

# A MINOR PROJECT REPORT

## ON

## SOCIAL MIRROR

SUBMITTED IN PARTIAL FULFILLMENT FOR THE AWARD OF  
DEGREE OF BACHELOR OF TECHNOLOGY  
IN  
ELECTRONICS AND COMMUNICATION ENGINEERING



*Student Name:*

Name: Abhishek Jain

Enrollment No. : 9917102234

Name: Harshit Sachdev

Enrollment No. : 9917102235

Name: Nikhil Jain

Enrollment No. : 9917102258

*Under the guidance of :*

Mr. VARUN GOEL

DEPARTMENT OF ELECTRONICS AND COMMUNICATION ENGINEERING

JAYPEE INSTITUTE OF INFORMATION TECHNOLOGY, NOIDA,

JUNE 2020

# **CERTIFICATE**

This is to certify that the minor project report entitled, “SOCIAL MIRROR” submitted by “Abhishek Jain, Harshit Sachdev, Nikhil Jain” in partial fulfillment of the requirements for the award of Bachelor of Technology Degree in Electronics and Communication Engineering of the Jaypee Institute of Information Technology, Noida is an authentic work carried out by them under my supervision and guidance. The matter embodied in this report is original and has not been submitted for the award of any other degree.

Signature of Supervisor

Name of Supervisor : Mr. VARUN GOEL

ECE Department,

JIIT, Sec-128,

Noida-201304

Date:

# DECLARATION

We hereby declare that this written submission represents our own ideas in our own words and where other's ideas or words have been included, have adequately cited and referenced the original sources. We also declare that we have adhered to all principles of academic honesty and integrity and have not misinterpreted or fabricated or falsified any idea/data/fact/source in our submission.

Place: Noida

Dated:

Name: Abhishek Jain

Enrollment No. : 9917102234

.

Name: Harshit Sachdev

Enrollment No. : 9917102235

.

Name: Nikhil Jain

Enrollment No. : 9917102258

# ABSTRACT

Social media have received more attention nowadays. Public and private opinion about a wide variety of subjects are expressed and spread continually via numerous social media. Twitter is one of the social media that is gaining popularity. Twitter offers organizations a fast and effective way to analyze customers' perspectives toward the critical to success in the market place. Developing a program for sentiment analysis is an approach to be used to computationally measure customers' perceptions. This paper reports on the design of a sentiment analysis, extracting a vast amount of tweets. Prototyping is used in this development. Results classify customers' perspective via tweets into positive and negative, which is represented in a pie chart and html page. Millions of people are using social network sites to express their emotions, opinion and disclose about their daily lives. However, people write anything such as social activities or any comment on products. Through the online communities provide an interactive forum where consumers inform and influence others. Moreover, social media provides an opportunity for business that giving a platform to connect with their customers such as social media to advertise or speak directly to customers for connecting with customer's perspective of products and services. In contrast, consumers have all the power when it comes to what consumers want to see and how consumers respond. With this, the company's success and failure is publicly shared and end up with word of mouth. However, the social network can change the behavior and decision making of consumers, for example, 87 percent of internet users are influenced in their purchase and decision by customer's review. So that, if organization can catch up faster on what their customer's think, it would be more beneficial to organize to react on time and come up with a good strategy to compete their competitors.

# ACKNOWLEDGEMENT

We would like to express our special gratitude to our esteemed college, Jaypee Institute of Information Technology, for giving us this wonderful opportunity of making the minor project of 2020, under the valuable supervision, timely suggestions and inspiration offered by our most respectable project mentor Mr. Varun Goel . It was because of his continuous encouragement that this report reached its successful completion. We also place on record and warmly acknowledge the contribution of our panel coordinators. Last but not the least we express our sincere thanks to all our friends who have patiently extended their support for accomplishing this undertaking.

Name: Abhishek Jain

Enrollment No. : 9917102234

Name:Harshit Sachdev

Enrollment No. : 9917102235

Name: Nikhil Jain

Enrollment No. : 9917102258

# Contents

<i>CERTIFICATE</i> . . . . .	i
<i>DECLARATION</i> . . . . .	ii
<i>ABSTRACT</i> . . . . .	iii
<i>ACKNOWLEDGEMENT</i> . . . . .	iv
<b>1 INTRODUCTION</b>	<b>1</b>
1.1 IMPORTANCE OF SOCIAL MEDIA . . . . .	1
1.2 PROBLEMS FACED . . . . .	1
1.3 OBJECTIVE . . . . .	2
1.4 APPROACHES . . . . .	3
1.4.1 Lexicon-based Approach . . . . .	3
1.4.2 Machine-learning-based Approach . . . . .	3
<b>2 LITERATURE REVIEW</b>	<b>4</b>
2.1 SOCIAL MEDIA . . . . .	4
2.1.1 ALGORITHM OF LEXICON METHOD . . . . .	5
2.1.2 ALGORITHM OF MACHINE LEARNING METHOD . . . . .	6
<b>3 PROJECT DESCRIPTION</b>	<b>8</b>
3.1 CREATION OF TWITTER DEVELOPER ACCOUNT . . . . .	8
3.2 BLOCK DIAGRAM . . . . .	10
3.3 DESCRIPTION OF BLOCK DIAGRAM . . . . .	10
<b>4 MACHINE LEARNING APPROACH</b>	<b>12</b>
4.1 DESCRIPTION OF FLOWCHART . . . . .	13
4.1.1 GATHER UNSTRUCTURED DATA . . . . .	13

4.1.2	DATA CLEANING AND PREPROCESSING . . . . .	13
4.1.3	VECTORIZED DATA . . . . .	13
4.1.4	ALGORITHM SELECTION . . . . .	14
<b>5</b>	<b>CHALLENGES FACED</b>	<b>15</b>
<b>6</b>	<b>CONCLUSION AND FUTURE ENHANCEMENTS</b>	<b>17</b>
6.1	CONCLUSION . . . . .	17
6.2	ACCURACY OF DIFFERENT ALGORITHMS . . . . .	17
6.3	FUTURE ENHANCEMENTS . . . . .	18
<b>7</b>	<b>APPENDIX</b>	<b>20</b>
7.1	CODE OF LEXICON BASED APPROACH . . . . .	20

