VISVESVARAYA TECHNOLOGICAL UNIVERSITY

JNANA SANGAMA, BELAGAVI – 590 018

An Internship Project Report
on

Twitter Sentiment Analysis

Submitted in partial fulfilment of the requirements for the VIII Semester of the degree of **Bachelor of Engineering in Information Science and Engineering** of Visvesvaraya Technological University, Belagavi

by

Richa Singh
1RN18IS086

Under the Guidance of

Mr. T S Bhagavath Singh
Associate
Professor
Department of ISE

ESTD:2001
An Institute with a Difference

Department of Information Science and Engineering

RNS Institute of Technology

Dr. Vishnuvaradhan Road, Rajarajeshwari Nagar post,
Channasandra, Bengaluru-560098

2021-2022

RNS INSTITUTE OF TECHNOLOGY

Dr. Vishnuvaradhan Road, Rajarajeshwari Nagar post,
Channasandra, Bengaluru - 560098

DEPARTMENT OF INFORMATION SCIENCE AND ENGINEERING

CERTIFICATE

Certified that the Internship work entitled ***Twitter Sentiment Analysis*** has been successfully completed by **Richa Singh (1RN18IS086)** a bonafide student of **RNS Institute of Technology, Bengaluru** in partial fulfilment of the requirements of 8th semester for the award of degree in **Bachelor of Engineering in Information Science and Engineering** of **Visvesvaraya Technological University, Belagavi** during the academic year **2021-2022**. The internship report has been approved as it satisfies the academic requirements in respect of internship work for the said degree.

Mr. T S Bhagavath Singh	**Dr. Suresh L**	**Dr. M K Venkatesha**
Internship Guide	Professor and HoD	Principal
Associate Professor	Department of ISE	RNSIT
Department of ISE	RNSIT	

External Viva

Name of the Examiners	**Signature with Date**
1. _____	**1.** _____
2. _____	**2.** _____

DECLARATION

I, **RICHA SINGH [USN: 1RN18IS086]** student of VII Semester BE, in Information Science and Engineering, RNS Institute of Technology hereby declare that the Internship work entitled ***Twitter Sentiment Analysis*** has been carried out by us and submitted in partial fulfillment of the requirements for the *VIII Semester degree of* ***Bachelor of Engineering in Information Science and Engineering*** *of Visvesvaraya Technological University, Belagavi* during the academic year 2021-2022.

Place: Bengaluru

Date: 9th January 2022

RICHA SINGH

(1RN18IS086)

ABSTRACT

Social media is increasingly used by humans to express their feelings and opinions in the form of short text messages. Detecting sentiments in the text has a wide range of applications including identifying anxiety or depression of individuals and measuring well-being or mood of a community. Sentiments can be expressed in many ways that can be seen such as facial expressions and gestures, speech, and written text. Sentiment Analysis in text documents is essentially a content-based classification problem involving concepts from the domains of Natural Language Processing as well as Machine Learning. In this paper, sentiment recognition based on textual data and the techniques used in sentiment analysis is discussed.

ACKNOWLEDGMENT

At the very onset, I would like to place our gratefulness to all those people who helped me in making the Internship a successful one.

Coming up, this internship to be a success was not easy. Apart from the sheer effort, the enlightenment of the very experienced teachers also plays a paramount role because it is they who guided me in the right direction.

First, I would like to thank the **Management of RNS Institute of Technology** for providing such a healthy environment for the successful completion of internship work.

In this regard, I express sincere gratitude to our beloved Principal **Dr. M K Venkatesha,** for providing us with all the facilities.

We are extremely grateful to our own and beloved Professor and Head of the Department of Information science and Engineering, **Dr. Suresh L**, for having accepted to patronize me in the right direction with all her wisdom.

We place our heartfelt thanks to **Mr. T S Bhagavath Singh,** Associate Professor, Department of Information Science and Engineering for having guided the internship and all the staff members of the Department of Information Science and Engineering for helping at all times.

