AMRITA VISHWA VIDYAPEETHAM

19MAT105-MATHEMATICS FOR INTELLIGENT SYSTEM 1

SENTIMENT ANALYSIS OF TWITTER DATA USING LOGISTIC REGRESSION

GROUP 6

MEGHA MOHAN AM.EN.U4AIE21176

ANUSHKA KAIMAL AM.EN.U4AIE21116

NAVAMI S K AM.EN.U4AIE21146

PARVATHY G PILLAI AM.EN.U4AIE21150

AISWARYA SHAJIL KUMAR AM.EN.U4AIE21107

S1 B.TECH CSE (AIE) B BATCH

AMRITA SCHOOL OF ENGINEERING

AMRITA VISHWA VIDYAPEETHAM
AMRITAPURI 690 525

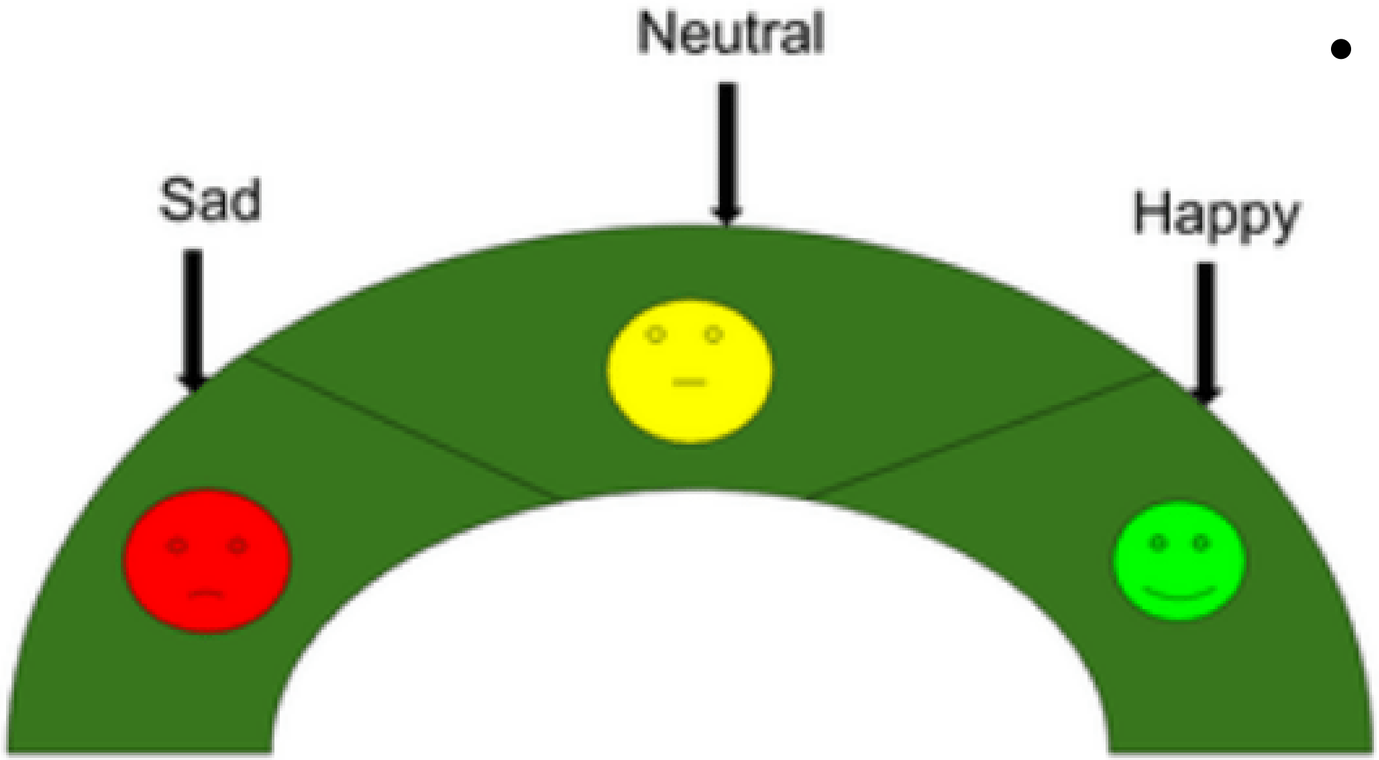
January 2022

INTRODUCTION

- The dramatic increase in internet usage around the world has increased the volume of feedback data making it difficult to organize and classify sentiments.
- Sentiment analysis also referred to as opinion mining, is a sub-machine learning task where we want to determine which is the general sentiment of a given document.
- For example, it has always been a great use for businesses to gain more meaningful insights from their feedback data, which will help them to improve their products and make it easier for customers to select the proper product in less time.