# List of Figures

2.1	Different types of social media platforms . . . . .	5
2.2	Lexicon Approach . . . . .	6
2.3	Block Diagram of Lexicon Approach . . . . .	6
2.4	Word Embedding . . . . .	7
3.1	Step 1 for creation of developer account . . . . .	8
3.2	Step 2 for creation of developer account . . . . .	9
3.3	Step 3 for creation of developer account . . . . .	9
3.4	Step 4 successful creation of developer account . . . . .	9
3.5	Project Description . . . . .	10
4.1	Flowchart . . . . .	12
5.1	Crashing of Ram due to large dataset . . . . .	16
5.2	Error showing unable to allocate memory . . . . .	16
5.3	Caption . . . . .	16
5.4	Google Colab Showing Memory Error . . . . .	16



# Chapter 1

## INTRODUCTION

### 1.1 IMPORTANCE OF SOCIAL MEDIA

Millions of people are using social network sites to express their emotions, opinion and disclose about their daily lives. However, people write anything such as social activities or any comment on products. Through the online communities provide an interactive forum where consumers inform and influence others. Moreover, social media provides an opportunity for business that giving a platform to connect with their customers such as social media to advertise or speak directly to customers for connecting with customer's perspective of products and services. In contrast, consumers have all the power when it comes to what consumers want to see and how consumers respond. With this, the company's success failure is publicly shared and end up with word of mouth. However, the social network can change the behavior and decision making of consumers, for example, 87 percent of internet users are influenced in their purchase and decision by customer's review. So that, if organization can catch up faster on what their customer's think, it would be more beneficial to organize to react on time and come up with a good strategy to compete their competitors

### 1.2 PROBLEMS FACED

Despite the availability of software to extract data regarding a person's sentiment on a specific product or service, organizations and other data workers still face issues

regarding the data extraction.

- Sentiment Analysis of Web Based Applications Focus on Single Tweet Only. With the rapid growth of the World Wide Web, people are using social media such as Twitter which generates big volumes of opinion texts in the form of tweets which is available for the sentiment analysis . This translates to a huge volume of information from a human viewpoint which make it difficult to extract a sentences, read them, analyze tweet by tweet, summarize them and organize them into an understandable format in a timely manner .
- Difficulty of Sentiment Analysis with inappropriate English Informal language refers to the use of colloquialisms and slang in communication, employing the conventions of spoken language such as ‘would not’ and ‘wouldn’t’. Not all systems are able to detect sentiment from use of informal language and this could hanker the analysis and decision making process.

Emoticons, are a pictorial representation of human facial expressions , which in the absence of body language and prosody serve to draw a receiver’s attention to the tenor or temper of a sender’s nominal verbal communication, improving and changing its interpretation . For example, indicates a happy state of mind. Systems currently in place do not have sufficient data to allow them to draw feelings out of the emoticons. As humans often turn to emoticons to properly express what they cannot put into words [6]. Not being able to analyze this puts the organization at a loss. Short-form is widely used even with short message service (SMS). The usage of short-form will be used more frequently on Twitter so as to help to minimize the characters used. This is because Twitter has put a limit on its characters to 140

## **1.3 OBJECTIVE**

The objectives of the study are first, to study the sentiment analysis in microblogging which in view to analyze feedback from a customer of an organization’s product; and second, is to develop a program for customers’ review on a product which allows an organization or individual to sentiment and analyzes a vast amount of tweets into a useful format.

## **1.4 APPROACHES**

### **1.4.1 Lexicon-based Approach**

Lexicon-based methods make use of predefined list of words where each word is associated with a specific sentiment . The lexicon methods vary according to the context in which they were created and involve calculating orientation for a document from the semantic orientation of texts or phrases in the documents . Besides, also states that a lexicon sentiment is to detect word-carrying opinion in the corpus and then to predict opinion expressed in the text.

### **1.4.2 Machine-learning-based Approach**

Machine learning methods often rely on supervised classification approaches where sentiment detection is framed as a binary which are positive and negative . This approach requires labeled data to train classifiers . This approach, it becomes apparent that aspects of the local context of a word need to be taken into account such as negative (e.g. Not beautiful) and intensification (e.g. Very beautiful) .

# Chapter 2

## LITERATURE REVIEW

This section reviews on the important elements in developing the our project "SOCIAL MIRROR" . The first part focuses on LEXICON method make use of predefined list of words where each word is asso-ciated with a specific sentiment . The second part discusses the MACHINE LEARNING METHOD methods often rely on supervised classification approaches where sentiment detection is framed as a binary which are positive and negative.

### 2.1 SOCIAL MEDIA

- Social media have emerged as one of the platforms to raise users' opinions and influence the way any business is commercialized.
- Opinion of people matters a lot to analyze how the propagation of information impacts the lives in a large-scale network like Twitter.
- Sentiment analysis of the tweets determine the polarity and inclination of vast population towards specific topic, item or entity.
- Thus, this project act as a mirror to many employers as well as to many users because of its immense usage



Figure 2.1: Different types of social media platforms

### 2.1.1 ALGORITHM OF LEXICON METHOD

- i. Preprocess each tweet, post by remove punctuation
- ii. Initialize a total polarity score (s) equal 0 -i, s=0
- iii. Check if token is present in a dictionary, then  
 If token is positive, s will be positive (+)  
 If token is negative, s will be negative (-)
- iv. Look at the total polarity score of tweet post  
 If s  $\geq$  threshold, tweet post as positive  
 If s  $\leq$  threshold, tweet post as negative.

