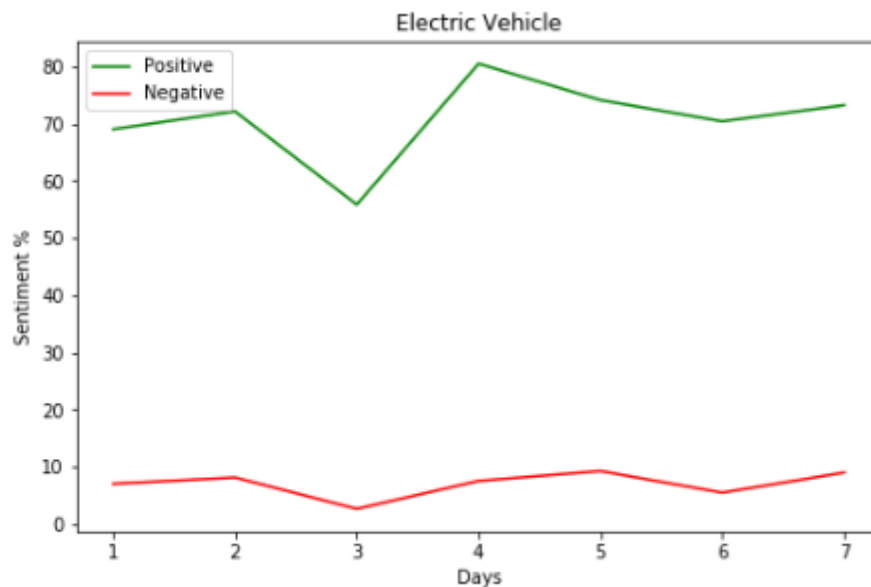
## 2.3 Electric Vehicle

**Table 4.16** Variations of tweets for one week

Day	Positive	Negative
1	69.08	7.02
2	72.19	8.12
3	55.88	2.63
4	80.54	7.53
5	74.17	9.25
6	70.45	5.47
7	73.28	8.98

**Table 4.17** Mathematical Analysis

	Positive	Negative
<b>Average</b>	70.80	7.0
<b>Std Dev</b>	7.53	2.31
<b>Max</b>	80.54	9.25
<b>Min</b>	55.88	2.63



**Fig 4.23:** Variation in tweets

**Observation:** Due to pollution caused by the conventional vehicles, there is an increasing trend in support for switching to electric vehicles. This trend is widely discussed on social media platforms. The fig shows positive support to this idea.

**Inference:**

From the daily analysis of these popular trends in technology, we observe an average of positive sentiments for these upcoming technologies.

**Table 4.18** Comparison of support for various technologies

<b>Electric Vehicle</b>	70.80%
<b>5G</b>	49.27%
<b>Bitcoin</b>	44.37%

It can be observed that among the upcoming technologies, the concept of Electric vehicles is very popular. Other technologies are also trending and have gained good support.