- The main goal of sentiment analysis is to classify and determine the polarity of material on the Internet.
- We use LOGISTIC REGRESSION for the effective accuracy and the prediction of the data set.
- Logistic Regression is a Machine Learning algorithm that is used for classification problems, it is a predictive analysis algorithm based on the concept of probability.

Background Information



- Sentiment analysis is the process of classifying whether a block of text is positive, negative, or neutral. Sentiment analysis is contextual mining of words that indicates the social sentiment of a brand.
- The goal which Sentiment analysis tries to gain is to analyze people's opinions in a way that can help the businesses expand. It focuses not only on polarity (positive, negative & neutral) but also on emotions (happy, sad, angry, etc.).

Why perform Sentiment Analysis?

According to the survey, 80% of the world's data is unstructured. The data needs to be analyzed and be in a structured manner whether it is in the form of emails, texts, documents, articles, and many more.

- Sentiment Analysis is required as it stores data in an efficient, cost-friendly.

- Sentiment analysis solves real-time issues and can help you solve all the real-time scenarios.

WHY DO WE NEED SENTIMENT ANALYSIS ?

Track sentiments real-time

Monitor brand perception

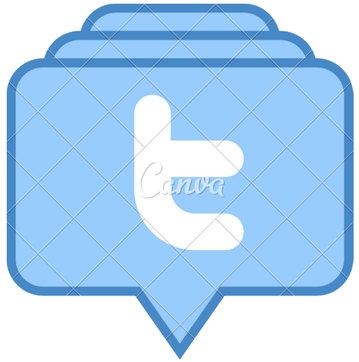
Do better marketing campaigns

Boost your products and services

Improve customer service

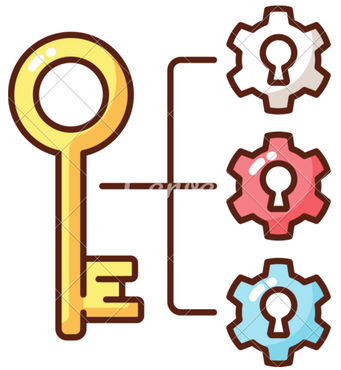
SOLUTION APPROACH

As a part of building a sentiment classifier using logistic regression, we train the model on the Twitter sample dataset.



Thus we do some data pre-processing and cleaning to break down it to an easily understandable format.

The dataset available is in its natural format of tweets(hard for a model to understand).

