**MINI PROJECT REPORT**

**on**

**Twitter Sentiment Analysis**

**(CSE VI Semester Mini Project)**

**2021-2022**

****

**Submitted to: Submitted by:**

Mr. Saumitro Siddharth verma

(CC-CSE-J-VI-Sem) Roll. No.: 1918726

CSE-J-VI-Sem

Session: 2021-22

**DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING**

**GRAPHIC ERA HILL UNVERSITY, DEHRADUN**

**TABLE OF CONTENTS**

1. Introduction 1

1.1 Context 1

1.2 Motivation 1

1.3 Idea 2

2. Tweepy 4

2.1 Twitter API 4

2.2 Tweepy 4

2.3 Creating Twitter API Authentication Credentials 5

3. TextBlob 6

3.1 About 6

3.2 Features 6

3.3 NLTK 7

4. Overview 8

4.1 Diagram 8

4.2 workflow 9

5. References 10

**1.Introduction**

**1.1 Context**

This project has been done as a part of my course for the B.tech at grapgic era hill university .I had three months to fullfill the requirements in order to succeed the Project. This project of analyzing sentiments of tweets comes under the domain of “Pattern Classification” and “Data Mining”. Both of these terms are very closely related and intertwined, and they can be formally defined as the process of discovering “useful” patterns in large set of data, either automatically (unsupervised) or semiautomatically (supervised)

**1.2 Motivations**

Being extremely interested in everything having a relation with the Machine Learning, the independant project was a great occasion to give me the time to learn and confirm my interest for this field. The fact that we can make estimations, predictions and give the ability for machines to learn by themselves is both powerful and limitless in term of application possibilities. We can use Machine Learning in Finance, Medicine, almost everywhere. That’s why I decided to conduct my project around the Machine Learning.

**1.3 Idea**

Millions of tweets get posted per day and their hastags can be used see the performance of a particular asset in market.This project was motivated by my desire to investigate the sentiment analysis field of machine learning since it allows to approach natural language processing which is a very hot topic actually. using tweets and try to figure out which is positive or negative.

**2. Tweepy**

2.1 Twitter API

The Twitter API gives developers access to most of Twitter’s functionality. You can use the API to read and write information related to Twitter entities such as tweets, users, and trends.

Technically, the API exposes dozens of HTTP endpoints related to:

* Tweets
* Retweets
* Likes
* Direct messages
* Favorites
* Trends
* Media

Tweepy, as we’ll see later, provides a way to invoke those HTTP endpoints without dealing with low-level details.

The Twitter API uses [OAuth](https://oauth.net/), a widely used open authorization protocol, to authenticate all the requests. Before making any call to the Twitter API, you need to create and configure your authentication credentials. Later in this article, you’ll find detailed instructions for this.

You can leverage the Twitter API to build different kinds of automations, such as bots, analytics, and other tools. Keep in mind that Twitter imposes certain restrictions and policies about what you can and cannot build using its API. This is done to guarantee users a good experience. The development of tools to spam, mislead users, and so on is forbidden.

The Twitter API also imposes **rate limits** about how frequently you’re allowed to invoke API methods. If you exceed these limits, you’ll have to wait between 5 and 15 minutes to be able to use the API again. You must consider this while designing and implementing bots to avoid unnecessary waits.

You can find more information about the Twitter API’s policies and limits in its official documentation:

* [Twitter Automation](https://help.twitter.com/en/rules-and-policies/twitter-automation)
* [Rate limits](https://developer.twitter.com/en/docs/basics/rate-limits)

2.2 Tweepy

[**Tweepy**](https://github.com/tweepy/tweepy) is an open source Python package that gives you a very convenient way to access the Twitter API with Python. Tweepy includes a set of classes and methods that represent Twitter’s models and API endpoints, and it transparently handles various implementation details, such as:

* Data encoding and decoding
* HTTP requests
* Results pagination
* OAuth authentication
* Rate limits
* Streams

If you weren’t using Tweepy, then you would have to deal with low-level details having to do with HTTP requests, data serialization, authentication, and rate limits. This could be time consuming and prone to error. Instead, thanks to Tweepy, you can focus on the functionality you want to build.

Almost all the functionality provided by Twitter API can be used through Tweepy. The only current limitation, as of version 3.7.0, is that Direct Messages don’t work properly due to some recent changes in the Twitter API.

### 2.3 Creating Twitter API Authentication Credentials

As we have previously seen, the Twitter API requires that all requests use OAuth to authenticate. So you need to create the required authentication credentials to be able to use the API. These credentials are four text strings:

1. Consumer key
2. Consumer secret
3. Access token
4. Access secret

If you already have a Twitter user account, then follow these steps to create the key, token, and secrets. Otherwise, you have to sign up as a Twitter user before proceeding.

3 TextBlob

3.1 About

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. When computing a sentiment for a single word, TextBlob employs the “averaging” technique, which is applied to polarity values to calculate a polarity score for a single word, and thus a similar procedure applies to every single word, resulting in a combined polarity for larger texts.

3.2 Features

Noun phrase extraction

Part-of-speech tagging

Sentiment analysis

Classification (Naive Bayes, Decision Tree)

Tokenization (splitting text into words and sentences)

Word and phrase frequencies

Parsing

n-grams

Word inflection (pluralization and singularization) and lemmatization

Spelling correction

Add new models or languages through extensions

WordNet integration

3.3 NLTK

NLTK is a leading platform for building Python programs to work with human language data. It provides easy-to-use interfaces to over 50 corpora and lexical resources such as WordNet, along with a suite of text processing libraries for classification, tokenization, stemming, tagging, parsing, and semantic reasoning, wrappers for industrial-strength NLP libraries, and an active discussion forum.

**4 Overview**

4.1 workflow diagram

Twitter database

Authentication tweet-extraction

Tweepy

Set polarity

TextBlob

Tokenized text

tweets

4.2 Workflow

financial data services provider carried out a study which concluded that from December 2011 to November 2013, companies with positive social media sentiment had their stock prices rise by 76% compared to -14% from negative sentiment stocks. Paying attention to social media can pay well on Wall Street.From this we can assume that how crucial these social media platform plays role in market manipulation.

Twitter has become one of the most used social media platform . People use the hashtag symbol (#) before a relevant keyword or phrase in their Tweet to categorize those Tweets and help them show more easily in Twitter search. Twitter allows use the database to extract the tweets of particular hastags by setting up the developer account.

After setting up the tweeter developer account by using different keys provided by the twitter we can extract tweets form twitter database and by asking user what hastag and how many tweets should be extracted for particular analysis.

Once the tweets gets extracted we tokenize the tweets and by providing these tokens to textblob easily the polarity can be checked.TextBlob assign polarity of 1 -1 or 0 based on positive,negative or neutral behaviour of particular word.

These polarity of each word can be used to calculate overall polarity of tweete and these tweets can be used used to determine the performace or current behaviour of particular asset e.g.-$BTC,$ETH,$SNAP,etc**REFERENCES**

1. How NLP works-

https://www.youtube.com/watch?v=CMrHM8a3hqw

1. Model of sentiment analysis–

https://www.youtube.com/watch?v=1gQ6uG5Ujiw&ab\_channel=freeCodeCamp.org

1. Extract tweets from Tweepy

https://www.askpython.com/python/examples/extracting-tweets-using-twitter-api

1. Twitter REST API

https://dev.twitter.com/docs/api

1. Twitter Sentiment, an online application performing sentiment classification of Twitter.

http://twittersentiment.appspot.com/

1. Introduction to textblob-

https://textblob.readthedocs.io/en/dev/quickstart.html