### 3. COVID-19 Pandemic

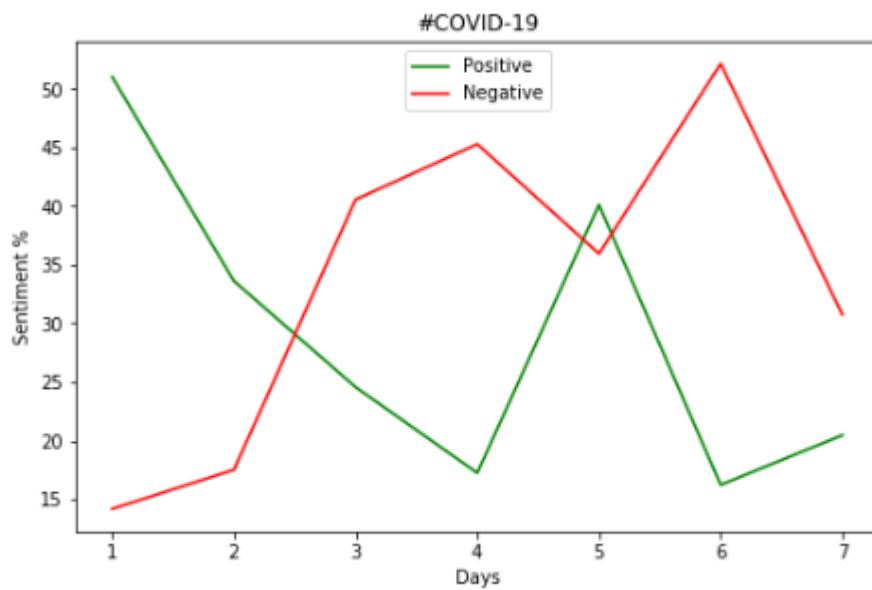
#### 3.1 #COVID-19

**Table 4.19** Variations of tweets for one week

Day	Positive	Negative
1	51.02	14.18
2	33.62	17.52
3	24.57	40.56
4	17.25	45.29
5	40.12	35.94
6	16.23	52.15
7	20.47	30.75

**Table 4.20** Mathematical Analysis

	Positive	Negative
<b>Average</b>	29.04	33.77
<b>Std Dev</b>	13.06	14.01
<b>Max</b>	51.02	52.15
<b>Min</b>	16.23	14.18



**Fig 4.24:** Variation in tweets

**Observation:** The pandemic has become a widespread issue of difficulties all over the world, the countries have gone into long lockdowns. Therefore, we observe a negative sentiment towards it over social media.

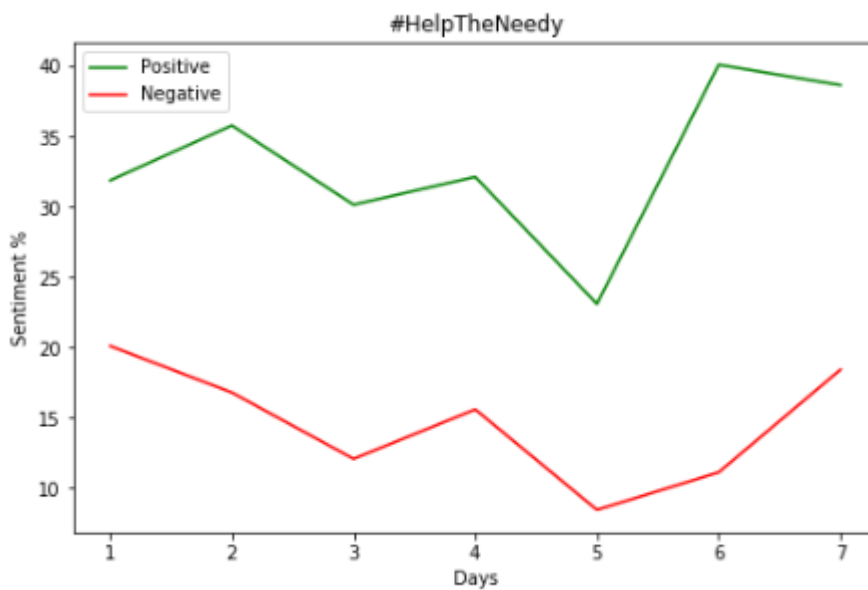
### 3.2 #HelpTheNeedy

**Table 4.21** Variations of tweets for one week

Day	Positive	Negative
1	31.89	20.11
2	35.80	16.80
3	30.16	12.08
4	32.14	15.60
5	23.10	8.45
6	40.12	11.12
7	38.67	18.42

**Table 4.22** Mathematical Analysis

	Positive	Negative
<b>Average</b>	33.13	14.65
<b>Std Dev</b>	5.75	4.22
<b>Max</b>	40.12	20.11
<b>Min</b>	23.10	8.45



**Fig 4.25:** Variation in tweets

**Observation:** The pandemic caused months long lockdown in economies, resulting in a lot of trouble all over to the masses. The poor mostly who depend on daily wages have been hit the hardest. Therefore, there was a trend to help the needy in times of trouble.