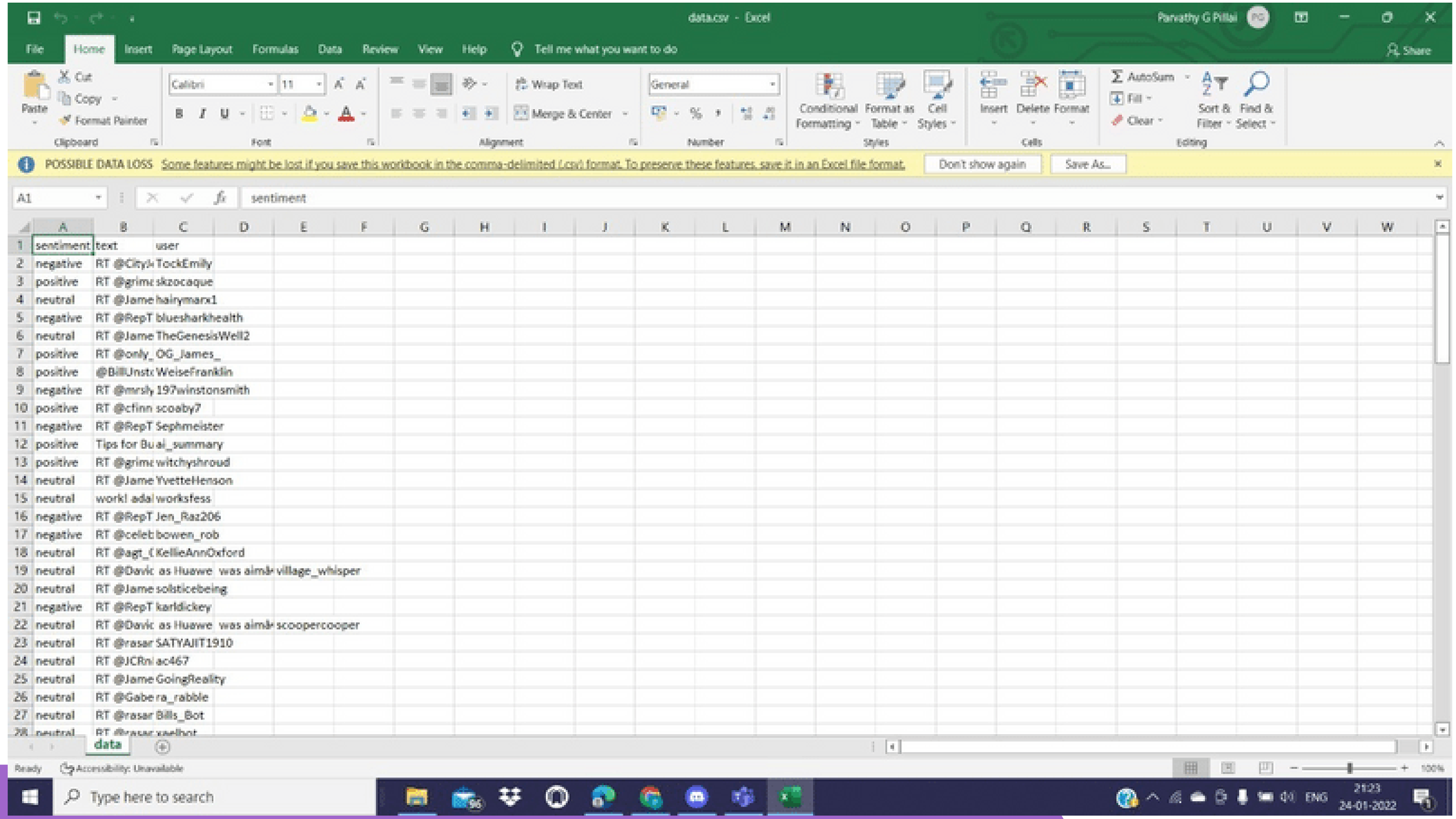


Pre-processing of tweets includes the following steps:

1. Removing punctuations , hyperlinks, and hashtags
2. Tokenization — Converting a sentence into a list of words
3. Removing stop words like and, is, on...
4. Converting all the words to lower cases
5. Lemmatization /stemming — Transforming a word to its root word

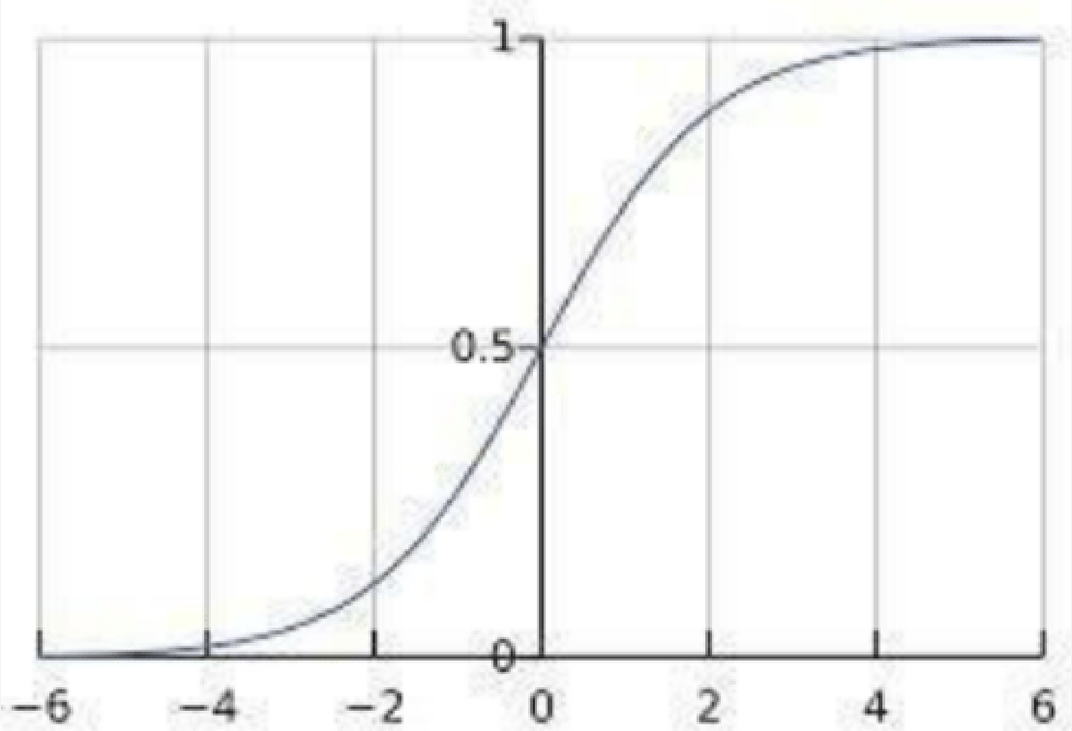
After preprocessing the tweets, it is converted into a feature vector.