I thank **Mr. Aman Upadhyay, NASTECH**, for providing the opportunity to be a part of the Internship program and has guided me to complete the same successfully.

I also thank our internship coordinator **Dr. R Rajkumar,** Associate Professor, Department of Information Science and Engineering. I would thank my friends for having supported me with all their strength and might. Last but not the least, I thank my parents for supporting and encouraging me throughout. I have made an honest effort in this assignment.

RICHA SINGH

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS

NLP	-	Natural Language Processing
ML	-	Machine Learning
AI	-	Artificial Intelligence
NLTK	-	Natural Language Toolkit
NB	-	Naïve Bayes
KNN	-	K Nearest Neighbors
UI	-	User Interface

Chapter 1

INTRODUCTION

Background

Natural language processing (NLP) is a subfield of linguistics, computer science, and artificial intelligence concerned with the interactions between computers and human language, how to program computers to process and analyze large amounts of natural language data. The goal is a computer capable of "understanding" the contents of documents, including the contextual nuances of the language within them. The technology can then accurately extract information and insights contained in the documents as well as categorize and organize the documents themselves.

Machine learning is enabling computers to tackle tasks that have, until now, only been carried out by people. At a very high level, machine learning is the process of teaching a computer system how to make accurate predictions when fed data.

Those predictions could be answering whether a piece of fruit in a photo is a banana or an apple, spotting people crossing the road in front of a self-driving car, whether the use of the word *book* in a sentence relates to a paperback or a hotel reservation, whether an email is a spam, or recognizing speech accurately enough to generate captions for a YouTube video.

The key difference from traditional computer software is that a human developer hasn't written code that instructs the system how to tell the difference between the banana and the apple.

Instead, a machine-learning model has been taught how to reliably discriminate between the fruits by being trained on a large amount of data, in this instance likely a huge number of images labeled as containing banana or an apple.

Data, and lots of it, is the key to making machine learning possible.

Existing System

Social networking sites like Twitter have millions of people sharing their thoughts day by day as tweets. Predictions could be sentiment inferred from social media posts and product reviews. AI/ML – based sentiment analysis is crucial for companies to automatically predict whether their customers are happy or not.

.

Proposed System

The process of sentiment analysis could be done automatically without having humans manually review thousands of tweets and customer reviews. In Twitter Sentiment Analysis, we will analyze thousands of Twitter tweets to predict people's sentiment.

There are also various names and having different tasks, e.g., sentiment analysis, opinion extraction, opinion mining, sentiment mining, affect analysis, subjectivity analysis, review mining, etc. Levels of Analysis: In general, sentiment analysis is categorized into mainly three different levels:

A. **Document Level Analysis**: This level classifies that whether the complete document gives a positive sentiment or negative sentiment. The document is on a single topic is considered. Thus, texts which comprise comparative learning cannot be considered under document level.

B. **Sentence Level Analysis**: The task of this level is to decide if each sentence represents opinion into negative, positive, or neutral. Neutral, if the sentence does not give any opinion means it is neutral. Sentence level analysis is related to subjectivity classification. That expresses factual information from sentences that gives subjective aspect and opinions. i.e. good-bad terms.

C. **Entity/Aspect Level Analysis**: Both the document and the sentence level analysis don't find people's likes and dislikes. Entity/Aspect level gives throughout the analysis