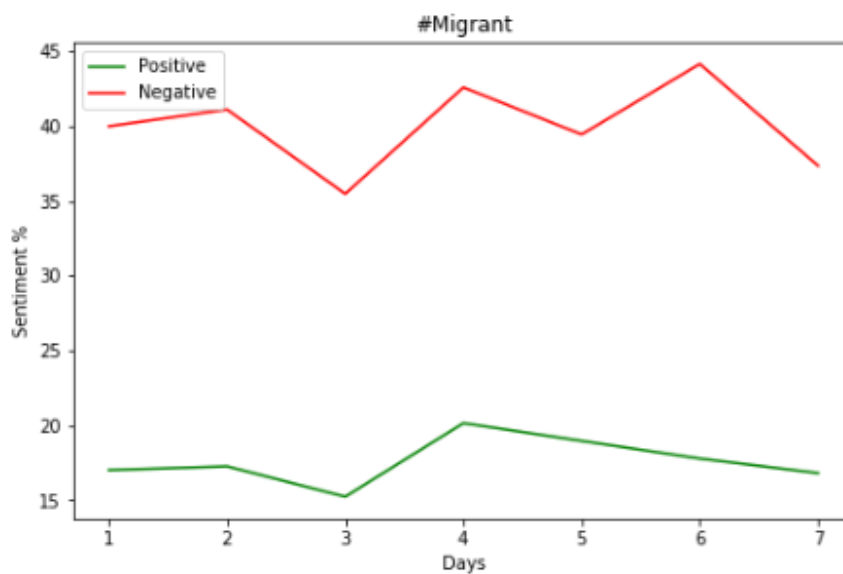
### 3.3 #Migrant

**Table 4.23** Variations of tweets for one week

Day	Positive	Negative
1	17.00	40.00
2	17.25	41.12
3	15.22	35.48
4	20.14	42.60
5	18.95	39.45
6	17.77	44.17
7	16.80	37.36

**Table 4.24** Mathematical Analysis

	Positive	Negative
<b>Average</b>	17.59	44.02
<b>Std Dev</b>	1.59	2.97
<b>Max</b>	20.14	44.17
<b>Min</b>	15.22	35.48



**Fig 4.26:** Variation in tweets

**Observation:** Migrant labourers who have been hit hard, are unable to live in big cities due to higher cost of living. Without a source of livelihood, they have started shifting back to their native places, these migrants have been a topic of debate as becoming a possible carriers of coronavirus.

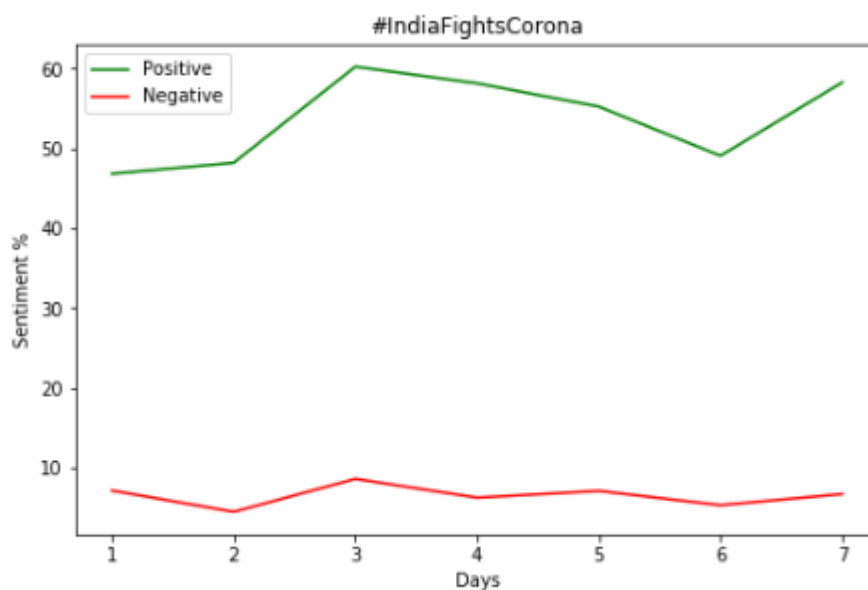
### 3.4 #IndiaFightsCorona

**Table 4.25** Variations of tweets for one week

Day	Positive	Negative
1	46.82	7.19
2	48.20	4.52
3	60.22	8.65
4	58.14	6.30
5	55.23	7.14
6	49.07	5.36
7	58.25	6.75

**Table 4.26** Mathematical Analysis

	Positive	Negative
<b>Average</b>	53.70	6.55
<b>Std Dev</b>	5.54	1.34
<b>Max</b>	60.22	8.65
<b>Min</b>	46.82	4.52



**Fig 4.27:** Variation in tweets

**Observation:** Coronavirus has caused hardships to countries all over the world. In India, the doctors and other frontline corona warriors are giving their best to control its spread. The users on social media have expressed their support for Indians fighting the pandemic.



