CLOUD MICROSERVICES PROJECT REPORT

TWEETILYSE

(Twitter Sentiment Analysis)

Submitted in partial fulfillment of the requirements for the degree of

Bachelor of Engineering

in

Computer Science and Business System

Submitted By

Abhimanyu Raghuvanshi 102018017

Himanshi Pahuja 102018009

Divija Arora 102018056

Paras Thapar 102018014

Under the supervision of :

_____

Dr. Sumana Maiti

Designation of Supervisors: Assistant professor

COMPUTER SCIENCE AND ENGINEERING DEPARTMENT

THAPAR INSTITUTE OF ENGINEERING AND TECHNOLOGY

PATIALA – 147004

December 2022

Table of Contents

1. Introduction

A Twitter sentiment analysis identifies negative, positive, or neutral emotions within the text of a tweet. It is a text analysis using natural language processing (NLP) and machine learning. It identifies and extracts subjective information from original data, providing a company with a better understanding of the social sentiment of its brand, product, or service. At the same time, analyze the online conversations of customers.

Sentiment analysis is frequently used to analyze customer feedback, survey responses, and product reviews.

Some relevant areas of sentiment analysis are monitoring the activities in existing communities within social networks, reputation management, and customer experience.

What is Twitter Sentiment Analysis?

A Twitter sentiment analysis is the process of determining the emotional tone behind a series of words. A sentiment analysis tool is an automated technique that extracts meaningful customer information related to their attitudes, emotions, and opinions.

The classification of customer conversations around a brand is crucial and can follow the lineaments listed below:

1. The most relevant characteristics of a brand's product or service for customers.
2. The customer behavior and intentions around those brand characteristics.

These characteristics are an essential tool for analyzing brand reputation among customers and their conversations concerning the brand with a more human approach.

1.2 Problem Statement

The main goal of the project is to perform sentimental analysis on the tweets of a particular user and i.e. determine whether the sentiments/feelings associated with a particular tweet are positive,

negative, or neutral. Also, to perform various kinds of graphical analysis in the data i.e. subjectivity, no, of likes, retweets, etc.

2. Background

Social media have received more attention nowadays. Public and private opinion about a wide variety of subjects is 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 critical success in the marketplace. Developing a program for sentiment analysis is an approach to be used to computationally measure customers' perceptions.

This paper reports on the design of sentiment analysis, extracting a vast amount of tweets. Python is used in this development along with various modules such as Tweepy, NumPy, pandas, and Textblob. Results classify customers' perspectives via tweets into positive and negative, which is represented in a pie chart and tabular form.

2.1 Proposed Layout

We will be following the steps given below:

1. Import Libraries
2. Tweets Mining
3. Data Cleaning
4. Tweets Processing
5. Data Exploration
6. Sentiment Analysis

2.2 Motivation

We have chosen to work with Twitter since we feel it is a better approximation of public sentiment as opposed to conventional internet articles and web blogs. The reason is that the amount of

relevant data is much larger for Twitter, as compared to traditional blogging sites. Moreover, the response on Twitter is more prompt and also more general (since the number of users who tweet is substantially more than those who write web blogs on a daily basis). Sentiment analysis of the public is highly critical in macro-scale socioeconomic phenomena like predicting the stock market rate of a particular firm. This could be done by analyzing overall public sentiment towards that firm with respect to time and using economics tools for finding the correlation between public sentiment and the firm's stock market value. Firms can also estimate how well their product is responding in the market, which areas of the market is it having a favorable response, and in which a negative response (since Twitter allows us to download streams of geo-tagged tweets for particular locations. If firms can get this information they can analyze the reasons behind the geographically differentiated responses, and so they can market their product in a more optimized manner by looking for appropriate solutions like creating suitable market segments. Predicting the results of popular political elections and polls is also an emerging application to sentiment analysis.

3. Objectives

The main objectives of this project are:

- To analyze the crucial feelings related to the varied tweets and
- Obtain and basic graphical analysis of assorted tweets. This could be useful in the opinion of an outsized quantity of individuals.
- To generate a word cloud of the most used words.
- To display a graph between subjectivity and polarity

4. Methodology

Methods of Sentiment Analysis:

- Information assortment Consumers sometimes specific their sentiments on public forums just as blogs, discussion boards, and product reviews also as on personal logs – Social network sites like Facebook and Twitter. Sentiments are asserted in numerous methods, with totally contrasting terminology and meaning of writing, creating the info immense and scrambled.

Manual analysis of sentiment information is nearly not possible. Therefore, special programming languages like ‘R’ are accustomed to methods and analyzing the info.

