Title: Automated System for Sentimental Analysis of Streaming Data

A Project Report Submitted
in Fulfilment of the Requirements
for the Course of
Major Project - I
In
Fourth year – Seventh Semester of
Bachelor of Technology
Specialization
In
Big Data

Under

Dr. Shaurya Gupta

By

500067177	**R172218064**	**Lakshay Vasuja**
500068760	**R172218063**	**Lakshay Sharma**
500068095	**R172218045**	**Divyansh Chandna**
500068030	**R172218055**	**Janvi Arya**

DEPARTMENT OF INFORMATICS
SCHOOL OF COMPUTER SCIENCE
UNIVERSITY OF PETROLEUM AND ENERGY STUDIES, BIDHOLI,
DEHRADUN, UTTRAKHAND, INDIA

October, 2021

TABLE OF CONTENT

1. Introduction	3
2. Problem Statement	3
3. Motivation	3
4. Methodology	3-4
5. Flow Chart	4-5
6. Results/Outputs	5-6
7.Scope of Project	7
8.Pert Chart	7
9. Contribution of each member	7
10. References	8

1. Introduction

- Sentiment analysis is the process of detecting positive or negative sentiment in text. It's often used by businesses to detect sentiment in social data, gauge brand reputation, and understand customers.
- Big companies take decision based on data now not solely on their intuition and they are getting better results. Analysis of any kind has proven to be very useful in growth of large multinational corporations.
- Companies maintain their social presence by analyzing whether the trend has a positive or a negative impact because sometimes things get viral for the wrong reasons so if the response is positive, they hop on it and use that for acquiring new customers.

2. Problem Statement

- Businesses nowadays analyse and hop on trends using their social media teams. Every big brand has a social media presence and they maintain that by hopping on trends that are being widely loved by public.
- In this project, we try to create a general purpose twitter sentimental analysis system where a user can analyze both tweets posted by a particular user and tweets related to a specific hashtag or keyword. It will help to overcome the challenges of identifying the sentiments of the tweets.

3. Motivation

This system will help in

- Automate the whole process.
- Gathering of Streaming Data.
- Enhance the process of Sentimental Analysis.
- Reduces the Line of code for end user.
- Saves the end user time by collectively perform all the steps together rather than separately (Gathering of Data, Sentimental Analysis).

4. Methodology

We divided our work into different modules for the ease of (work)

1. Fetching the data

First and foremost, data is fetched from the twitter. For that a developer account was created which in turn will provide us with certain keys that will be used to access the tweets. Tweets will be fetched according to the needs and the input entered by the user. The input can be a twitter username and keywords or hashtag.

2. Pre-processing the fetched data

The data that is now fetched needs to be cleaned because the data contains many fields that are not needed and might cause hindrance while analysing the tweets. Different operations like removal of digits and lowering the text were performed. Punctuations were also removed; lemmatization and tokenization were performed. The last pre-processing step is removing stop words from tweets.

3. Performing Sentimental Analysis

Sentimental analysis is performed based on three sentiments neutral, positive and negative. The tweets will be classified based on the text used in that particular tweet. It calculates average polarity and subjectivity over each word in a given text using a dictionary of adjectives and their hand-tagged scores. A sentiment score is calculated using NaiveBayesAnalyzer and based off of those different types of tweets (positive, negative and neutral tweets) are filtered out.

4. Storing results in Data frames

The filtered-out tweets are then stored in three different data frames (positive, negative and neutral). These are the resultant tweets. Now in order to showcase the results that are visually appealing and more understandable to a user we need to plot it on a graph but before that we calculate the percentages of the three different types of tweets using some mathematical formulas.

5. Plotting of Results on graph

Now that we have the different percentage of data we will plot it accordingly on the pie chart using python libraries like matplotlib. Adding to this, we can visualize the extracted result by plotting it. Finally, in the transition phase, the system is tested and modified according to the requirements.

