

<b>TABLE OF CONTENTS</b>	<b>Page Number</b>
<i>Certificate</i>	<i>i</i>
<i>Acknowledgement</i>	<i>ii</i>
Abstract	1
<b>Chapter 1: INTRODUCTION</b>	
1.1 Project Outline	2
1.2 Tools/ Platform	2
1.3 Introduction	2
1.4 Packages	3
<b>Chapter 2: MATERIALS AND METHODS</b>	
2.1 Description	7
2.2 Take Input	7
2.3 Encode	7
2.4 Generate QR Code	7
2.5 Decode and Display	7
<b>Chapter 3: RESULT</b>	
3.1 Output	8
<b>Chapter 4: CONCLUSION</b>	9
<b>References</b>	10

## **ABSTRACT**

The objective of our project is to analyze the sentiment of twitter. Social media websites 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. These days, the applications of such analysis can be easily observed during public elections, movie promotions, brand endorsements and many other fields. In this project, we used search API to extract saved tweets and perform sentiment analysis. The primary aim is to provide a method for analyzing sentiment score in noisy twitter streams. This project on the design of a sentiment analysis, extracting vast number of tweets. Results classify user's perception via tweets into positive and negative.

# **CHAPTER 1: INTRODUCTION**

## **1.1 Project Outline**

TITLE OF THE PROJECT: - TWITTER SENTIMENT ANALYSIS

## **1.2 Tools / Platform**

- 1 Operating System :- WINDOWS 8
- 2 Language :- PYTHON
- 3 Software Used :- PYCHARM

## **1.3 Introduction**

As internet is growing bigger, its horizons are becoming wider. Social Media and Micro blogging platforms like Facebook, Twitter, Tumblr dominate in spreading encapsulated news and trending topics across the globe at a rapid pace. A topic becomes trending if more and more users are contributing their opinion and judgments, thereby making it a valuable source of online perception. These topics generally intended to spread awareness or to promote public figures, political campaigns during elections, product endorsements and entertainment like movies, award shows. Large organizations and firms take advantage of people's feedback to improve their products and services which further help in enhancing marketing strategies. One such example can be leaking the pictures of upcoming iPhone to create a hype to extract people's emotions and market the product before its release. Thus, there is a huge potential of discovering and analyzing interesting patterns from the infinite social media data for business-driven applications.

Sentiment analysis is the prediction of emotions in a word, sentence or corpus of documents. It is intended to serve as an application to understand the attitudes, opinions and emotions expressed within an online mention. The intention is to gain an overview of the wider public opinion behind certain topics. Precisely, it is a paradigm of categorizing conversations into positive, negative or neutral labels. Many people use social media sites for networking with other people and to stay up-to-date with news and current events. These sites (Twitter, Facebook, Instagram, google+) offer a platform to people to voice their opinions. For example, people quickly post their reviews online as soon as they watch a movie and then start a series of comments to discuss about the acting skills depicted in the movie. This kind of information forms a basis for people to evaluate, rate about the performance of not only any movie but about other products and to know about whether it will be a success or not. This type of vast information on these

sites can be used for marketing and social studies. Therefore, sentiment analysis has wide applications and includes emotion mining, polarity, classification and influence analysis.

Twitter is an online networking site driven by tweets which are 140 character limited messages. Thus, the character limit enforces the use of hashtags for text classification. Currently around 6500 tweets are published per second, which results in approximately 561.6 million tweets per day. These streams of tweets are generally noisy reflecting multitopic, changing attitudes information in unfiltered and unstructured format. Twitter sentiment analysis involves the use of natural language processing to extract, identify and characterize the sentiment content. Sentiment Analysis is often carried out at two levels 1) coarse level and 2) fine level. In coarse level, the analysis of entire documents is done while in fine level, the analysis of attributes is done. The sentiments present in the text are of two types: Direct and Comparative. In comparative sentiments, the comparison of objects in the same sentence is involved while in direct sentiments, objects are independent of one another in the same sentence.

However, doing the analysis of tweets expressed is not an easy job. A lot of challenges are involved in terms of tonality, polarity, lexicon and grammar of the tweets. They tend to be highly unstructured and non-grammatical. It gets difficult to interpret their meaning. Moreover, extensive usage of slang words, acronyms and out of vocabulary words are quite common while tweeting online.

Following python modules have been used in the program:

- tweepy
- tkinter
- datetime
- TextBlob
- Matplotlib

The functions that are used are as follows:

- Description
- Registering App
- Accessing Data
- Storing Data
- Preparing Data

## 1.4 MODULES

### 1. tweepy

Tweepy is open-sourced, hosted on GitHub and enables Python to communicate with Twitter platform and use its API. Tweepy supports accessing Twitter via Basic Authentication and the newer method, OAuth. Twitter has stopped accepting Basic Authentication so OAuth is now the only way to use the Twitter API. The main difference between Basic and OAuth authentication are the consumer and access keys. With Basic Authentication, it was possible to provide a username and password and access the API, but since 2010 when the Twitter started requiring OAuth, the process is a bit more complicated. An app has to be created at dev.twitter.com. OAuth is a bit more complicated initially than Basic Auth, since it requires more effort, but the benefits it offers are very lucrative:

Tweets can be customized to have a string which identifies the app which was used.

It doesn't reveal user password, making it more secure.

It's easier to manage the permissions, for example a set of tokens and keys can be generated that only allows reading from the timelines, so in case someone obtains those credentials, he/she won't be able to write or send direct messages, minimizing the risk.

The application doesn't rely on a password, so even if the user changes it, the application will still work.

After logging in to the portal, and going to "Applications", a new application can be created which will provide the needed data for communicating with Twitter API.

