**VIVEKANAND EDUCATION SOCIETY'S
INSTITUTE OF TECHNOLOGY**

Department of Computer Engineering

Project Report on

Twitter Sentiment Analysis

In partial fulfillment of the Final Year, Bachelor of Engineering (B.E.) Degree in Computer Engineering at the University of Mumbai Academic Year 2020-2021.

Submitted by

Rahul Motwani D17B 39

Naveed Sarguroh D17B 56

Ankush Shetty D17B 58

Project Mentor

Mrs.Vidya Zope

(2020-2021)

VIVEKANAND EDUCATION SOCIETY'S INSTITUTE OF TECHNOLOGY
Department of Computer Engineering

Certificate

This is to certify that **Rahul Motwani, Naveed Sarguroh, Ankush Shetty** of Final Year Computer Engineering studying under the University of Mumbai have satisfactorily completed the mini project on "**Twitter Sentiment Analysis**" as a part of their coursework of Mini Project for Semester-VIII under the guidance of their mentor **Mrs. Vidya Zope** in the year 2020-2021.

This mini project report entitled (**Twitter Sentiment Analysis**) by (**Rahul Motwani, Naveed Sarguroh, Ankush Shetty**) is approved for the degree of _____ (**NLP**).

Programme Outcomes	Grade
PO1,PO2,PO3,PO4,PO5,PO6,PO7, PO8, PO9, PO10, PO11, PO12 PSO1, PSO2	

Date:

Project Guide: Mrs. Vidya Zope

INDEX

CHAPTER 1: INTRODUCTION

1.1 Problem Definition

Sentiment analysis is the technique of capturing the emotion behind the text. It applies natural language processing (NLP) and machine learning to detect, extract, and study people's or customers' perceptions about a product, topic, news, etc. In our project we aimed to analyse the sentiment of the tweets of various Twitter users. Such analysis is helpful for understanding the sentiment of various Twitter users towards various topics, and also for understanding whether the general consensus towards something is positive or negative. This can then be used to analyse and visualise the sentiments of a particular person's latest tweets for better understanding of the data.

In this way, a particular person's sentiment towards recent topics or issues can be understood by analysing their tweets' sentiments. By doing so for a number of users, the sentiment analysis would help to understand the common consensus and sentiment at a particular point of time.

1.2 Scope of the project

1. Projects like these have a wide variety of scope when it comes to analysing reviews regarding products or services.
2. Social unrest regarding issues can be avoided by detecting negative sentiment of Twitter users in advance.
3. Similar platforms can be built for various other social media platforms to get more analysis.
4. A person's opinion regarding certain products or current issues can be understood by analysing his or her most recent tweets.
5. The project can further be used to compare sentiments of different users.

1.3 Users and their requirements

Users of this system could be:

1. Various media organisations trying to understand sentiments regarding current affairs

2. Government trying to avoid social unrest by detecting high negative sentiments on current affairs.

The requirements of the users include:

1. An interface where they can enter the username and get sentiments regarding the latest tweets of that user.

2. An interface where the sentiments of the entered username are summarised and visualised with the help of graphs.

1.4 Technologies to be used

- Python 3.6
- TextBlob
- Matplotlib
- Flask Framework
- Pandas
- Natural Language Toolkit (nltk)
- Tweepy

CHAPTER 2: LITERATURE SURVEY

2.1 Literature Survey

1. **A Study on Sentiment Analysis Techniques of Twitter Data Abdullah Alsaeedi1 , Mohammad Zubair Khan2 Department of Computer Science, College of Computer Science and Engineering Taibah University Madinah, KSA**

   Abstract -- This paper is a comparison of all the known methods and their working and accuracy. These models used either the twitter API or datasets of old existing tweets. Twitter is an enormously popular microblog on which clients may voice their opinions. In recent years, researchers in the field of sentiment analysis have been concerned with analyzing opinions on different topics such as movies, commercial products, and daily societal issues. Opinion investigation of Twitter data is a field that has been given much attention over the last decade and involves dissecting "tweets" (comments) and the content of these expressions. As such, this paper explores the various sentiment analysis applied to Twitter data and their outcomes.