5. Flow Chart

1. Activity diagram

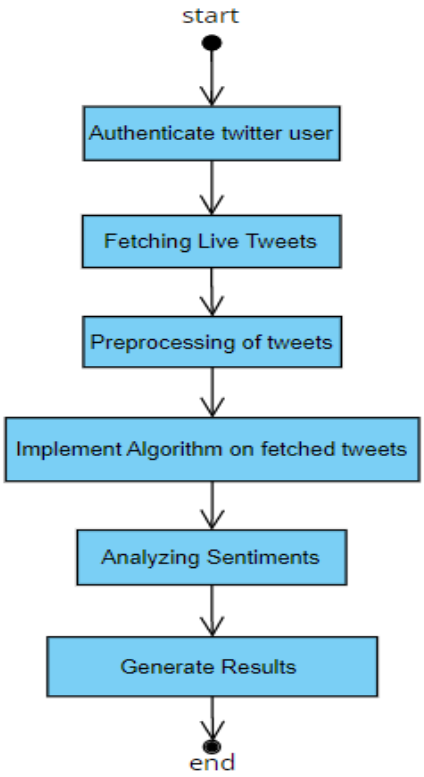

Figure 1: Activity Diagram

6. Results

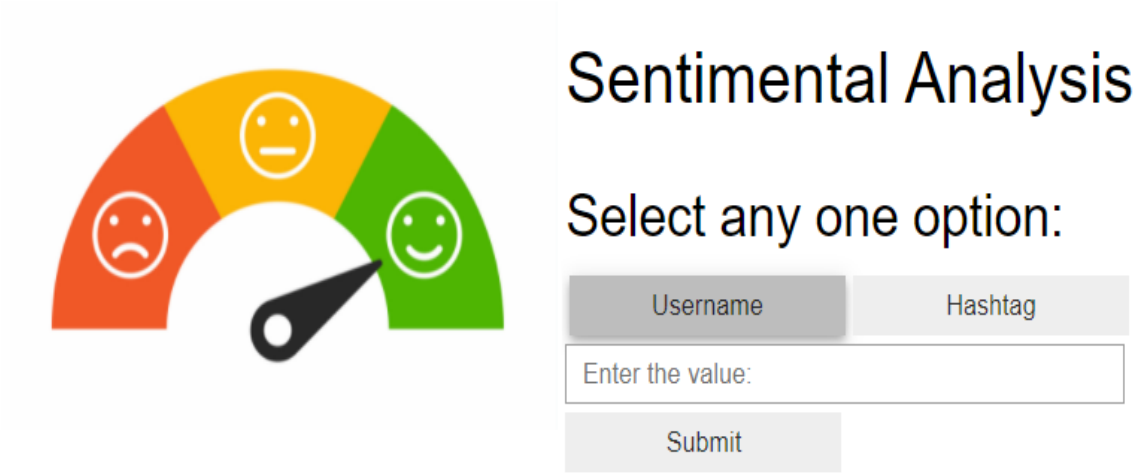

Figure 2: Represent Graphical user interface.

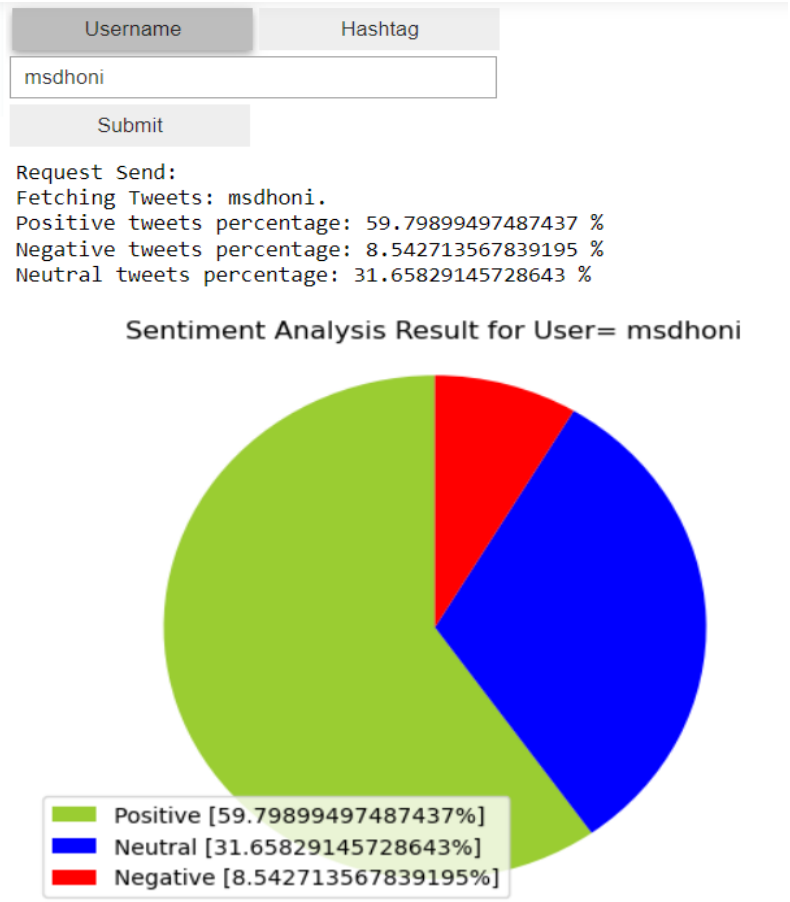

Figure 3: Analyse the sentiment of msdhoni.

