A

Project Report

On

**“Twitter Sentiment Analysis”**

For the Course

**Intro. To Data Science.**

**Submitted by**

Mahad Naveed

<https://whatthemahad.com/>

**Introduction:**

This project will be useful in analyzing the sentiments i.e., positivity and negativity from the provided tweets. Sentiment analysis refers to the use of natural language processing, text analysis, computational linguistics, and biometrics to systematically identify, extract, quantify, and study affective states and subjective information.

**Scope:**

**Sentiment analysis** (or opinion mining) is a natural language processing technique **used to** determine whether data is positive, negative, or neutral. **Sentiment analysis** is often performed on textual data to help businesses monitor brand and product **sentiment** in customer feedback and understand customer needs.

This analyze can have multiple uses such as:

* Marketing Strategy
* More Insightful Data
* Measurement of Marketing Campaign
* In terms of politics, find the like and dislike ratio of the candidates.

Trends in Twitter and other Social Media are very interesting to follow. Following the trends and analyzing them can clearly presents a picture of what people are really up-to. This analyze can be beneficial for marketing strategies and campaigns and other useful purposes

**Development Tools**

* Python
* Google Collab
* Octaparse
* Twitter

**Methods/Algorithms:**

* Data Pre-Processing
* Data Cleaning
* Sentiment Analysis using Textblob
* Graphical Representation of Data

**Dataset:**

<https://twitter.com/whatthemahad>

Dataset also included in Drive Folder.

**Source Code:**

<https://colab.research.google.com/drive/1a5lysvsLVGPwSSRaXp9VL9y-mi_Auc3A?usp=sharing>

**Libraries:**

* Textblob
* csv
* Wordcloud
* Pandas
* NumPy
* Re
* Matplotlib.pyplot

**Implementation:**

1. **Data Pre-Processing**

* **Tweets Scrape**

For this project, [personal Twitter](https://twitter.com/whatthemahad) handle was used to extract tweets dataset. The tool used for extraction of Tweets is [Octaparse](https://www.octoparse.com/).Graphical user interface, application, table, Excel

Description automatically generated

* **Data Import in Python Program**

Tweets from the file was read by Python library CSV.

**Graphical user interface, text, application, email

Description automatically generated**

* **Data frame**

A data frame is two-dimensional tabular sized table data structure. Showing default Data Frame method to recent first 5 tweets in tabular form.

Graphical user interface, text, application

Description automatically generated

1. **Data Cleaning**

As tweets have mentions, retweets and redirecting links to different platforms, tweets does not make good sense inorder to find sentiment.

To overcome, tweets data has been cleaned to remove any:

* **Mentions**
* **Retweets**
* **Hashtags**
* **Hyperlink**

Graphical user interface, text, application, email

Description automatically generated

1. **Sentiment Analysis using 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.

* **Subjectivity** **and Polarity:**

The sentiment function of **textblob** returns two properties, **polarity**, and **subjectivity**. ... **Subjective** sentences generally refer to opinion, emotion, or judgment whereas objective refers to information. **Subjectivity** is also a float which lies in the range of [0,1].

**Polarity** is float which lies in the range of [-1,1] where 1 means positive statement and -1 means a negative statement.

Graphical user interface, text, application

Description automatically generated

* **Wordcloud:**

Visual representation of text data typically used to depict keyword metadata.

**Graphical user interface, text, application

Description automatically generated**

* **Showing Positivity and Negativity of Tweets:**

**Graphical user interface, text, application, email

Description automatically generated**

1. **Graphical Representation of Data:**

Representing graph between Subjectivity, Polarity.

The tweets on right side are positive and negative on left side. And in mid are neutral.

**A picture containing graphical user interface

Description automatically generated**

* **Count of Positive, Negative Tweets:**

Positive Tweets: 14 = 27.5

Negative Tweets: 10 = 19.6%

Neutral Tweets: 27 = 52%

**Graphical user interface, application

Description automatically generated**

* **Final Output of Positive, Negative Tweets:**

Now the result has been shown in the following graph which represents the number of Positive, Negative and Neutral tweets which were used as input.

**Graphical user interface, application

Description automatically generated**