**Inference:**

COVID-19 which started from China has spread all over the world. It has taken the form of a pandemic affecting lives worldwide causing hardships. The nations are putting forward their best efforts to control it. Here we have analysed the response of various issues related to this pandemic.

**Table 4.27** Comparison of sentiments for issues related to pandemic

<b>#IndiaFightsCorona</b>	53.70%
<b>#HelpTheNeedy</b>	33.13%
<b>#COVID-19</b>	29.04%
<b>#Migrant</b>	17.59%

We can clearly see that India supports its corona warriors and help its needy citizens in this hour of need. There is a negative sentiment for the spread of coronavirus.

## **CHAPTER 5: CONCLUSION**

### **5.3.1 Conclusions:**

This project presents an insight into a different way of analyzing various topics like popularity and support for a leader or a business or a particular product. The social media sites are quite popular as they empower a common citizen to express their views. The businesses or political groups can benefit from this by analyzing the sentiments of public over a demographic region or group of people on these social media platforms.

Earlier the businesses relied on conducting surveys to gather a taste of the public and launched marketing campaigns or products based on that information. This process took a lot of time and was tedious. But using social media analytics has proved to save a lot of time and are better in giving favourable outcomes in real time. Sentiment analysis is one such field of analytics.

Through this project we have analyzed various trending issues, concerning majority of population over different geographical regions. The project aims to provide the subjectivity and polarity of tweets collected and performed sentiment analysis on this gathered data.

### **5.3.2 Future Scope:**

This project further be implemented with OS kernel to create a better user-interface. Moreover algorithms implemented can be bettered by further improving the data mining tools and objectivity of a particular tweet to give better result terms of polarity and subjectivity.

### **5.3.3 Applications:**

5.3.3.1 Political campaign strategy

5.3.3.2 Product launch

5.3.3.3 Trend Analysis

5.3.3.4 Most popular sentiments

## REFERENCES:

- [1]:B. Pang and L. Lee(2008), “Opinion mining and sentiment analysis” . Foundations and Trends in Information Retrieval 2(1-2), pages 1-135.
- [2]: Y. Kim and S. Myaeng. “Opinion analysis based on lexical clues and their expansion” . In Proceedings of 6th NTCIR Evaluation Workshop, 2007.
- [3]: M. Cha, H. Haddadi, F. Benevenuto, and K. P. Gummadi. “Measuring User Influence in Twitter: The Million Follower Fallacy”. Int’l AAAI Conference on Weblogs and Social Media (ICWSM), 2010.
- [4]: Rodrigo Sandoval-Almazan, David Valle-Cruz , “Facebook impact and sentiment analysis on political campaigns on Political rallies”,19th Annual International Conference -May 2018.
- [5]: Shahana P.H,Bini Ommanb , “Evaluation of Features on Sentimental Analysis”, International Conference on Information and Communication Technologies, Pg 1585-1592, 2014.
- [6]: Rahul Rajput, Arun Kumar Solanki, “Review of Sentimental Analysis Methods using Lexicon Based Approach”, International Journal of Computer Science and Mobile Computing, IJCSMC, Vol. 5, Issue. 2, February 2016, pg.159 – 166.
- [7]: Shubham Goyal, “Sentimental Analysis of Twitter Data using Text Mining and Hybrid Classification Approach”, International Journal of Advanced Research, Ideas and Innovations in Technology, Volume2, Issue 5, 2016.
- [8]: Gurneet Kaur, Abhinash Singla, “Sentimental Analysis of Flipkart reviews using Naïve Bayes and Decision Tree algorithm”, International Journal of Advanced Research in Computer Engineering & Technology (IJARCET) Volume 5, Issue 1, January 2016.