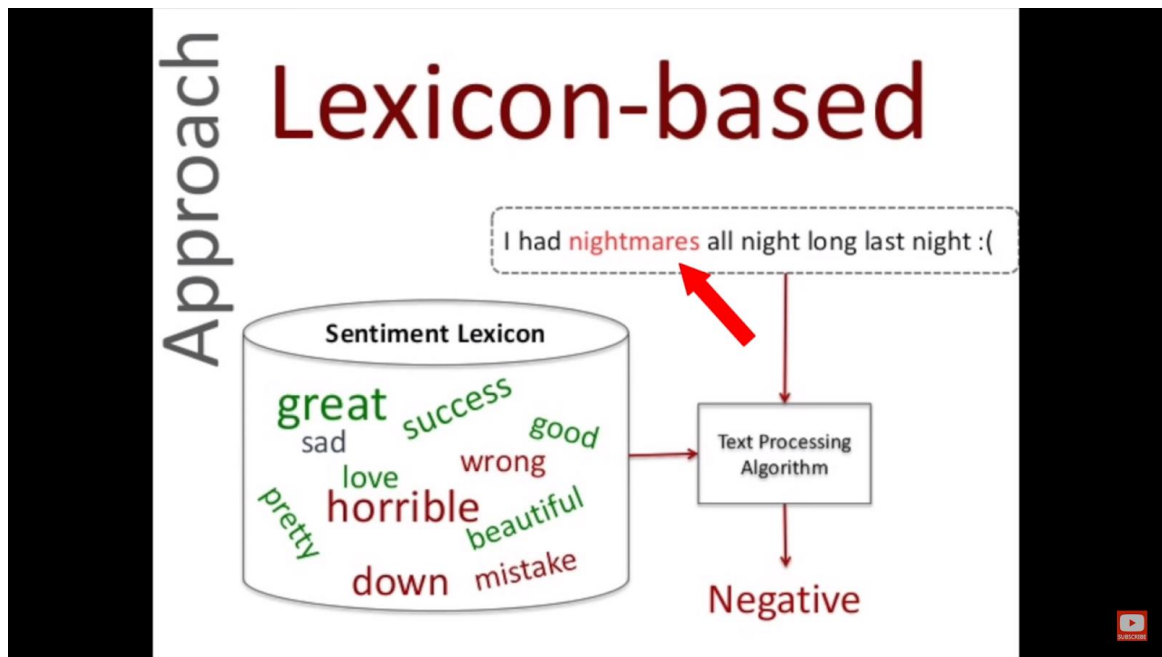


Figure 2.2: Lexicon Approach

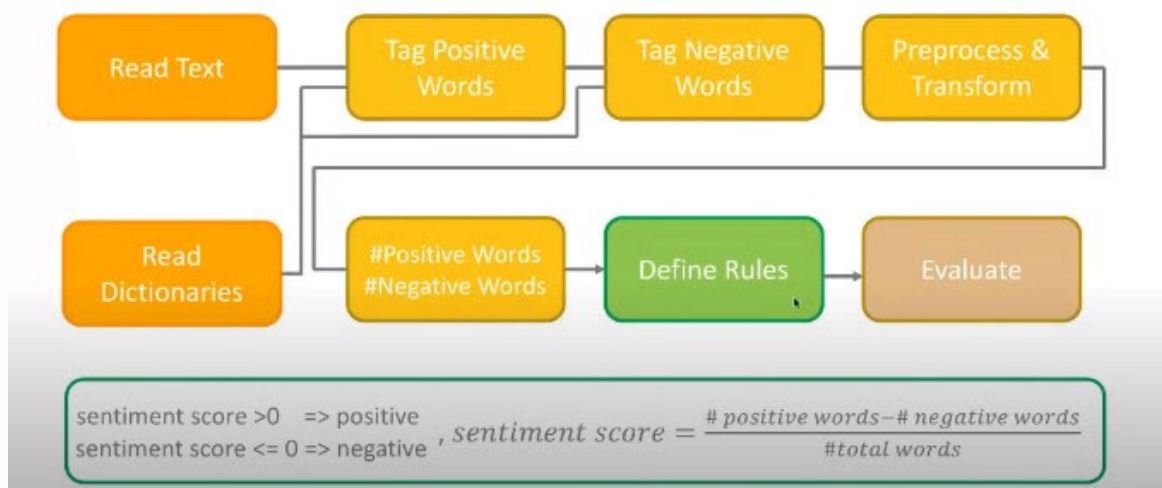


Figure 2.3: Block Diagram of Lexicon Approach

## 2.1.2 ALGORITHM OF MACHINE LEARNING METHOD

- i. Apply a part of speech tagger to each tweet post
- ii. Collect all the adjective for entire tweet posts
- iii. Make a popular word set composed of the top N adjectives
- iv. Navigate all of the tweets in the experimental set to create the following:
  - Number of positive words
  - Number of negative words