2. **A. Sarlan, C. Nadam and S. Basri, "Twitter sentiment analysis," Proceedings of the 6th International Conference on Information Technology and Multimedia, Putrajaya, Malaysia, pp. 212-216, doi: 10.1109/ICIMU.2014.7066632.**

   Abstract -- This paper first discusses the two main approaches for extracting sentiment automatically which are the lexicon-based approach and machine-learning-based approach. The model will categorize sentiment into positive and negative (and null), which is represented in a pie chart and html page. 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 marketplace. 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. However, the program has planned to develop on a web application system, but due to

limitations of Django which can be worked on a Linux server or LAMP, further this approach needs to be done.

3. **S. A. El Rahman, F. A. AlOtaibi and W. A. AlShehri, "Sentiment Analysis of Twitter Data," 2019 International Conference on Computer and Information Sciences (ICCIS), 2019, pp. 1-4, doi: 10.1109/ICCISci.2019.8716464.**

Abstract -- The aim of this paper was to present a model that can perform sentiment analysis of real data collected from Twitter. Data in Twitter is highly unstructured which makes it difficult to analyze. However, the model proposed here is different from prior work in this field because it combined the use of supervised and unsupervised machine learning algorithms. The process of performing sentiment analysis done here was follows: Tweet extracted directly from Twitter API. This was then followed by cleaning and discovery of data that was performed. After that the data was fed into several models for the purpose of training. Each tweet was classified based on its sentiment whether it is a positive, negative or neutral. Data was collected on two subjects, McDonalds and KFC to show which restaurant has more popularity. Different machine learning algorithms were used. The result from these models were tested using various testing metrics like cross validation and f-score. Moreover, this model demonstrates strong performance on mining texts extracted directly from Twitter. For validation and testing, several metrics were used such as recall, precision and f score.

4. **M. S. Neethu and R. Rajasree, "Sentiment analysis in twitter using machine learning techniques," 2013 Fourth International Conference on Computing, Communications and Networking Technologies (ICCCNT), 2013, pp. 1-5, doi: 10.1109/ICCCNT.2013.6726818.**

Abstract :- This paper uses machine learning techniques for general sentiment analysis. They are using 1200 tweets as their dataset which they collected from a period of 1 month. After the dataset is created, the features are extracted in two parts. The first part extracts information from tweets which are in the form of emoticons and hashtags. The second part extracts information from the text of the tweet. For text preparation, they are

using the unigram approach which results in collection of words. Then they created a list of positive keywords, negative keywords and a list of different words that represent negation. The various keywords generated while list creation are not treated equally. So to overcome that, the authors are selecting a special character from the tweet. If a relevant part of speech can be determined for a keyword, then that is taken as a special keyword. Otherwise a keyword is selected randomly from the available keywords as a special keyword. Thus the feature vector is composed of 8 relevant features. The 8 features used are part of speech (pos) tag, special keyword, presence of negation, emoticon, number of positive keywords, number of negative keywords, number of positive hashtags and number of negative hashtags.

5. **Vishal.A.Kharde, Prof. Sheetal.Sonawane, "Sentiment Analysis of Twitter Data: A Survey of Techniques", International Journal of Computer Applications (2016)**

Abstract -- Sentiment analysis can be defined as a process that automates mining of attitudes, opinions, views and emotions from text, speech, tweets and database sources through Natural Language Processing (NLP). Sentiment analysis involves classifying opinions in text into categories like "positive" or "negative" or "neutral". Preprocessing of the dataset here includes removing all URLs, hashtags, targets, stopwords, punctuations, symbols, numbers and also all non-english tweets. Also, all acronyms are replaced with their sentiments and the acronyms are expanded. This entire process is then followed by feature extraction. Part of speech tags, words and their frequencies, opinion words and frequencies, position terms, etc are extracted. Finally various machine algorithms such as Naive Bayes, Maximum Entropy, Support Vector Machine are applied or lexicon based approaches like dictionary based or corpus based are used.

CHAPTER 3: CONCEPTUAL SYSTEM DESIGN
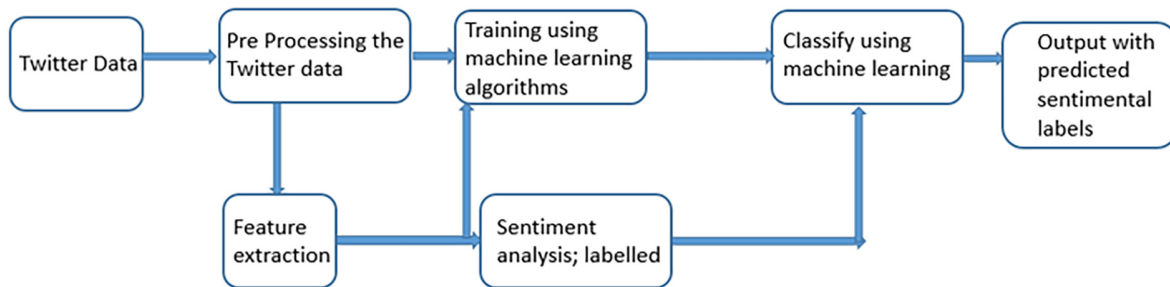
## 3.1 Conceptual system diagram