DATASET


data.csv - Excel
Parvathy G Pillai
File Home Insert Page Layout Formulas Data Review View Help Tell me what you want to do
Share
Paste Cut Copy Format Painter
Calibri 11
Wrap Text Merge & Center
General
Conditional Formatting Format as Table Cell Styles
Insert Delete Format
AutoSum Fill Clear
Sort & Filter Find & Select
Clipboard Font Alignment Number Styles Cells Editing
POSSIBLE DATA LOSS Some features might be lost if you save this workbook in the comma-delimited (.csv) format. To preserve these features, save it in an Excel file format.
Don't show again
Save As...
A1
sentiment
sentiment text user
negative RT @City/ TockEmily
positive RT @grimc skzocaque
neutral RT @Jame hairymarx1
negative RT @RepT bluesharkhealth
neutral RT @Jame TheGenesisWell2
positive RT @only_ OG_James_
positive @BillUnstc WeiseFranklin
negative RT @mrsly 197winstonsmith
positive RT @cfinn scoaby7
negative RT @RepT Sephmeister
positive Tips for Bu ai_summary
positive RT @grimc witchyshroud
neutral RT @Jame YvetteHenson
neutral workl ada worksfess
negative RT @RepT Jen_Raz206
negative RT @celeb bowen_rob
neutral RT @agt_(KellieAnnOxford
neutral RT @Davic as Huawe was aimé village_whisper
neutral RT @Jame solsticebeing
negative RT @RepT karldickey
neutral RT @Davic as Huawe was aimé scoopercooper
neutral RT @rasar SATYAJIT1910
neutral RT @JCRn ac467
neutral RT @Jame GoingReality
neutral RT @Gabe ra_rabble
neutral RT @rasar Bills_Bot
data
Ready Accessibility: Unavailable
100%
Type here to search
ENG 21:23 24-01-2022