- Presence, absence or frequency of each word

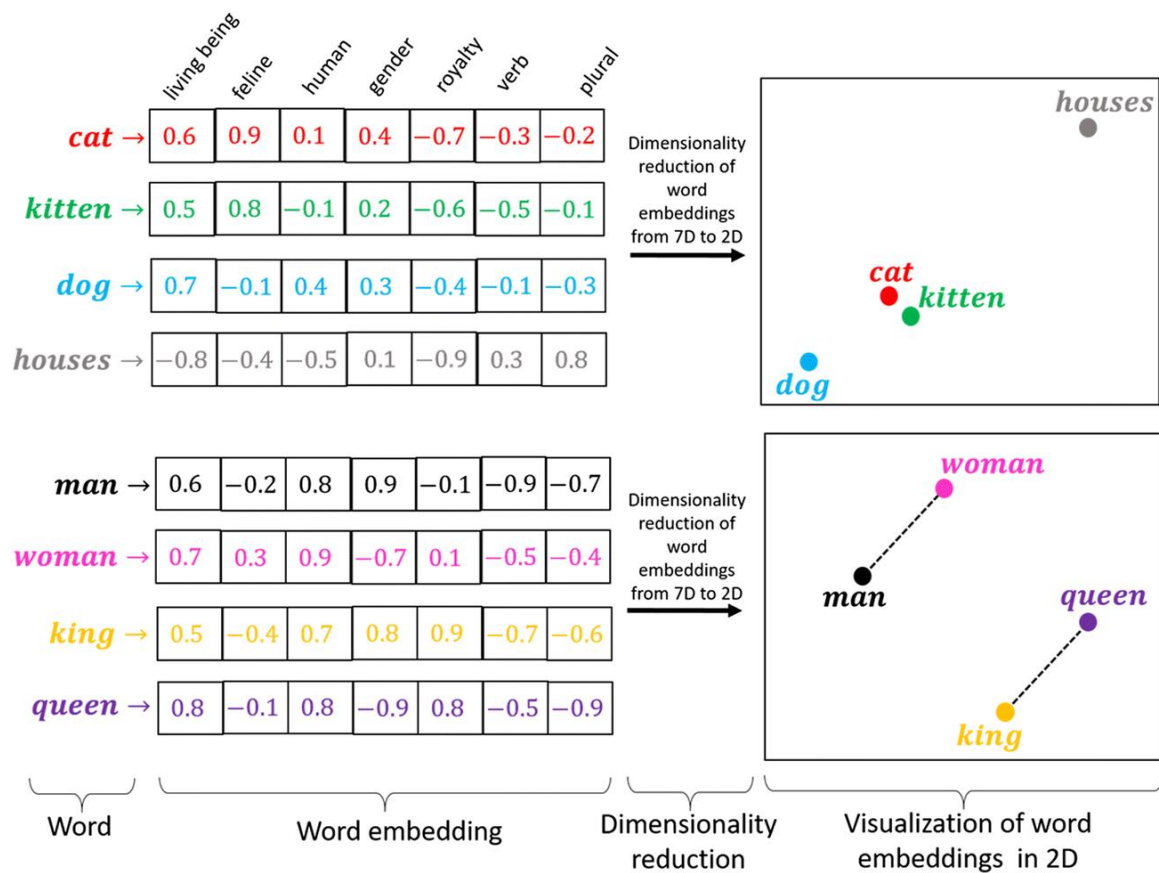


Figure 2.4: Word Embedding

# Chapter 3

## PROJECT DESCRIPTION

- To study the sentiment analysis in microblogging which in view to analyze feedback from a customer of an organization's product
- To develop a program for customers' review on a product which allows an organization or individual to sentiment and analyzes a vast amount of tweets into a useful format.

### 3.1 CREATION OF TWITTER DEVELOPER ACCOUNT

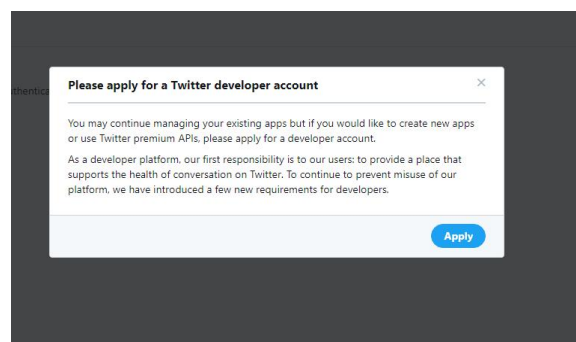


Figure 3.1: Step 1 for creation of developer account



### How will you use the Twitter API or Twitter data?

All fields are required

#### In your words

In English, please describe how you plan to use Twitter data and/or APIs. For students and teachers, please include the name of the school, the name of the instructor and the course number (if available). The more detailed the response, the easier it is to review and approve.

Minor Project is a project one doing graduation needs to make, which he makes in a group of 3 people. Twitter-API will be used by me to do search queries, and then to analyse and to do sentiment analysis. The project we are creating has no commercial objective. It is purely for academic purpose.

Figure 3.2: Step 2 for creation of developer account

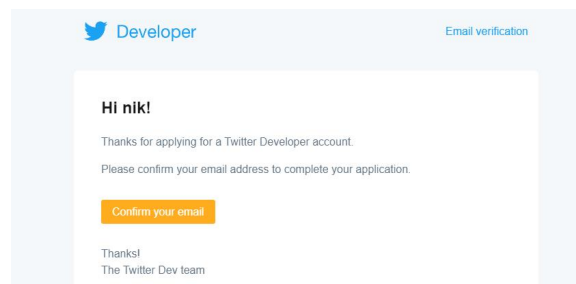


Figure 3.3: Step 3 for creation of developer account

#### Application description (required)

Share a description of your app. This description will be visible to users so this is a good place to tell them what your app does.

Using Machine Learning to analyze tweets. All of the algorithms of Machine Learning will be based on classification problem. Some algorithms will be Neural Network, Logistic Regression.



Between 10 and 200 characters

Figure 3.4: Step 4 successful creation of developer account

## 3.2 BLOCK DIAGRAM

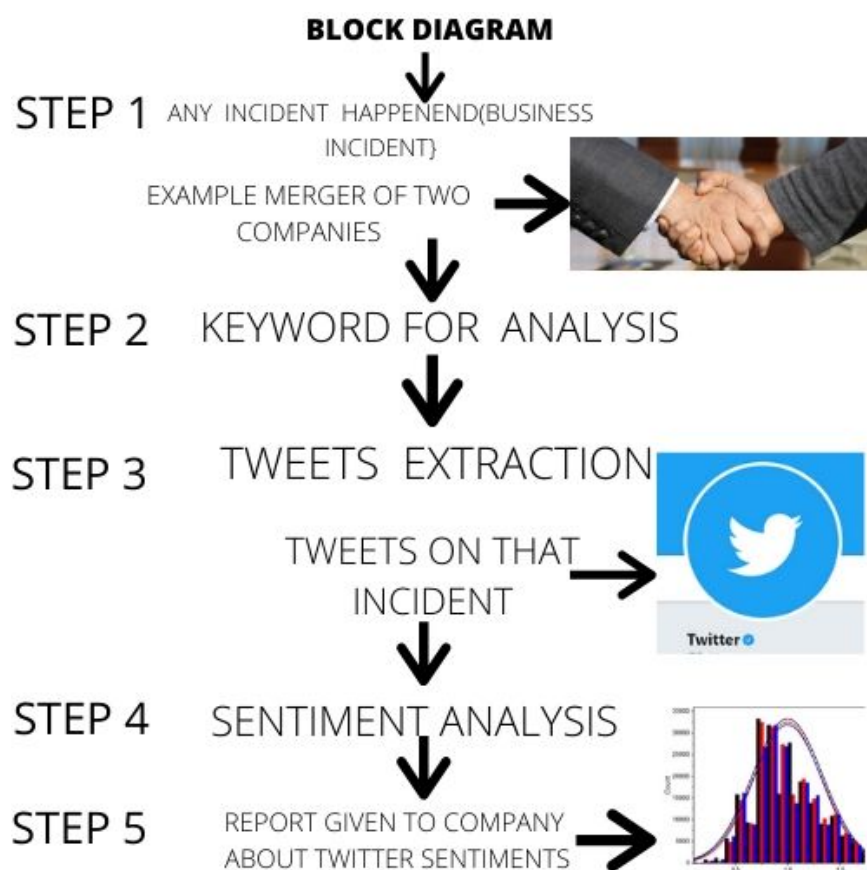


Figure 3.5: Project Description

## 3.3 DESCRIPTION OF BLOCK DIAGRAM

### Step1

The program “Social Mirror” is being developed to analyse the emotions and sentiments, of any incident. That incident could be business or, in general. For eg. a merger a new product launch, or any general, if a government wants to know if public is happy, in the lockdown situation, or should they increase the lockdown duration. Anything, which requires, to know the sentiments of the public, this program can be used live using twitter

### Step 2

For searching for the tweets, the program will need a keyword, for example “Coronavirus Lockdown”, or “Apple smart watch series 5”. After getting the keyword, the program will be able to use twitter’s search functionality to search latest tweets, in past seven days for that keyword.