Fig 1.Conceptual Diagram

The above diagram shows the conceptual design of our system. The first step includes loading the dataset to our system. The dataset we have used was based on a review dataset which had 3 labels in it, i.e. positive, neutral and negative. It includes 4 attributes: text, polarity, score and subjectivity.

The next step includes preprocessing of data using various methods of Lemmatization, removal of stop words and cleaning the data thoroughly so that it is ready for processing. After successfully preprocessing the data, we proceed to train our model.
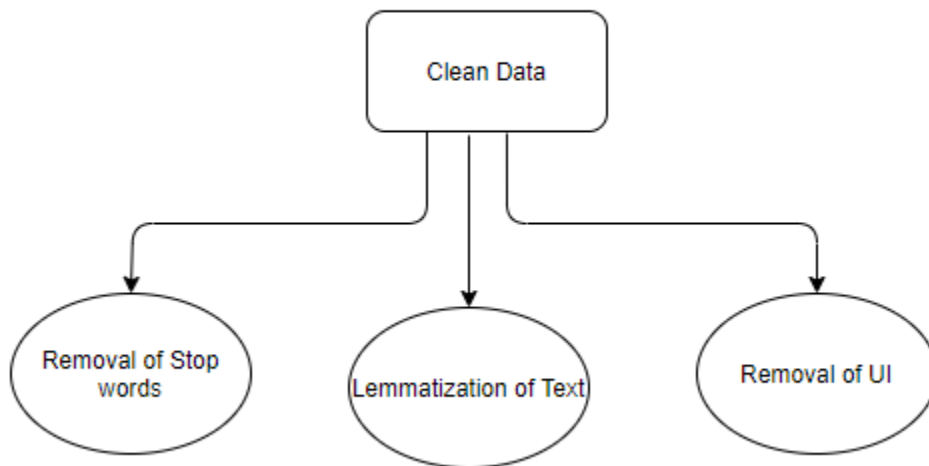
Modelling is the process of representing the patterns in an effective manner and then using the model and data for visualization. After that, we carry out successful testing for the data we have fed and obtain results**.**
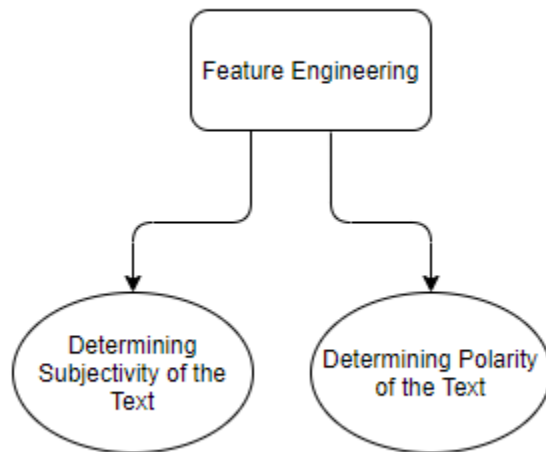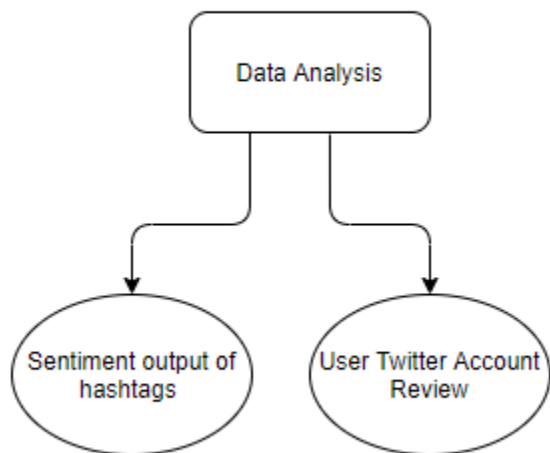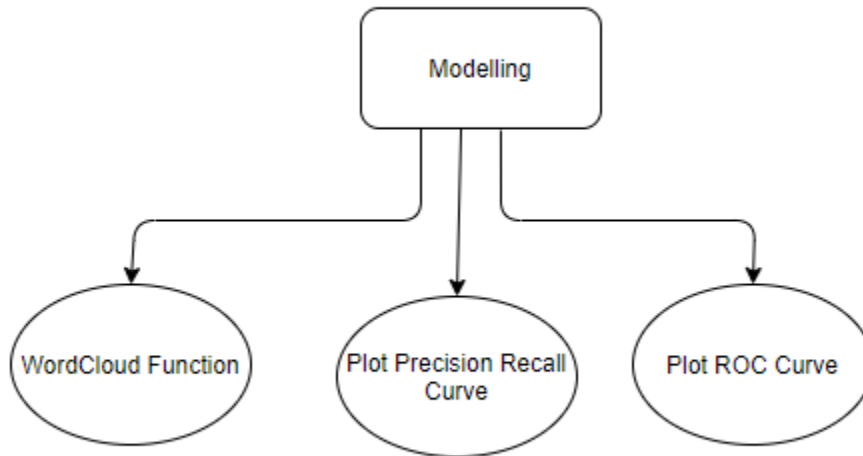
3.2 Design of Modules in Detail