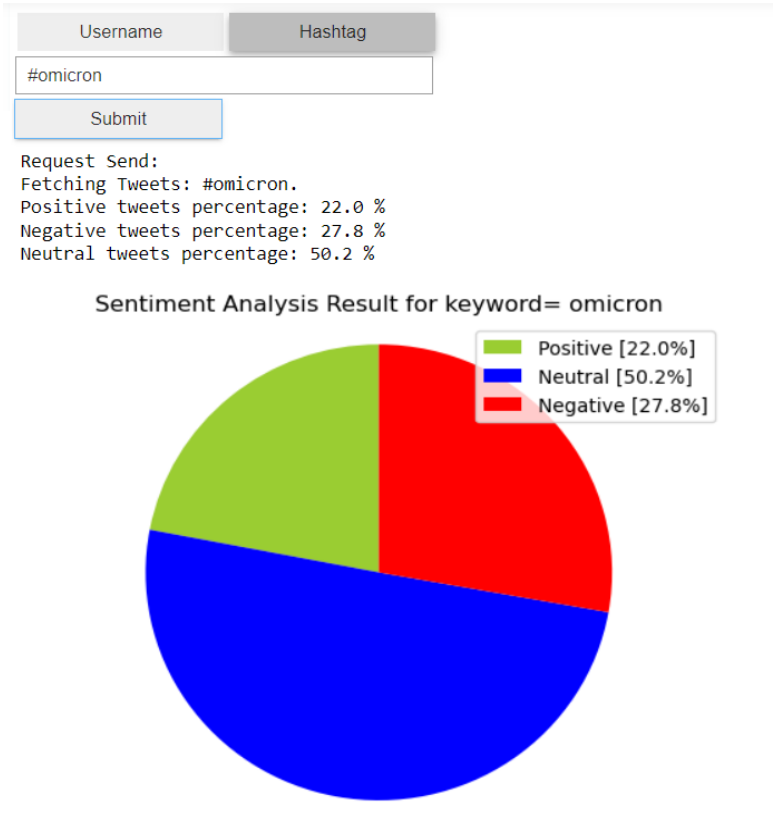

Figure 4: Analysis the sentiments of people on #onicron

7. Scope of Project

- Companies build their own sentimental analysis systems for various purposes but they had to code manually and invest their time and workforces towards it but with this project all these things are just a click away. The interactive GUI will help them do the sentiment analysis of the tweets without any hassle.
- People from various fields who don't have any technical knowledge but want to research also take the help of this GUI to analyse.

8. PERT Chart

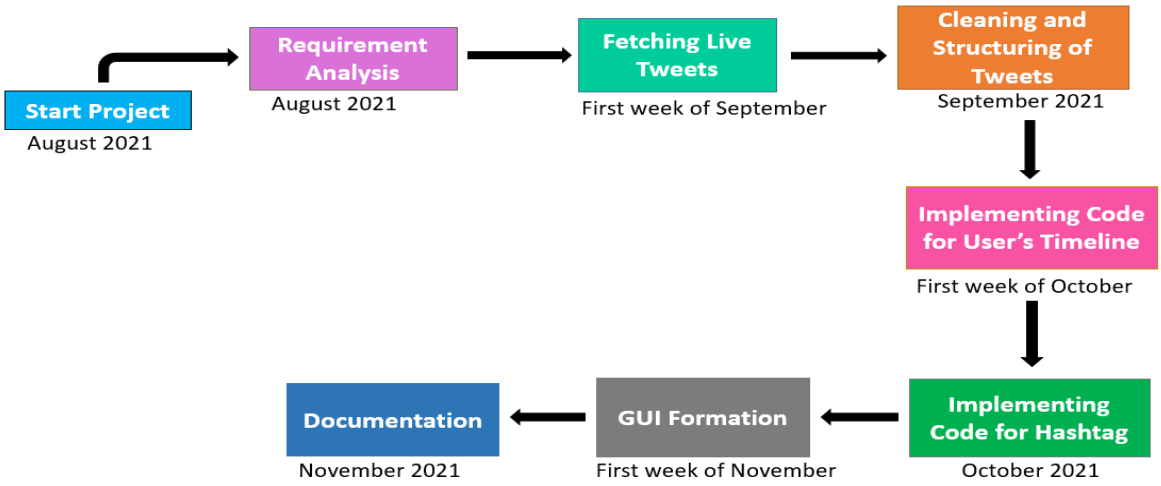

Figure 4: Represents the workflow of the Project.

9. Contribution by each member

- **Lakshay Vasuja**
  Worked on the idea behind the project, implementation of the code and building the GUI and helped in various parts of documentation.
- **Lakshay Sharma**
  Worked on generating twitter keys and tokens and implementation of the code and responsible for generating outputs
- **Divyansh Chandna**
  Worked on the implementation of the code, responsible for hashtag or keyword searching part of the code and helped in various parts of documentation.
- **Janvi Arya**
  Worked on documentation part of the project and also helped in the testing phase.

10.References

- Hossam ELzayady, Khaled M.Badran, Gouda I. Salama,(2018). Sentiment Analysis on Twitter Data using Apache Spark Framework, 10.1109/ICCES.2018.8639195.
- Dr. Khalid N. Alhayyan & Dr. Imran Ahmad (2017). “Discovering and Analyzing Important Real-Time Trends in Noisy Twitter Stream” n.p
- Shashank Gupta (2018, January 7). Sentiment Analysis: Concept, Analysis and Applications. Retrieved From: https://towardsdatascience.com/sentiment-analysis-concept-analysis-and-applications-6c94d6f58c17
- Sandeep Dayananda (2019, May 22). Spark Streaming Tutorial – Sentiment Analysis Using Apache Spark. Retrieved From: https://www.edureka.co/blog/spark-streaming/
- Nikhil Raj (2021, June 15). Starters Guide to Sentiment Analysis using Natural Language Processing. Retrieved From: https://www.analyticsvidhya.com/blog/2021/06/nlp-sentiment-analysis/