### Step 3

For that keyword, the program will be extracting tweets from twitter. We have used Twitter API, called “tweepy” to extract tweets. We will also be able to search tweets, in a particular area, and for a particular language.

### Step 4–

The program will be able to perform sentiment analysis, for all the tweets extracted.

The main types of algorithms used include:

Rule-based systems that perform sentiment analysis based on a set of manually crafted rules. Automatic systems that rely on machine learning techniques to learn from data.

We will be using different classification models like - logistic regression, rnn, lstm etc.

A program we have created using Rule- based method is attached in the report itself.

### Step 5–

A report will be made. Report will contain, quantitative analysis of the sentiments of the tweets. It will also contain a word cloud, of the sentiments. A word cloud is a image, having words, in which those words(sentiments) which are highly used are there in big words, and less used sentiments, are used in small size.

## Chapter 4

# MACHINE LEARNING APPROACH

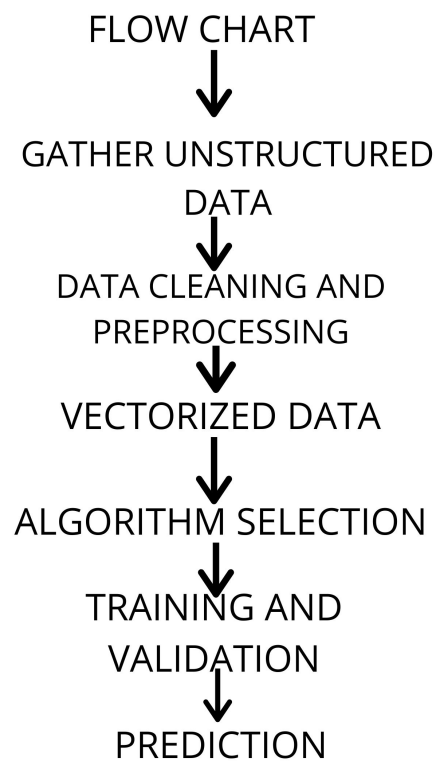


Figure 4.1: Flowchart

## 4.1 DESCRIPTION OF FLOWCHART

### 4.1.1 GATHER UNSTRUCTURED DATA

- We collect data from large databases like kaggle
- Kaggle is an online community of data scientists and machine learners.
- From here tweets are collected

### 4.1.2 DATA CLEANING AND PREPROCESSING

It has following parts

#### 1. STOPWORDS

- Removal of unwanted words are done like single letter words (a ,and ,an)etc

#### 2. STEMMING

- It is the process of reducing inflected words to their word stem, base or root form—generally a written word form.

#### 3.LEMMATISATION

- It is the process of grouping together the inflected forms of word so they can be analysed as a single item, identified by word's lemma, or dictionary form.

### 4.1.3 VECTORIZED DATA

It has following parts

#### 1.TOKENIZATION

- Process of substituting a sensitive data element with a nonsensitive equivalent, referred to as a token It has no extrinsic or exploitable meaning or value. The token is a reference that maps back to the sensitive data through a tokenization system.

#### 2.COUNT VECTORIZER

- Provides a simple way to both tokenize a collection of text documents and builds a vocabulary of known words Encodes new documents using that vocabulary.

Call the fit() function in order to learn a vocabulary from one or more documents

### 3. TF IDF

- In information retrieval, tf-idf or TFIDF, short for term frequency Inverse document frequency, is a numerical statistic that is intended to reflect how important a word is to a document in a collection or corpus

## 4.1.4 ALGORITHM SELECTION

We have used four algorithm techniques

### 1.RANDOM FOREST

- Also called random decision forests
- A type of supervised machine learning algorithm based on ensemble learning
- It is one of the most ensemble classifiers which Combines multiple algorithms of the same type i.e Multiple decision trees, resulting in a forest of trees
- Very fast and robust against over fitting which makes it possible to form as many trees as user wants

### 2.LOGISTIC REGRESSION

- A classification algorithm that is used to assign observations to a discrete set of classes.

Logistic regression transforms its output using the logistic sigmoid function to return a probability value.

### 3.MULTINOMIAL NB

- It is a specialized version of Naive Bayes that is designed more for text documents.
- Simple naive Bayes would model a document as the presence and absence of particular words Multinomial naive bayes explicitly models the word counts and adjusts the underlying calculations to deal with in.

### 4.LSTM

- Long short-term memory (LSTM) is an artificial recurrent neural network (RNN) architecture used in the field of deep learning.LSTM networks are well-suited to classifying, processing and making predictions based on time series data, since there can be lags of unknown duration between important events.

# Chapter 5

## CHALLENGES FACED

- First Challenge, was to list down, and research the forms, and types of sentiment analysis, and different ways, in which it can be done.
- Another challenge was about the data. We did some research, about twitter, and usual behaviour of people, on twitter, and how they communicate. We then took the data from the standard Kaggle data repository and found the data very useful.
- Just to be sure, about the quality of data, we manually checked the labelled data. in batches, and found it to be correct, and good quality
- The steps, of sentiment analysis, are many and varied. To keep a check on the quality, we pre-processed the data, so that the accuracy comes out to be good.
- To apply Deep Neural Network on the tweets, was another big and major step in the pipeline.+ To study and research about twitter API, and new features of twitter API, and to be able to use them
- The biggest challenge amongst all was that the system disk space was running over-time and crashing.
- As a result we tried on google colaboratory but in google colaboratory also RAM was crashing again and again.
- Difficulty in memory allocation due to darge dataset of tweets.