1.Load Data
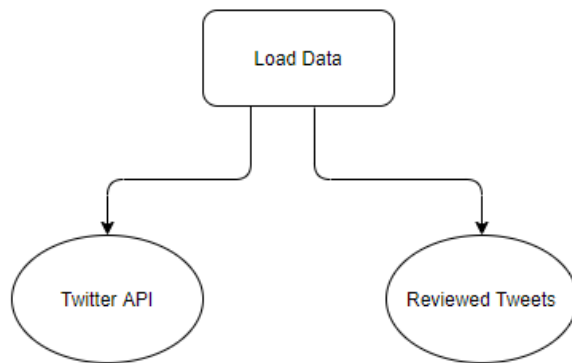
2.Clean Data

3.Feature Engineering

```
                    ┌──────────────────────┐
                    │ Feature Engineering  │
                    └──────────────────────┘
                        │              │
                        ▼              ▼
            ╭───────────────╮   ╭───────────────────╮
            │  Determining  │   │ Determining Polarity │
            │ Subjectivity of the │ │   of the Text     │
            │     Text      │   │                   │
            ╰───────────────╯   ╰───────────────────╯
```

4.Data Analysis

```
                    ┌──────────────────────┐
                    │    Data Analysis     │
                    └──────────────────────┘
                        │              │
                        ▼              ▼
            ╭───────────────╮   ╭───────────────────╮
            │ Sentiment output of │ │ User Twitter Account │
            │    hashtags    │   │      Review        │
            ╰───────────────╯   ╰───────────────────╯
```

5.Modelling Reviews

CHAPTER 4: IMPLEMENTATION AND EVALUATION

4.1 Implementation

Determining Subjectivity and Polarity of tweets

```
In [10]: print(df.head(20))

                                         Tweet  Subjectivity  Polarity
0   Incredible effort by the team to pull that off...     0.500000  0.450000
1   From listening to you coach me on the field, t...     0.000000  0.000000
2   Thank you guys for your kind wishes. Let's win...     0.650000  0.700000
3   Our reefs are the heart and souls of our ocean...     0.550000  0.500000
4   The other cause that's extremely close to my h...     0.638889 -0.204861
5   It's ElClásico time! \n\nVery excited for the ...     0.687500  0.043750
6   Yesterday when I walked on to the field it was...     0.450000  0.050000
7            The CAUSES I support ❤ Stay tuned! 😊     0.000000  0.000000
8   Team effort isse kehte hai. Just like our   te...     0.000000  0.000000
9   Hello Paltan! We're getting closer to the open...     0.300000  1.000000
10  Another day another set 😠Lights, camera ..... ...     0.000000  0.000000
11  Wishing all  fans a happy &amp; safe Holi.\n\n...     0.750000  0.650000
12  Hitman ko lagta hai ki yeh anthem superhit hai...     0.000000  0.000000
13  Outstanding presentation of skills and mindset...     0.637500  0.650000
14  Great finish to a fantastic test series. We no...     0.825000  0.600000
15  This is us ringing Pujara's door bell and scam...     0.000000  0.000000
16  Good win for team. Huge compliments to  for re...     0.633333  0.633333
17               Welcome to the city of heroics!     0.900000  1.000000
18                   Great work so far  👍😆     0.875000  0.450000
19  Fingers seem to be fine. Sammy and I giving ma...     0.500000  0.114583
```

User Interface:

Sentiment Analyzer   Home About

Topic related results

[Click Here to classify another piece of news]

WHY DO WE NEED THIS APP?