- [9]: K. R. Aida-zade, S. S. Rustamov, E. E. Mustafayev, and N. T. Aliyeva, "Human-Computer Dialogue Understanding Hybrid System, " IEEE Xplore, International Symposium on Innovations in Intelligent Systems and Applications. Trabzon, Turkey, July 2012, pp. 1-
- [10]: X. Lei, X. Qian and G. Zhao, "Rating Prediction Based on Social Sentiment From Textual Reviews," in IEEE Transactions on Multimedia, vol. 18, no. 9, pp. 1910-1921, Sept. 2016, doi: 10.1109/TMM.2016.2575738.
- [11]: Zhen Hai, "Analyzing Sentiments in One Go: A Supervised Joint Topic Modeling Approach", IEEE Transactions on Knowledge and Data Engineering, Vol 29, pg: 1172-1185.
- [12]: S. Sun, H. Liu, H. Lin and A. Abraham, "Twitter part-of-speech tagging using pre-classification Hidden Markov model," 2012 IEEE International Conference on Systems, Man, and Cybernetics (SMC), Seoul, 2012, pp. 1118-1123, doi: 10.1109/ICSMC.2012.6377881.
- [13]: M. Bouazizi and T. Ohtsuki, "Sentiment analysis: From binary to multi-class classification: A pattern-based approach for multi-class sentiment analysis in Twitter," 2016 IEEE International Conference on Communications (ICC), Kuala Lumpur, 2016, pp. 1-6, doi: 10.1109/ICC.2016.7511392.
- [14]: R. Zhao and K. Mao, "Fuzzy Bag-of-Words Model for Document Representation," in IEEE Transactions on Fuzzy Systems, vol. 26, no. 2, pp. 794-804, April 2018, doi: 10.1109/TFUZZ.2017.2690222.
- [15]: Manvee Chauhan ,Divakar Yadav , "Sentimental Analysis of Product Based Reviews Using Machine Learning Approaches", Journal of Network Communications and Emerging Technologies (JNCET) Volume 5, Special Issue 2, December (2015).

[16]: R. Wagh and P. Punde, "Survey on Sentiment Analysis using Twitter Dataset," 2018 Second International Conference on Electronics, Communication and Aerospace Technology (ICECA), Coimbatore, 2018, pp. 208-211, doi: 10.1109/ICECA.2018.8474783.

[17] [http://docs.tweepy.org/en/latest/getting\\_started.html#introduction](http://docs.tweepy.org/en/latest/getting_started.html#introduction)

[18] [https://textblob.readthedocs.io/en/dev/advanced\\_usage.html#sentiment-analyzers](https://textblob.readthedocs.io/en/dev/advanced_usage.html#sentiment-analyzers)

[19] [https://matplotlib.org/api/\\_as\\_gen/matplotlib.pyplot.html?highlight=pyplot#module-matplotlib.pyplot](https://matplotlib.org/api/_as_gen/matplotlib.pyplot.html?highlight=pyplot#module-matplotlib.pyplot)