Chapter 2

LITERATURE REVIEW

Now a day's social media became so popular that it has a significance impact on trade and marketing. People share their precious views, opinion, experiences on social sites so that others can get benefit from these. Twitter is one of such platforms where common people share their reviews in short blogs i.e. in 140 characters. These reviews can be for anything like any product or service such as movie, stock market, schools, colleges, politics and much more. Here, people share the unbiased opinions about anything they wanted, that's why one can consider these reviews as more generalized and real one. There are five basic steps involved to implement this whole system- first step is selection of training data which is selected on the basis of type of problem, second step is pre-processing of training data which means removing irrelevant information like URLs, user names, slang words, symbols etc., third step is to establish a connection with Twitter database using twitter API from where recent tweets can be learning algorithms like Naive Bayes, Support Vector Machine are used for the classification of tweets into different classes and in the final step results are displayed on the basis of polarities of tweets after their classification. SentiWordNet2 is lexical resource which is broadly used for opinion mining, and it provides sentiment score to each synset of WordNet. These scores are of three types; positivity score, negativity score and objectivity score. So instead of using Naive Bayes in traditional way, these scores can be taken into account so that more accuracy can be attained.

Prabhsimran Singh, Ravindra Singh and Karanjeeet Singh Kalhon, [7] they have examined this government policy the demonetization from the ordinary person’s viewpoint with the use of the approach of sentiment analysis and using Twitters data, Tweets are collected using certain hashtag (#demonetization).

The fundamental task in Sentiment Analysis is classifying the polarity of a given tweets feature. The polarity is in three classes i.e. Positive, Negative and Neutral. Polarity identification is done by using different lexicons e.g. Bing Lui sentiment lexicon, SentiWordNet etc. which help to calculate sentiment strength, sentiment score, etc. [13]

The limitations of available systems are not sufficient to deal with the complex structure of the big data. In this section, we present some of the limitations that are present in the existing system.

1) The available stems like Twitter-Monitor and Real Time Twitter Trend Mining System require extensive data cleaning, data scraping and integration strategies that will ultimately increase the overhead [9].

2) For real time analytics, the available system is inefficient.

3) It is very time-consuming process to analyze the huge amount of data in a short period of time.

The Existing Database is not able to process the big amount of data within specified amount of time. Also, this type of database is limited for processing of structured data and has a limitation when dealing with a large amount of data. So, the traditional solution cannot help an organization to manage and process unstructured data. With the use of Big Data technologies like Hadoop [4] is the best way to solve Big Data challenges.

The prepared tweets were passed to further process using domain relevant packages in order to generate ground truth for tweets. After the tweets were classed in terms of polarity (positive and negative) and emotion (anger, joy, sadness, disgust, fear, and surprise)

In sentiment analysis using a multi-dataset in different languages to understand which can give a better result when used with ML algorithms. Multi-language datasets such as English, modern standard Arabic and dialect Arabic are collected for the experiment. We evaluate is based on two parameters which are accuracy and runtime. Sentiment analysis is language and culture dependent. Any successful sentiment analysis tool should consider two main aspects (culture and language). The result for the English language and Modern Standard Arabic result is better than Dialect Arabic, that is because ML algorithms still needs an enhancement for Dialect Arabic. However, the ML algorithms need to improve to get a better result with Dialect Arabic than 50.80% accuracy. Future Enhancement can be done for Sentiment analysis for Dravidian Languages

Twitter about iPhone and iPad in certain locations: Chicago, Los Angeles, San Francisco, and New York. The data then was passed through a sequence of pre-processing and cleaning to extract relevant data which was the cleaned tweets

It was found that the K-NN using Manhattan distance has achieved the highest accuracy rates 96.58 % and 99.94 % for iPad and iPhone emotion data sets by using the cross validation technique.

In addition, the KNN was also outperformed the NB and SVM because of two main reasons, firstly, the KNN has minimum trained time to build model. Secondly, with using KNN, the highest accuracy rates 98.8% and 99.95% were achieved for the iPad and iPhone emotion data sets respectively

Removing similar tweets in a percentage will help researcher to build data sets more properly. Moreover, adding more features for the data sets especially for the polarity data set such as neutral very positive, very negative

Chapter 3

ANALYSIS

Introduction

Modern society runs over social media most time of every day. Web users spend the most time on social media and they share many details with their friends. Such information obtained from their chat has been used in several applications. The sentiment analysis is the one which has been applied with Twitter data set toward identifying the emotion of any user and based on those different problems can be solved.
In the marketing field, companies use it to develop their strategies, to understand customers' feelings towards products or brands, how people respond to their campaigns or product launches and why consumers don't buy some products.