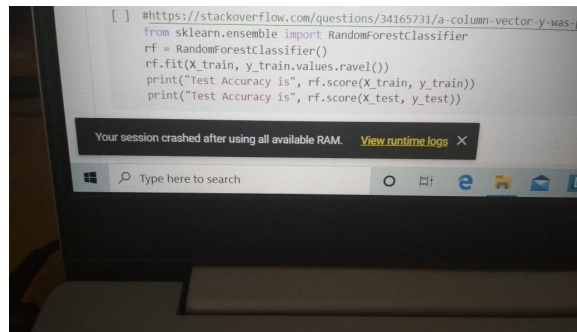


Figure 5.1: Crashing of Ram due to large dataset

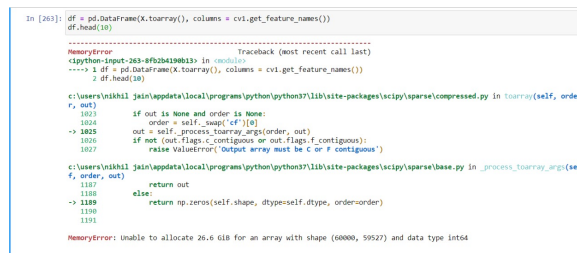


Figure 5.2: Error showing unable to allocate memory

Figure 5.3: Caption



Figure 5.4: Google Colab Showing Memory Error



## Chapter 6

# CONCLUSION AND FUTURE ENHANCEMENTS

### 6.1 CONCLUSION

Twitter is one of the social media that is gaining popularity. Twitter offers organizations a fast and effective way to analyze customers' perspectives toward the critical to success in the market place. Developing a program for sentiment analysis is an approach to be used to computationally which measure customers' perceptions. This report is on the design of a sentiment analysis, extracting a vast amount of tweets

### 6.2 ACCURACY OF DIFFERENT ALGORITHMS

Method	Logistic Regression	Random Forest	Naïve Bayes
<u>CountVectorizer</u>	Training – 95.06% Testing – 76.3%	Training – 99.97% Testing – 72.8%	Training – 93.32% Testing – 75.95%
TF-IDF	Training – 89.55% Testing – 75.15%	Training – 99.97% Testing – 73.45%	Training – 93.32% Testing – 75.95%
Word Embedding	LSTM Training Accuracy- 99.53% Testing Accuracy- 83.09%		

## 6.3 FUTURE ENHANCEMENTS

- The applications of high end sentiment analysis, are huge. It can be used, anywhere, and everywhere. Some applications include - Company review, and share holders, and public's sentiment about the company, and to see the effect on stock prices, and other decisions in the company.
- Future scope, of this project could also, be in determining the fake news, for Eg. Fake News about Covid-19. To determine, public sentiments, for government decision, which is very important now-a-days. For Eg. To check, if people want the Covid-19 pandemic Lockdown to be removed or not, and in which area.
- Certainly, this requires, much more computational power, and data collection resources, and can be done, on a medium scale, if data is collected

# Bibliography

- [1] K. N. Alhayyan et al., "Discovering and Analyzing Important Real-Time Trends in Noisy Twitter Stream" n.p
- [2] J. Ramteke et al., "Election result prediction using Twitter sentiment analysis," in Inventive Computation Technologies (ICICT), International Conference on, 2016, vol. 1, pp. 1–5.
- [3] M. Desai et al., "Techniques for sentiment analysis of Twitter data: A comprehensive survey", 2016 International Conference on Computing, Communication and Automation (ICCCA), 2016.
- [4] A. Pak et al., "Twitter as a corpus for sentiment analysis and opinion mining". In Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10), may 2010.
- [5] R. Mehta et al., "Sentiment analysis and influence tracking using twitter," International Journal of Advanced Research in Computer Science and Electronics Engineering (IJARCSEE), vol. 1, no. 2, p. pp-72, 2012.

# Chapter 7

## APPENDIX

### 7.1 CODE OF LEXICON BASED APPROACH

```
In [20]: pip install nltk
```

```
Requirement already satisfied: nltk in /usr/local/lib/python3.6/dist-packages (3.2.5)  
Requirement already satisfied: six in /usr/local/lib/python3.6/dist-packages (from nltk) (1.12.0)
```

```
In [8]: pip install tweepy
```

```
Requirement already satisfied: tweepy in /usr/local/lib/python3.6/dist-packages (3.6.0)  
Requirement already satisfied: requests-oauthlib>=0.7.0 in /usr/local/lib/python3.6/dist-packages (from tweepy) (1.3.0)  
Requirement already satisfied: six>=1.10.0 in /usr/local/lib/python3.6/dist-packages (from tweepy) (1.12.0)  
Requirement already satisfied: requests>=2.11.1 in /usr/local/lib/python3.6/dist-packages (from tweepy) (2.21.0)  
Requirement already satisfied: PySocks>=1.5.7 in /usr/local/lib/python3.6/dist-packages (from tweepy) (1.7.1)  
Requirement already satisfied: oauthlib>=3.0.0 in /usr/local/lib/python3.6/dist-packages (from requests-oauthlib>=0.7.0->tweepy) (3.1.0)  
Requirement already satisfied: idna<2.9,>=2.5 in /usr/local/lib/python3.6/dist-packages (from requests>=2.11.1->tweepy) (2.8)  
Requirement already satisfied: urllib3<1.25,>=1.21.1 in /usr/local/lib/python3.6/dist-packages (from requests>=2.11.1->tweepy) (1.24.3)  
Requirement already satisfied: certifi>=2017.4.17 in /usr/local/lib/python3.6/dist-packages (from requests>=2.11.1->tweepy) (2019.11.28)  
Requirement already satisfied: chardet<3.1.0,>=3.0.2 in /usr/local/lib/python3.6/dist-packages (from requests>=2.11.1->tweepy) (3.0.4)
```

```
In [9]: pip install textblob
```

```
Requirement already satisfied: textblob in /usr/local/lib/python3.6/dist-packages (0.15.3)  
Requirement already satisfied: nltk>=3.1 in /usr/local/lib/python3.6/dist-packages (from textblob) (3.2.5)  
Requirement already satisfied: six in /usr/local/lib/python3.6/dist-packages (from nltk>=3.1->textblob) (1.12.0)
```

```
In [0]: from textblob import TextBlob
```

```
In [26]: import nltk
nltk.download()
nltk.download('punkt')
```

NLTK Downloader

d) Download l) List u) Update c) Config h) Help q) Quit

Downloader> d

Download which package (l=list; x=cancel)?