- Text Preparation Preparing the text is however anyway separating the extricated data before the examination. It includes differentiating and deleting non-text content and content that is extraneous to the world of study from the information that is present.

1. Analyzing the sentiment: At this stage, every sentence of the review is checked for sound judgment basis. Sentences with subjective expressions are kept and tweets that convey objective expressions are discarded.

2. Classification Sentiments are broadly speaking classified into 2 things i.e positive and negative. At this stage of sentiment analysis methodology, every tweet detected is assessed as positive, negative, and neutral.

3. Results The main plan of sentiment analysis is to convert JSON text into a pandas data frame. Upon completing the project, we will visualize the data using matplotlib and will display the sentiments.

5. Observations and Findings

An accurate text processing data library text blob was used to carry out a sentiment analysis of tweets using tweepy and wordCloud. The accuracy of the model and its output of sentiment were found to be good.

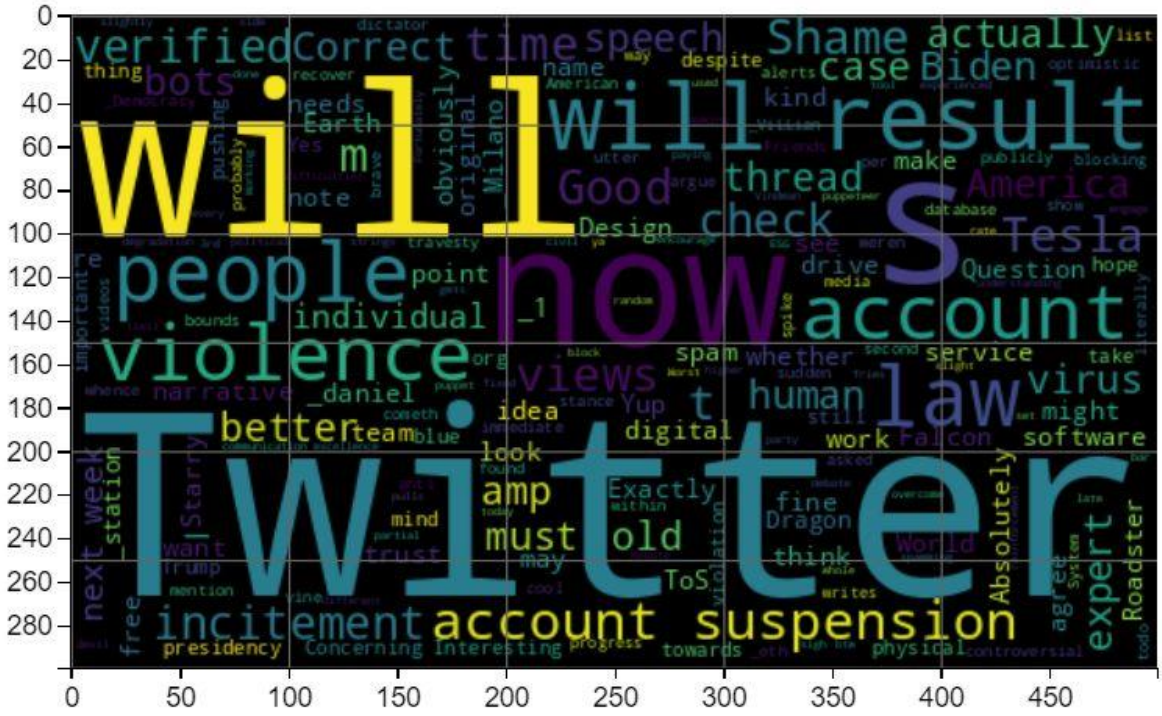
The final output of the Twitter sentiment analysis is the categorization of tweets into positive, negative, and neutral along with some charts and graphs between subjectivity and polarity for the graphical reference of the analysis.