[20] <https://docs.python.org/3.4/library/re.html>

# APPENDIX

## A1. Computer Program

```
import sys,tweepy, csv, re
from textblob import TextBlob
import matplotlib.pyplot as plt

class SentimentAnalysis:

    def __init__(self):
        self.tweets = []
        self.tweetText = []

    def DownloadData(self):
        # authenticating
        consumerKey = "# Your key here "
        consumerSecret = "# Your key here"
        accessToken = "# Your key here"
        accessTokenSecret = "# Your key here "
        auth = tweepy.OAuthHandler(consumerKey, consumerSecret)
        auth.set_access_token(accessToken, accessTokenSecret)
        api = tweepy.API(auth)

        # input for term to be searched and how many tweets to search
        searchTerm = input("Enter Keyword/Tag to search about: ")
        NoOfTerms = int(input("Enter how many tweets to search: "))

        # searching for tweets
        self.tweets = tweepy.Cursor(api.search, q=searchTerm, lang = "en").items(NoOfTerms)

        # Open/create a file to append data to
        csvFile = open("C:/Users/hp/Desktop/8th SEM/Final/tweetstore.csv", 'a')

        # Use csv writer
        csvWriter = csv.writer(csvFile)

        # creating some variables to store info
        polarity = 0
        positive = 0
        wpositive = 0
        spositive = 0
        negative = 0
        wnegative = 0
        snegative = 0
        neutral = 0

        # iterating through tweets fetched
        for tweet in self.tweets:
            #Append to temp so that we can store in csv later. I use encode UTF-8
            self.tweetText.append(self.cleanTweet(tweet.text).encode('utf-8'))

            analysis = TextBlob(tweet.text)
            print("Tweet : {}".format(tweet.text))
            print("Tweet\'s Polarity: {}".format(analysis.sentiment)) # print tweet's polarity

            polarity += analysis.sentiment.polarity # adding up polarities to find the average later
```

```

# adding reaction of how people are reacting to find average later
if (analysis.sentiment.polarity == 0):
    neutral += 1

elif (analysis.sentiment.polarity > 0 and analysis.sentiment.polarity <= 0.3):
    wpositive += 1

elif (analysis.sentiment.polarity > 0.3 and analysis.sentiment.polarity <= 0.6):
    positive += 1

elif (analysis.sentiment.polarity > 0.6 and analysis.sentiment.polarity <= 1):
    spositive += 1

elif (analysis.sentiment.polarity > -0.3 and analysis.sentiment.polarity <= 0):
    wnegative += 1

elif (analysis.sentiment.polarity > -0.6 and analysis.sentiment.polarity <= -0.3):
    negative += 1

elif (analysis.sentiment.polarity > -1 and analysis.sentiment.polarity <= -0.6):
    snegative += 1

csvWriter.writerow(self.tweetText)
csvFile.close()

# finding average of how people are reacting
positive_wt = self.percentage(positive, NoOfTerms)
wpositive_wt = self.percentage(wpositive, NoOfTerms)
spositive_wt = self.percentage(spositive, NoOfTerms)
negative_wt = self.percentage(negative, NoOfTerms)
wnegative_wt = self.percentage(wnegative, NoOfTerms)
snegative_wt = self.percentage(snegative, NoOfTerms)
neutral_wt = self.percentage(neutral, NoOfTerms)

# finding average reaction
polarity = polarity / NoOfTerms

# printing out data
print("How people are reacting on " + searchTerm + " by analyzing " + str(NoOfTerms) + " tweets.")
print()
print("General Report: {}".format(polarity))

if (polarity == 0):
    print("Neutral")
elif (polarity > 0 and polarity <= 0.3):
    print("Weakly Positive")
elif (polarity > 0.3 and polarity <= 0.6):
    print("Positive")
elif (polarity > 0.6 and polarity <= 1):
    print("Strongly Positive")
elif (polarity > -0.3 and polarity <= 0):
    print("Weakly Negative")
elif (polarity > -0.6 and polarity <= -0.3):
    print("Negative")
elif (polarity > -1 and polarity <= -0.6):
    print("Strongly Negative")

print()
print("Detailed Report: ")
print(str(positive_wt) + "% people thought it was positive")
print(str(wpositive_wt) + "% people thought it was weakly positive")
print(str(spositive_wt) + "% people thought it was strongly positive")
print(str(negative_wt) + "% people thought it was negative")
print(str(wnegative_wt) + "% people thought it was weakly negative")
print(str(snegative_wt) + "% people thought it was strongly negative")
print(str(neutral_wt) + "% people thought it was neutral")

```

```

self.plotPieChart(positive_wt, wpositive_wt, spositive_wt, negative_wt, wnegative_wt,
                  snegative_wt, neutral_wt, searchTerm, NoOfTerms)

def cleanTweet(self, tweet):
    # Remove Links, Special Characters etc from tweet
    return ' '.join(re.sub("(@[A-Za-z0-9]+)|([^0-9A-Za-z \t]) | (\w +:\ / \ / \S +)", "", tweet).split())

# function to calculate percentage
def percentage(self, part, whole):
    temp = 100 * float(part) / float(whole)
    return format(temp, '.2f')

def plotPieChart(self, positive, wpositive, spositive, negative, wnegative, snegative, neutral, searchTerm, noOfSearchTerms):
    labels = ['Positive [' + str(positive) + '%]', 'Weakly Positive [' + str(wpositive) + '%]',
              'Strongly Positive [' + str(spositive) + '%]', 'Neutral [' + str(neutral) + '%]',
              'Negative [' + str(negative) + '%]', 'Weakly Negative [' + str(wnegative) + '%]',
              'Strongly Negative [' + str(snegative) + '%]']

    sizes = [positive, wpositive, spositive, neutral, negative, wnegative, snegative]
    colors = ['yellowgreen', 'lightgreen', 'darkgreen', 'gold', 'red', 'lightsalmon', 'darkred']
    patches, texts = plt.pie(sizes, colors=colors, startangle=90)
    plt.legend(patches, labels, loc="best")
    plt.title("How people are reacting on " + searchTerm + " by analyzing " + str(noOfSearchTerms) + " Tweets.")
    plt.axis('equal')
    plt.tight_layout()
    plt.show()

if __name__ == "__main__":
    sa = SentimentAnalysis()
    sa.DownloadData()