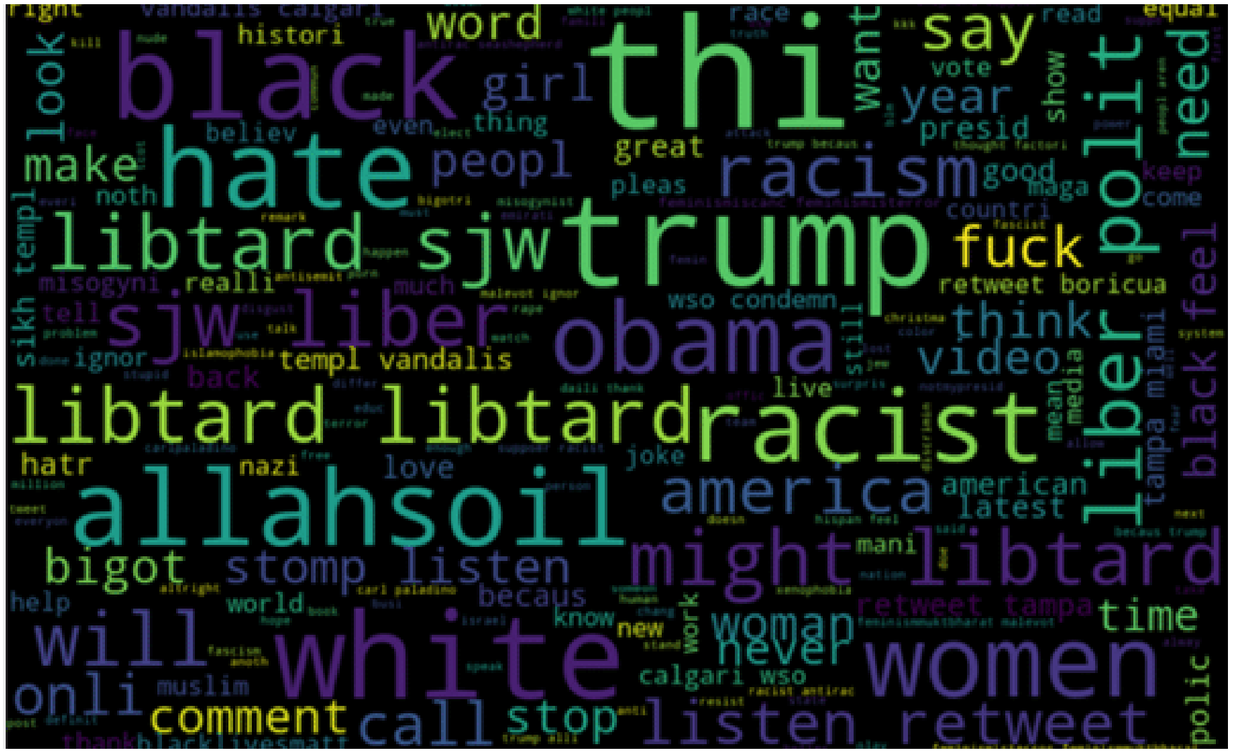
IMPLEMENTATION

- The objective of this task is to detect hate speech in tweets.
- The algorithm used is logistic regression.
- To predict the tweets into negative and positive, we have used a dataset with probability value 0 for negative and 1 for positive
- Logistic regression, uses a logistic function, for instance, a sigmoid function to estimate probabilities between positive or negative tweets