### 2. tkinter

Python offers multiple options for developing GUI (Graphical User Interface). Out of all the GUI methods, tkinter is most commonly used method. It is a standard Python interface to the Tk GUI toolkit shipped with Python. Python with tkinter outputs the fastest and easiest way to create the GUI applications. Creating a GUI using tkinter is an easy task.

To create a tkinter:

1. Importing the module – tkinter
2. Create the main window (container)
3. Add any number of widgets to the main window

#### 4. Apply the event Trigger on the widgets.

Importing tkinter is same as importing any other module in the python code. There are two main methods used you the user need to remember while creating the Python application with GUI.

**Tk(screenName=None, baseName=None, className='Tk', useTk=1):** To create a main window, tkinter offers a method 'Tk(screenName=None, baseName=None, className='Tk', useTk=1)'. To change the name of the window, you can change the className to the desired one. The basic code used to create the main window of the application is:

`m=tkinter.Tk()` where m is the name of the main window object

**mainloop():** There is a method known by the name mainloop() is used when you are ready for the application to run. mainloop() is an infinite loop used to run the application, wait for an event to occur and process the event till the window is not closed.

`m.mainloop()`

### 3. TextBlob

TextBlob is a Python (2 and 3) library for processing textual data. It provides a simple API for diving into common natural language processing (NLP) tasks such as part-of-speech tagging, noun phrase extraction, sentiment analysis, classification, translation, and more.

### 4. Matplotlib

matplotlib.pyplot is a collection of command style functions that make matplotlib work like MATLAB. Each pyplot function makes some change to a figure: e.g., creates a figure, creates a plotting area in a figure, plots some lines in a plotting area, decorates the plot with labels, etc. In matplotlib.pyplot various states are preserved across function calls, so that it keeps track of things like the current figure and plotting area, and the plotting functions are directed to the current axes (please note that “axes” here and in most places in the documentation refers to the axes part of a figure and not the strict mathematical term for more than one axis).

## CHAPTER 2: MATERIALS AND METHODS

## 2.1 Description:-

To be able to access Twitter data programmatically we need to create and register an app on twitter developers website for authentication and thereafter we can access data by using Twitter API.

## 2.2 Registering App:-

To register the twitter app, we need to create a new app <https://apps.twitter.com/>. On registering the app we will receive consumer\_key and consumer\_secret\_key. Next, From the configuration page of the app, we will get access\_token and access\_token\_secret, which will be used to get access to twitter on behalf of our application. We must keep these authentication tokens private as they can be misused. Best practice is to create a separate config file and keep these tokens.

## 2.3. Accessing Data :-

Twitter provides REST API's to connect with their service. We will use one python library to access the twitter REST API's called Tweepy. It provides wrapper methods to easily access twitter REST API. to install Tweepy we can use below command.

```
pip install tweepy
```

## 2.4 Storing Data:-

Now we will access all tweet data from personal profile and store it for our analysis steps. Tweepy library provides simple cursor interface to iterate through all the tweets and store them in file.

## 2.5 Preparing Data:-

Before we begin to analyze the twitter data, it's important to understand the structure of the tweet as well as pre-process the data to remove non-useful terms called stop-words. Preprocessing of data in data analysis is the very important step. Preprocessing is in the simple term means to take in the data and prepare the data for optimal output considering our requirement.

**Tokenizing** the Tweet Tokenization is one of the most basic, yet most important, steps in text analysis. The purpose of tokenization is to split a stream of text into smaller units called tokens, usually words or

phrases. We will use python NLTK library to tokenize the tweets. Even NLTK library needs some preprocessing steps to correctly tokenize @mentions and #hashtags. We use regular expressions to provide exceptions for mentions and hashtags.

**Removing Stop-Words** Stop-word removal is one important step that should be considered during the pre-processing stages. Stop-words are most popular and common words of any language. While their use in the language is crucial, they don't usually convey a particular meaning, especially if taken out of context. This is the case of articles, conjunctions, some adverbs, etc. which are commonly called stop-words. Some libraries provide default stop-words for different languages. NLTK library provides default stop-words for English language.

## CHAPTER 3: RESULT



### **3.1 OUTPUT:-**

## **CHAPTER 4: CONCLUSION**

In this project, we started with very basics of Twitter data analysis. We explained for twitter app authentication using OAuth and Tweepy. Then we explained steps to collect historical data as well as streaming data. We then preprocessed the data using tokenizers. In the final step, we tried to execute a number of use cases to analyze the stored data. We represented results of analyzing most used terms for a data set, most used hashtags, most used mentions of user accounts on twitter and we also represented the bigrams i.e. two terms used frequently in our dataset.

This project is introductory in nature and hence deals with basics of twitter data analysis using python. In future work, we will try to represent more advanced data analysis patterns decision making with more accurate results.

## **REFERENCES**

1. J He, W Shen, P Divakaruni, L Wynter, R Lawrence, “Improving Traffic Prediction with Tweet Semantics”, Proceedings of the Twenty-Third International Joint Conference on Artificial Intelligence, pp. 1387–1393, August 3-9 2013.
2. A. Agarwal, B. Xie, I Vovsha, O. Rambow, R. Passonneau “Sentiment Analysis of Twitter Data” In the proceedings of Workshop on Language in Social Media, ACL, 2011
3. S Kumar, F Morstatter, H Liu, “Twitter Data Analytics” Springer Book 2013
4. A Mittal, A Goel, “Stock Prediction Using Twitter Sentiment Analysis”, Stanford University, 2011
5. “D Ediger, K Jiang, J Riedy, D. A. Bader “Massive Social Network Analysis: Mining Twitter for Social Good”, 39th International Conference on Parallel Processing 2010, pp. 583-593
6. <https://github.com/vivekwisdom/TwitterAnalysisApp>, Code repository of the sample application