To know about our inspiration, kindly refer this documentation

KNOW MORE

To know more about in depth implementation, kindly refer this link which explains the methodology behind each feature.

USEFUL LINKS

About Us

Have a feedback?

Help

CONTACT

🏠 Mumbai, Maharashtra, India

✉️ info@gmail.com

📞 + 01 234 567 88

4.2 Evaluation

WordCloud:

```
In [17]: words = ' '.join([tweet for tweet in df['Tweet']])
         wordCloud = WordCloud(width=600, height=400).generate(words)

         plt.imshow(wordCloud)
         plt.show()
```



```
In [34]: import seaborn as sns

         for x in [0,1]:
             if x<0:
                 label="Negative"
             elif x==0:
                 label="Neutral"
             else:
                 label="Positive"

         sns.distplot(df['Polarity'], hist=True, label=label)
```

```
c:\users\ankus\appdata\local\programs\python\python37\lib\site-packages\seaborn\distributions.py:2557: FutureWarning: `distplot
` is a deprecated function and will be removed in a future version. Please adapt your code to use either `displot` (a figure-le
vel function with similar flexibility) or `histplot` (an axes-level function for histograms).
  warnings.warn(msg, FutureWarning)
c:\users\ankus\appdata\local\programs\python\python37\lib\site-packages\seaborn\distributions.py:2557: FutureWarning: `distplot
` is a deprecated function and will be removed in a future version. Please adapt your code to use either `displot` (a figure-le
vel function with similar flexibility) or `histplot` (an axes-level function for histograms).
  warnings.warn(msg, FutureWarning)
```
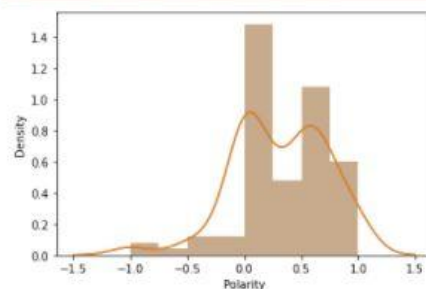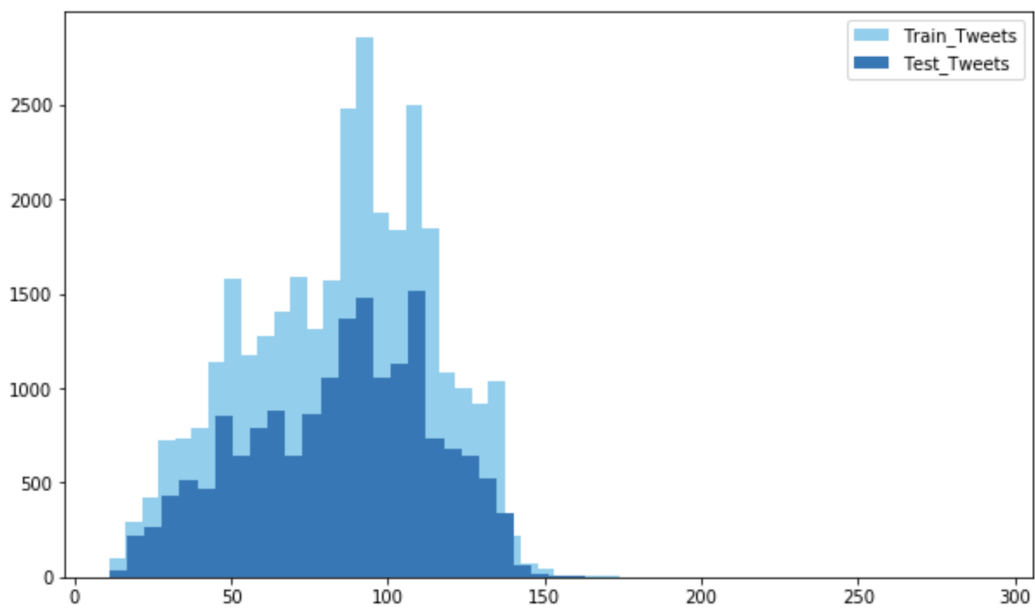
ROC Curve:
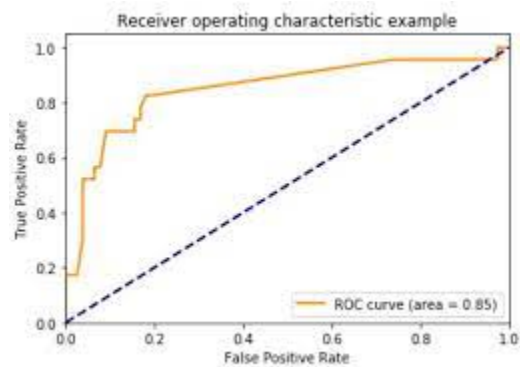
Precision Recall Curve:

```
length_train = train['tweet'].str.len()
length_test = test['tweet'].str.len()
plt.figure(figsize=(10,6))
plt.hist(length_train, bins=50, label="Train_Tweets", color = "skyblue" )
plt.hist(length_test, bins=50, label="Test_Tweets")
plt.legend()
```