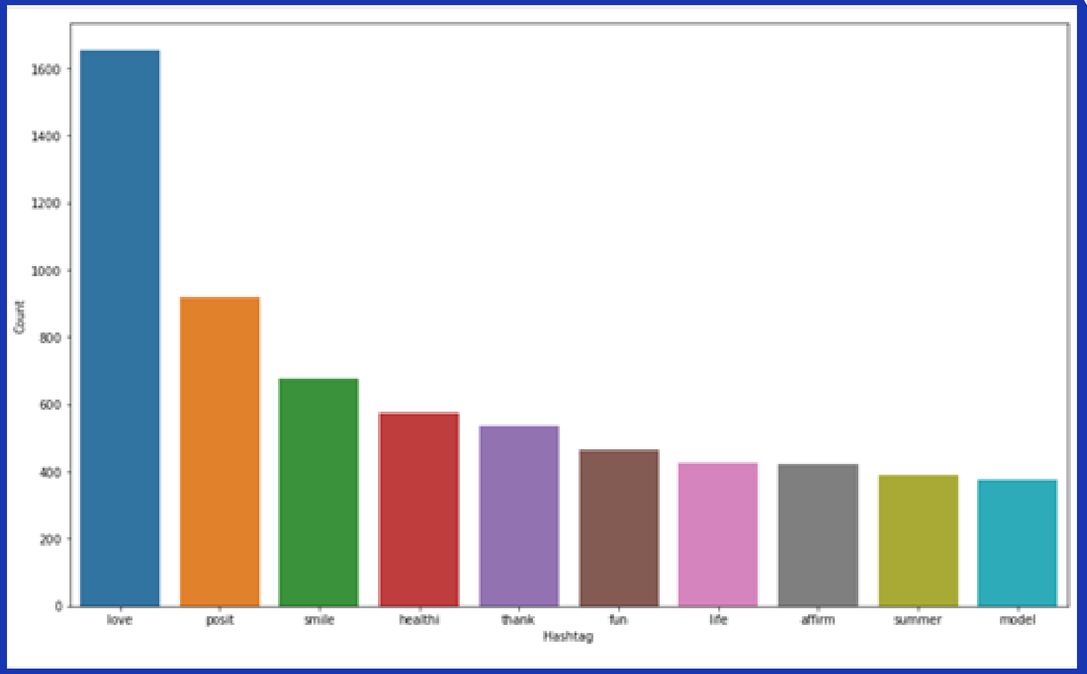
Frequent words visualization for +ve

Frequent words of visualization for -ve

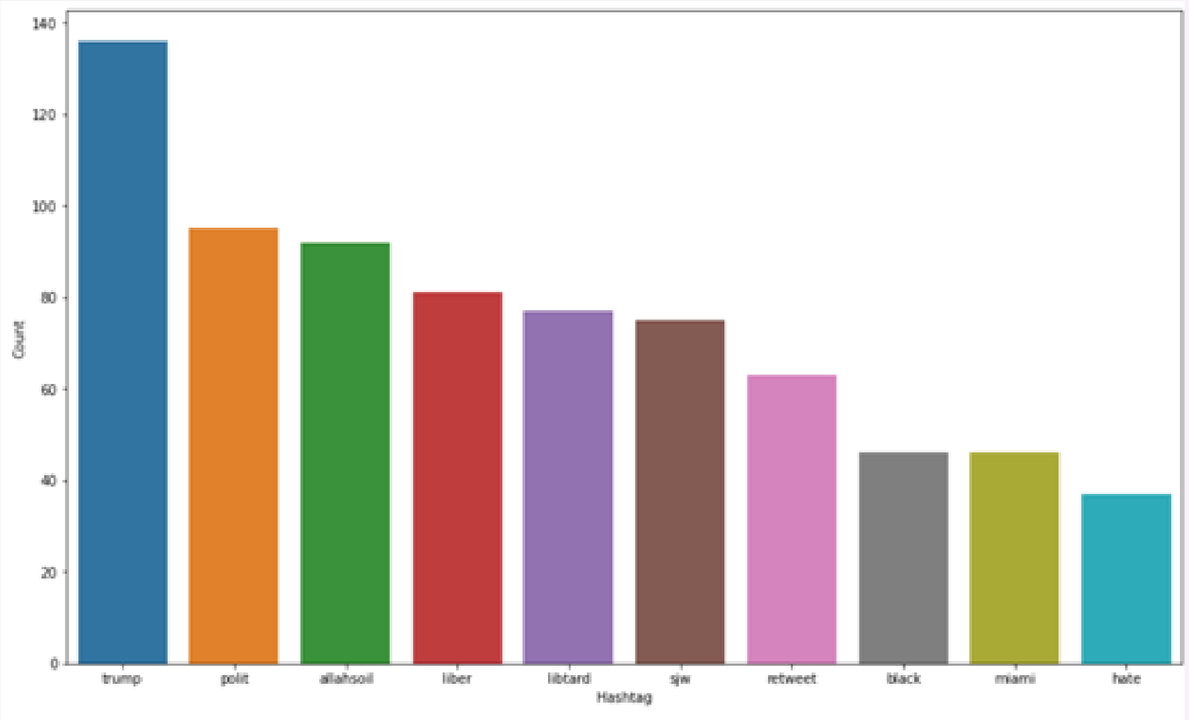


GitHub link :