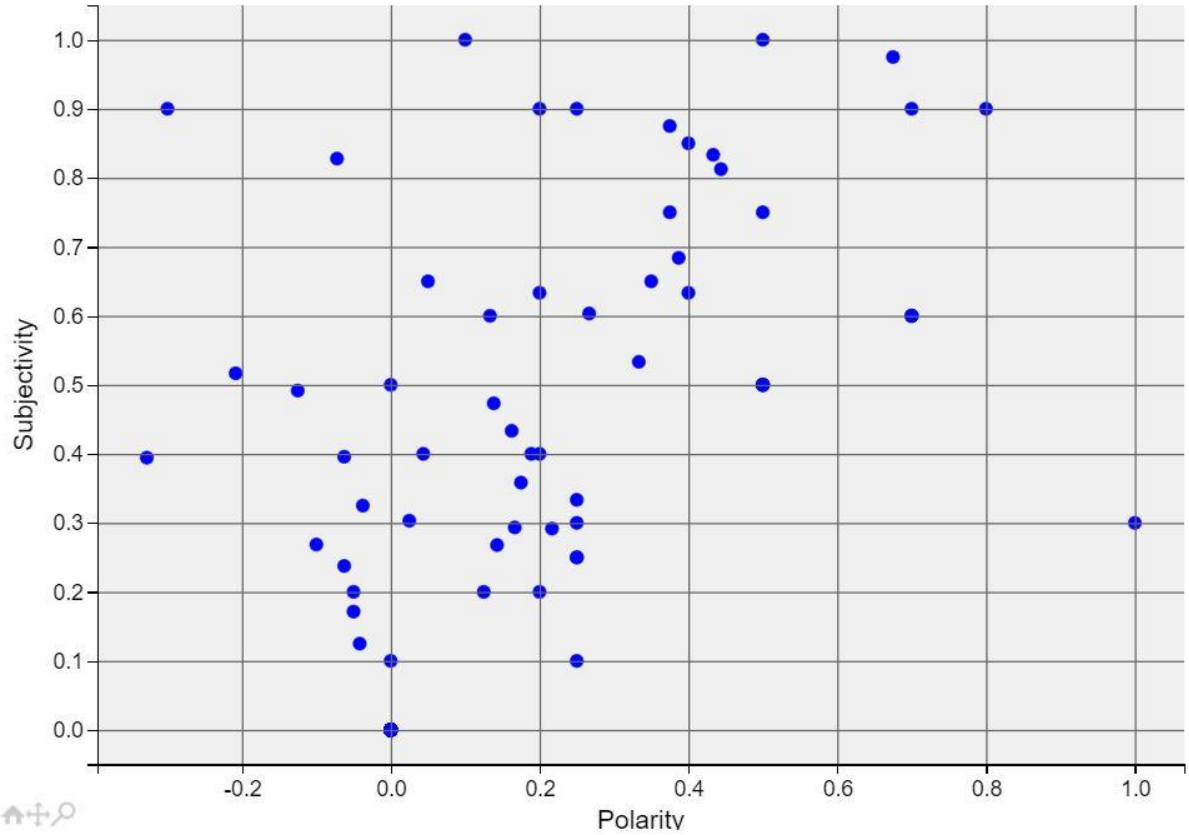
SENTIMENT ANALYSIS

POSITIVE

NEUTRAL

NEGATIVE





6. Limitations

Some of the limitations where the project could fail are:

- Sentiment analysis of short texts such as single sentences is challenging because of the limited contextual information that they normally contain.
- Since sentiment analysis needs textual data, tweets including videos, images, memes, emojis, or quotes-based posters will not be able to get analyzed by our Twitter sentiment analyzer.
- The sentiment analyzer works only with tweets in the English language. Tweets of any other languages will not be analyzed.
- As we have used the TextBlob library for processing textual data which is a Lexicon-based sentiment analyzer. It has some predefined rules or we can say word and weight dictionary, where it has some scores that help to calculate a sentence's polarity because of which it

has a limit to the type of words it can analyze. So in cases where the vocabulary used in the tweets is very rare or uncommon, it becomes difficult to analyze those tweets.

7. Conclusions and Future Work

Nowadays, sentiment analysis or opinion mining could be a hot topic in machine learning. We have still so much to find regarding the feelings of the corpus of texts terribly accurately thanks to the complexity of the English language.

In this project, we are tending to specialize in sentiment analysis. There is a capability of labor within the range of sentiment analysis with a slightly accepted background. For eg., we tend to see that clients, as a rule, utilize our site for explicit sorts of watchwords which can be partitioned into a couple of particular classes, to be specific: governmental issues/lawmakers, big names, items/brands, sports person, media and music.

Subsequently, we will attempt to perform a separate feeling investigation on tweets that exclusively have a place with 1 of those categories (for example the training data wouldn't be general anyway explicit to 1 of those classifications) and analyze the outcomes we tend to get if we apply general sentiment analysis on that instead.

Twitter's API is vastly helpful in data processing applications and may offer large insights into the general public opinion. Twitter API can be used in most of the difficult sentiment gathering, involving people, trends, and social graphs that are very different for the human mind to get. We can expand our project to other languages as well.

8. Bibliography/References

https://towardsdatascience.com/my-absolute-go-to-for-sentiment-analysis-textblob-3ac3a11d524 https://textblob.readthedocs.io/en/dev/ https://amueller.github.io/word_cloud/

https://pandas.pydata.org/docs/ https://matplotlib.org/stable/index.html

https://www.analyticsvidhya.com/blog/2022/07/sentiment-analysis-using-python/