Identifier> l

Packages:

```
[ ] abc..... Australian Broadcasting Commission 2006
[ ] alpino..... Alpino Dutch Treebank
[ ] averaged_perceptron_tagger Averaged Perceptron Tagger
[ ] averaged_perceptron_tagger_ru Averaged Perceptron Tagger (Russian)
[ ] basque_grammars..... Grammars for Basque
[ ] biocreative_ppi..... BioCreAtIvE (Critical Assessment of Information
Extraction Systems in Biology)
[ ] bllip_wsj_no_aux.... BLLIP Parser: WSJ Model
[ ] book_grammars..... Grammars from NLTK Book
[ ] brown..... Brown Corpus
[ ] brown_tei..... Brown Corpus (TEI XML Version)
[ ] cess_cat..... CESS-CAT Treebank
[ ] cess_esp..... CESS-ESP Treebank
[ ] chat80..... Chat-80 Data Files
[ ] city_database..... City Database
[ ] cmudict..... The Carnegie Mellon Pronouncing Dictionary (0.6)
[ ] comparative_sentences Comparative Sentence Dataset
[ ] comtrans..... ComTrans Corpus Sample
[ ] conll2000..... CONLL 2000 Chunking Corpus
[ ] conll2002..... CONLL 2002 Named Entity Recognition Corpus
Hit Enter to continue: averaged_perceptron_tagger
[ ] conll2007..... Dependency Treebanks from CoNLL 2007 (Catalan
and Basque Subset)
[ ] crubadan..... Crubadan Corpus
[ ] dependency_treebank. Dependency Parsed Treebank

[ ] dolch..... Dolch Word List
[ ] europarl_raw..... Sample European Parliament Proceedings Parallel
Corpus
[ ] floresta..... Portuguese Treebank
[ ] framenet_v15..... FrameNet 1.5
[ ] framenet_v17..... FrameNet 1.7
[ ] gazetteers..... Gazetteer Lists
[ ] genesis..... Genesis Corpus
[ ] gutenberg..... Project Gutenberg Selections
[ ] ieer..... NIST IE-ER DATA SAMPLE
[ ] inaugural..... C-Span Inaugural Address Corpus
[ ] indian..... Indian Language POS-Tagged Corpus
[ ] jeita..... JEITA Public Morphologically Tagged Corpus (in
ChaSen format)
[ ] kimmo..... PC-KIMMO Data Files
[ ] knbc..... KNB Corpus (Annotated blog corpus)
[ ] large_grammars..... Large context-free and feature-based grammars
for parser comparison
Hit Enter to continue:
[ ] lin_thesaurus..... Lin's Dependency Thesaurus
[ ] mac_morpho..... MAC-MORPHO: Brazilian Portuguese news text with
part-of-speech tags
[ ] machado..... Machado de Assis -- Obra Completa
[ ] masc_tagged..... MASC Tagged Corpus
[ ] maxent_ne_chunker... ACE Named Entity Chunker (Maximum entropy)
[ ] maxent_treebank_pos_tagger Treebank Part of Speech Tagger (Maximum entropy)
[ ] moosesample..... Moses Sample Models
[ ] movie_reviews..... Sentiment Polarity Dataset Version 2.0
[ ] mte_teip5..... MULTEXT-East 1984 annotated corpus 4.0
[ ] mwa_ppdb..... The monolingual word aligner (Sultan et al.
2015) subset of the Paraphrase Database.
[ ] names..... Names Corpus, Version 1.3 (1994-03-29)
[ ] nombank.1.0..... NomBank Corpus 1.0
[ ] nonbreaking_prefixes Non-Breaking Prefixes (Moses Decoder)
[ ] nps_chat..... NPS Chat
[ ] omw..... Open Multilingual Wordnet
[ ] opinion_lexicon..... Opinion Lexicon
[ ] panlex_swadesh..... PanLex Swadesh Corpora
[ ] paradigms..... Paradigm Corpus
[ ] pe08..... Cross-Framework and Cross-Domain Parser
```



```

-----
Evaluation Shared Task
Hit Enter to continue:
[ ] perluniprops..... perluniprops: Index of Unicode Version 7.0.0
                        character properties in Perl
[ ] pil..... The Patient Information Leaflet (PIL) Corpus
[ ] pl196x..... Polish language of the XX century sixties
[ ] porter_test..... Porter Stemmer Test Files
[ ] ppattach..... Prepositional Phrase Attachment Corpus
[ ] problem_reports..... Problem Report Corpus
[ ] product_reviews_1... Product Reviews (5 Products)
[ ] product_reviews_2... Product Reviews (9 Products)
[ ] propbank..... Proposition Bank Corpus 1.0
[ ] pros_cons..... Pros and Cons
[ ] ptb..... Penn Treebank
[ ] qc..... Experimental Data for Question Classification
[ ] reuters..... The Reuters-21578 benchmark corpus, ApteMod
                        version
[ ] rslp..... RSLP Stemmer (Removedor de Sufixos da Lingua
                        Portuguesa)
[ ] rte..... PASCAL RTE Challenges 1, 2, and 3
[ ] sample_grammars..... Sample Grammars
[ ] semcor..... SemCor 3.0
[ ] senseval..... SENSEVAL 2 Corpus: Sense Tagged Text
Hit Enter to continue:
[ ] sentence_polarity... Sentence Polarity Dataset v1.0
[ ] sentiwordnet..... SentiWordNet
[ ] shakespeare..... Shakespeare XML Corpus Sample
[ ] sinica_treebank..... Sinica Treebank Corpus Sample
[ ] smultron..... SMULTRON Corpus Sample
[ ] snowball_data..... Snowball Data
[ ] spanish_grammars.... Grammars for Spanish
[ ] state_union..... C-Span State of the Union Address Corpus
[ ] stopwords..... Stopwords Corpus
[ ] subjectivity..... Subjectivity Dataset v1.0
[ ] swadesh..... Swadesh Wordlists
[ ] switchboard..... Switchboard Corpus Sample
[ ] tagsets..... Help on Tagsets
[ ] timit..... TIMIT Corpus Sample
[ ] toolbox..... Toolbox Sample Files
[ ] treebank..... Penn Treebank Sample
-----

```

```

-----
d) Download  l) List    u) Update  c) Config  h) Help  q) Quit
-----

```

```

Downloader> q
[nltk_data] Downloading package punkt to /root/nltk_data...
[nltk_data] Package punkt is already up-to-date!