https://github.com/ANUSHKALA/Twitter_Sentimental_Analysis/tree/main

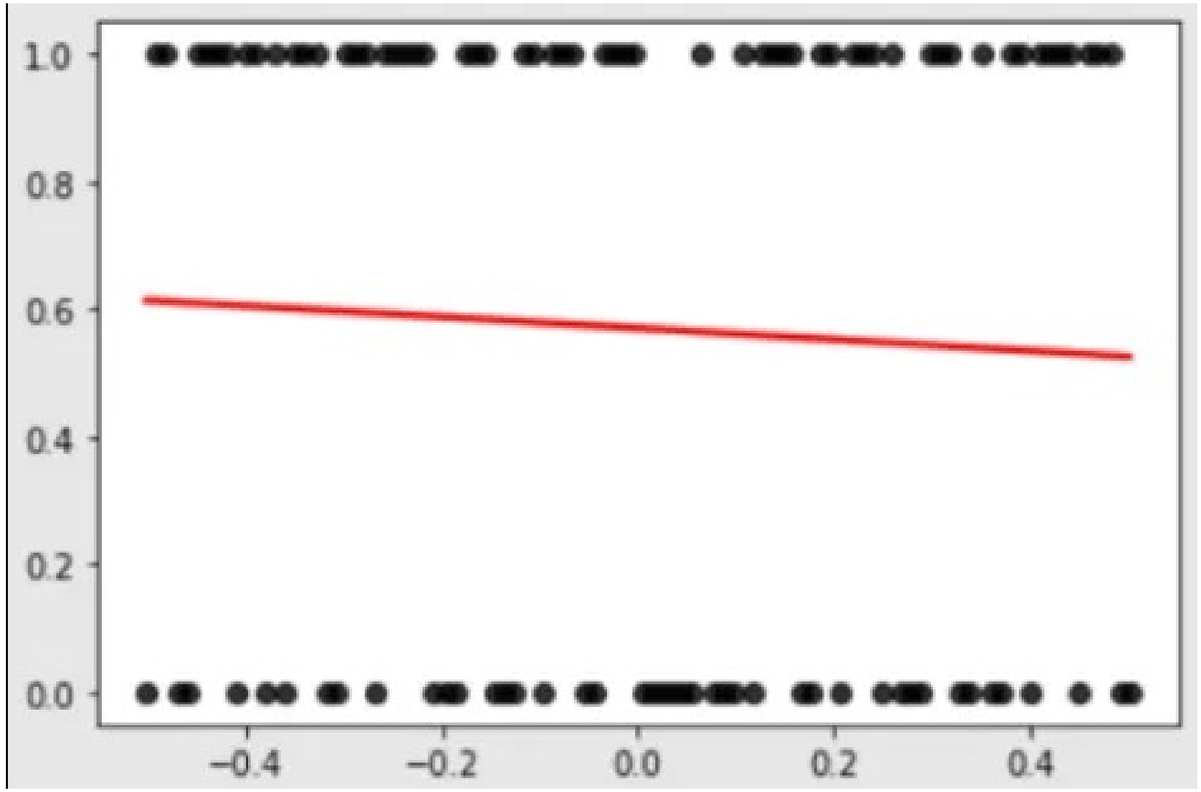


Top 10 hashtags (+ ve)

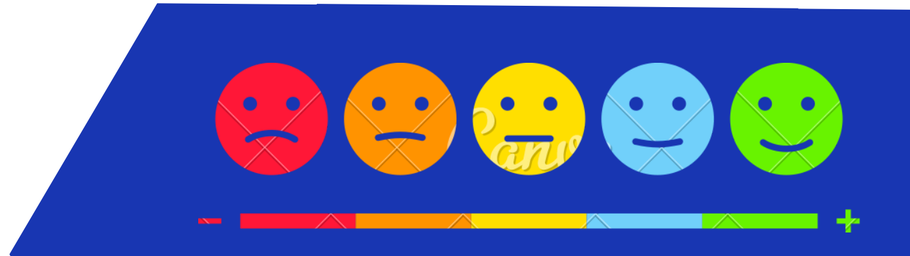


Top 10 hashtags (- ve)

RESULT



Hence we used Logistic regression to classify the tweets as positive and negative.Given below is the graph we obtained during implementation.The Logistic regression model's accuracy is 0.9950



REFERENCE

- https://www.researchgate.net/publication/325101249_Sentiment_Analysis_using_Logistic_Regression_and_Effective_Word_Score_Heuristic
- https://www.brandwatch.com/blog/understanding-sentiment-analysis/#:~:text=Sentiment%20analysis%20is%20extremely%20useful,public%20opinion%20behind%20certain%20topics.&text=Being%20able%20to%20quickly%20see,and%20plan%20for%20the%20future?utm_source=canva&utm_medium=iframely
- https://www.freecodecamp.org/news/what-is-sentiment-analysis-a-complete-guide-to-for-beginners/#:~:text=Sentiment%20analysis%20is%20a%20technique,(NLP)%20to%20achieve%20this.?utm_source=canva&utm_medium=iframely
- https://www.pyoneer.io/post/5-reasons-why-sentiment-analysis-is-important/?utm_source=linkedinstefan&utm_medium=linkedinarticle&utm_campaign=realreasonwhysentiment&utm_source=canva&utm_medium=iframely&utm_source=canva&utm_medium=iframely
- https://joannatrojak.medium.com/sentiment-analysis-with-logistic-regression-in-python-with-nltk-library-d5030b1d84e3

AMRITA
VISHWA VIDYAPEETHAM

Thank you!

AMRITA SCHOOL OF ENGINEERING

AMRITA VISHWA VIDYAPEETHAM
AMRITAPURI 690 525
January 2022