```



## SOCIAL MEDIA ANALYTICS

### ORIGINALITY REPORT

<b>12%</b>	<b>6%</b>	<b>4%</b>	<b>9%</b>
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

### PRIMARY SOURCES

<b>1</b>	<b>Submitted to Jaypee University of Information Technology</b> Student Paper	<b>6%</b>
<b>2</b>	<b>www.ijrte.org</b> Internet Source	<b>1%</b>
<b>3</b>	<b>www.enlight.ru</b> Internet Source	<b>&lt;1%</b>
<b>4</b>	<b>ieeexplore.ieee.org</b> Internet Source	<b>&lt;1%</b>
<b>5</b>	<b>"ICDSMLA 2019", Springer Science and Business Media LLC, 2020</b> Publication	<b>&lt;1%</b>
<b>6</b>	<b>M. Vadivukarassi, N. Puviarasan, P. Aruna. "A framework of keyword based image retrieval using proposed Hog_Sift feature extraction method from Twitter Dataset", Procedia Computer Science, 2018</b> Publication	<b>&lt;1%</b>
<b>7</b>	<b>de.scribd.com</b> Internet Source	<b>&lt;1%</b>

**JAYPEE UNIVERSITY OF INFORMATION TECHNOLOGY, WAKNAGHAT**

**PLAGIARISM VERIFICATION REPORT**

Date: 14/07/2020

Type of Document (Tick): **B.Tech Project Report**

Name: SUCHETAN PATHANIA

Department: CSE

Enrolment No: 161217

Contact No.: 9459845982

E-mail.: suchetanp.47@gmail.com

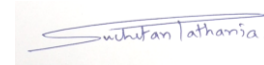
Name of the Supervisor: Dr Jagpreet Sidhu

Title of the Thesis/Dissertation/Project Report/Paper (In Capital letters): SOCIAL MEDIA ANALYTICS USING SENTIMENT ANALYSIS

**UNDERTAKING**

I undertake that I am aware of the plagiarism related norms/ regulations, if I found guilty of any plagiarism and copyright violations in the above thesis/report even after award of degree, the University reserves the rights to withdraw/revoke my degree/report. Kindly allow me to avail Plagiarism verification report for the document mentioned above.

- Total No. of Pages = 57
- Total No. of Preliminary pages = 10
- Total No. of pages accommodate bibliography/references = 3



(Signature of Student)

**FOR DEPARTMENT USE**

**12%**

We have checked the thesis/report as per norms and found **Similarity Index** at .....(%). Therefore, we are forwarding the complete thesis/report for final plagiarism check. The plagiarism verification report may be handed over to the candidate.



(Signature of Guide/Supervisor)

Signature of HOD

**FOR LRC USE**

The above document was scanned for plagiarism check. The outcome of the same is reported below:

Copy Received on	Excluded	Similarity Index (%)	Abstract & Chapters Details	
	<ul style="list-style-type: none"><li>• All Preliminary Pages</li><li>• Bibliography/ Images/Quotes</li><li>• 14 Words String</li></ul>		Word Counts	
<b>Report Generated on</b>			Character Counts	
		<b>Submission ID</b>	Page counts	
			File Size	

Checked by  
Name & Signature

Librarian

Please send your complete Thesis/Report in (PDF) & DOC (Word File) through your Supervisor/Guide at [plagcheck.juit@gmail.com](mailto:plagcheck.juit@gmail.com)