```

Out[26]: True

```
In [27]: wiki = TextBlob("Nikhil, Abhishek and Harshit are making this great minor project")
        wiki.tags
```

```
Out[27]: [('Nikhil', 'NNP'),
          ('Abhishek', 'NNP'),
          ('and', 'CC'),
          ('Harshit', 'NNP'),
          ('are', 'VBP'),
          ('making', 'VBG'),
          ('this', 'DT'),
          ('great', 'JJ'),
          ('minor', 'JJ'),
          ('project', 'NN')]

```

```
In [28]: wiki.words
```

```
Out[28]: WordList(['Nikhil', 'Abhishek', 'and', 'Harshit', 'are', 'making', 'this', 'great', 'minor',
                  'project'])

```

```
In [29]: wiki.sentiment.polarity
```

Out[29]: 0.375

```
In [30]: wiki2 = TextBlob("Nikhil, Abhishek and Harshit are making this bad minor project")
        wiki2.tags
```

```
Out[30]: [('Nikhil', 'NNP'),
          ('Abhishek', 'NNP'),

```

```

-----
d) Download  l) List    u) Update  c) Config  h) Help  q) Quit
-----
Downloader> q
[nltk_data] Downloading package punkt to /root/nltk_data...
[nltk_data] Package punkt is already up-to-date!

Out[26]: True

In [27]: wiki = TextBlob("Nikhil, Abhishek and Harshit are making this great minor project")
         wiki.tags

Out[27]: [('Nikhil', 'NNP'),
          ('Abhishek', 'NNP'),
          ('and', 'CC'),
          ('Harshit', 'NNP'),
          ('are', 'VBP'),
          ('making', 'VBG'),
          ('this', 'DT'),
          ('great', 'JJ'),
          ('minor', 'JJ'),
          ('project', 'NN')]

In [28]: wiki.words

Out[28]: WordList(['Nikhil', 'Abhishek', 'and', 'Harshit', 'are', 'making', 'this', 'great', 'minor',
                  'project'])

In [29]: wiki.sentiment.polarity

Out[29]: 0.375

In [30]: wiki2 = TextBlob("Nikhil, Abhishek and Harshit are making this bad minor project")
         wiki2.tags

Out[30]: [('Nikhil', 'NNP'),
          ('Abhishek', 'NNP'),
          ('and', 'CC'),
          ('Harshit', 'NNP'),
          ('are', 'VBP'),
          ('making', 'VBG'),
          ('this', 'DT'),
          ('bad', 'JJ'),
          ('minor', 'NN'),
          ('project', 'NN')]

In [31]: wiki2.sentiment.polarity

Out[31]: -0.37499999999999994

In [0]: import tweepy

In [0]: access_token = '740071211523264512-a4CDoXv6L3ZP8v41QyLnWo2rTW0Rum7'
         access_token_secret = 'Eu3qrJNGL3ocGZN0t7dm3bG4LvwjBxNfLeIsaveqPCmRT'

         api_key = '6CjV0v3F16301uhBKiuUbuHUux'
         api_secret_key = 'r1RxuU15nBRAG7hY3wAf1NeCQgNwa3bGtJgV9fgQDoZD12AQ0x'

In [0]: auth = tweepy.OAuthHandler(api_key, api_secret_key)
         auth.set_access_token(access_token, access_token_secret)

In [0]: api = tweepy.API(auth)

In [36]: public_tweets = api.search('Trump')
         for tweet in public_tweets:
             print(tweet.text)
             analysis = TextBlob(tweet.text)
             print(analysis.sentiment)

@JKNjenga Example in a sentence...china has opened the Pandora box and trump has promised to
close it.
Sentiment(polarity=0.0, subjectivity=0.0)

```

@ASlavitt The original sin was that Trump was elected in the first place...  
 Sentiment(polarity=0.3125, subjectivity=0.5416666666666666)  
 RT @BetteMidler: Donald Trump, you are a bald-faced liar. Dr Fauci said, no, we don't have the tests, and then you undercut him and say we...  
 Sentiment(polarity=0.0, subjectivity=0.0)  
 RT @atensnut: PLEASE READ RESPONSE BY PRES TRUMP'S RESPONSE TO JOE BIDEN..... Below. <https://t.co/ss26DItPfw>  
 Sentiment(polarity=0.0, subjectivity=0.0)  
 RT @SocialPowerOne1: Trump is breaking his own agency's coronavirus guidelines by not self-quarantining <https://t.co/Rd71ssqnyk>  
 Sentiment(polarity=0.6, subjectivity=1.0)  
 only for narcissistic people scapegoats really exist

#covid\_19 #corona #usa #trump #europeanunion  
 Sentiment(polarity=0.1, subjectivity=0.6)  
 RT @deepinmy\_bones: kendall jenner è la classica americana: vota trump segretamente, parla del clima e poi usa il jet per spostarsi di 2 me...  
 Sentiment(polarity=0.0, subjectivity=0.0)  
 RT @Based\_Tsunami: @XFLguru I think the timeline broke when Donald Trump got elected. That shit was not supposed to happen  
 Sentiment(polarity=-0.2, subjectivity=0.8)  
 RT @MattOswaltVA: there will be no better way to describe living under a Trump presidency to our grandchildren than by holding up a a roll...  
 Sentiment(polarity=-0.25, subjectivity=0.5)  
 RT @RealJamesWoods: What is truly stunning is the fact that the Speaker of the House is third in line of presidential succession. If Presid...  
 Sentiment(polarity=0.25, subjectivity=0.5)  
 RT @jimmylittle: Trump wants to travel-ban California and Washington for some fuck-dumb reason.

Let's recruit Oregon and make our own West...  
 Sentiment(polarity=0.4, subjectivity=0.55)  
 Trump's go-it-alone approach put to the test by global coronavirus pandemic <https://t.co/dA1RfnYunn>  
 Sentiment(polarity=0.0, subjectivity=0.0)  
 RT @DavidCornDC: Trump said virtually nothing about how to handle the public health crisis underway "in" the United States. This is primary...

---