<matplotlib.legend.Legend at 0x21853d3aa48>

CHAPTER 5: CONCLUSION AND FUTURE SCOPE

5.1 Conclusion

The popularity of social media is now at an all time high. People voice their opinions on these platforms and this becomes very important for various organisations and other parties involved. The analysis of sentiments of these opinions voiced on social media is a very critical task for a variety of reasons. This can be to keep a check on social unrest, to understand sentiment of users regarding certain products or issues and also to understand the general consensus regarding a variety of ongoing issues.

The project we have implemented is specifically built for analysing Twitter data. Twitter being a social media platform where the main method of interaction is through tweets which are in the form of text, the voicing of opinions and reviews is the maximum here. A particular person's sentiment towards recent topics or issues can be understood by analysing their tweets' sentiments. By doing so for a number of users, the sentiment analysis would help to understand the common consensus and sentiment at a particular point of time. This proves to be vital for a lot of stakeholders like various media outlets, product manufacturers and sometimes even the government.

5.2 Future Scope

The project developed here is specific to a certain social media platform, that is Twitter. The same system cannot be used for overall analysis of sentiments over a number of social media platforms. Thus, there is a scope of improvement in the sense that either a new system can be built for each of the other platforms like Facebook, Instagram etc which can then be aggregated to find overall sentiment across various platforms, or a single system can be built that works for different social media platforms at the same time.

At the same time with regards to Twitter itself, the project can be further extended to obtain sentiments pertaining to specific words or trends by obtaining general tweets from all over the

world instead of doing that for an individual user's tweets. There can also be a higher level of visualisation of these sentiments with more complex graphs taking into account more variables like frequency of tweets and intensity of the words used, etc.

CHAPTER 6. REFERENCES

1. A Study on Sentiment Analysis Techniques of Twitter Data Abdullah Alsaeedi1 , Mohammad Zubair Khan2 Department of Computer Science, College of Computer Science and Engineering Taibah University Madinah, KSA

2. A. Sarlan, C. Nadam and S. Basri, "Twitter sentiment analysis," Proceedings of the 6th International Conference on Information Technology and Multimedia, Putrajaya, Malaysia, pp. 212-216, doi: 10.1109/ICIMU.2014.7066632.

3. S. A. El Rahman, F. A. AlOtaibi and W. A. AlShehri, "Sentiment Analysis of Twitter Data," 2019 International Conference on Computer and Information Sciences (ICCIS), 2019, pp. 1-4, doi: 10.1109/ICCISci.2019.8716464.

4. M. S. Neethu and R. Rajasree, "Sentiment analysis in twitter using machine learning techniques," 2013 Fourth International Conference on Computing, Communications and Networking Technologies (ICCCNT), 2013, pp. 1-5, doi: 10.1109/ICCCNT.2013.6726818.

5. Vishal.A.Kharde, Prof. Sheetal.Sonawane, "Sentiment Analysis of Twitter Data: A Survey of Techniques", International Journal of Computer Applications (2016)

6. https://monkeylearn.com/blog/sentiment-analysis-of-twitter/

7. https://towardsdatascience.com/twitter-sentiment-analysis-classification-using-nltk-python-fa912578614c

8. https://www.analyticsvidhya.com/blog/2018/07/hands-on-sentiment-analysis-dataset-python/