Requirement Specification

The requirement that the system needs is categorized into functional and non-functional requirements. These requirements are listed below

Functional Requirements

Functional requirements are the functions or features that must be included in any system to satisfy the business needs and be acceptable to the users. Based on this, the functional requirements that the system must require are as follows: -System should be able to process new tweets stored in the database after retrieval -System should be able to analyze data and classify each tweet polarity

Non - Functional Requirements

A non-functional requirement is a description of features, characteristics, and attributes of the system as well as any constraints that may limit the boundaries of the proposed system. The non-functional requirements are essentially based on the performance, information, economy, control, and security efficiency and services.
Based on these the non-functional requirements are as follows:
-User friendly -System should provide better accuracy
-To perform with efficient throughput and response time

User Characteristics

User Characteristics Sentiment analysis can be defined as a process that automates mining of attitudes, opinions, views and emotions from text, speech, tweets and database sources through Natural Language Processing (NLP). Sentiment analysis involves classifying opinions in text into categories like "positive" or "negative" or "neutral". It's also referred as subjectivity analysis, opinion mining, and appraisal extraction. The words opinion, sentiment, view and belief are used interchangeably but there are differences between them. Opinion: A conclusion open to dispute (because different experts have different opinions) View: subjective opinion Belief: deliberate acceptance and intellectual assent Sentiment: opinion representing one’s feelings Sentiment Analysis is a term that include many tasks such as sentiment extraction, entiment classification, subjectivity classification, summarization of opinions or opinion spam detection, among others. It aims to analyze people's sentiments, attitudes, opinions emotions, etc. towards elements such as, products, individuals, topics, organizations, and services.

Software Requirements

- Operating system: Windows 7/8/10/11 or Linux
- Twitter dataset
- Modern web browser with internet access
- Python 3

Hardware Requirements

- Processor: Intel i5 or more
- RAM: 8 GB or more
- Cache: 512 KB
- Hard disk: 16 GB
- Monitor: 1024 x 720 display
- Speed: 2.7GHZ or more

Chapter 4

SYSTEM DESIGN

Introduction

tweet_id	Sentiment	content
1.96E+09	surprise	Got the news
1.96E+09	sadness	The storm is here and the electricity is gone
1.96E+09	sadness	So sleepy again and it's not even that late. I fail once again.
1.96E+09	fun	@IdleThumbs Up is out? I didn't get the memo It looks amazing.
1.96E+09	worry	Work day 7 of 7,

Table 4.1: Dataset description

The table above shows the raw text or un-processed text in .csv file.

High-level design

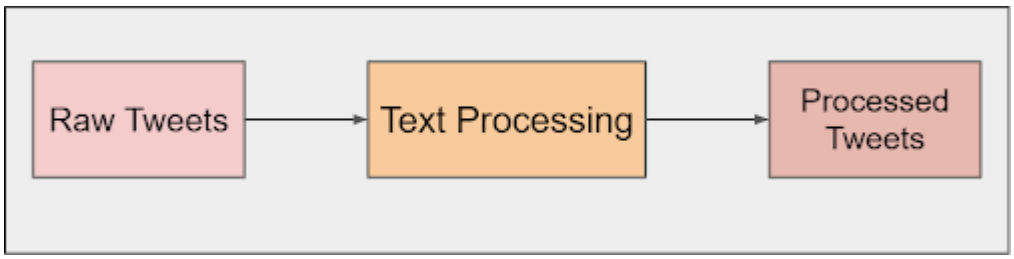

Figure 4.2.1: Diagram for text processing

The raw tweets have to go through a sequence of process to become processed tweets, this process is called data cleaning.

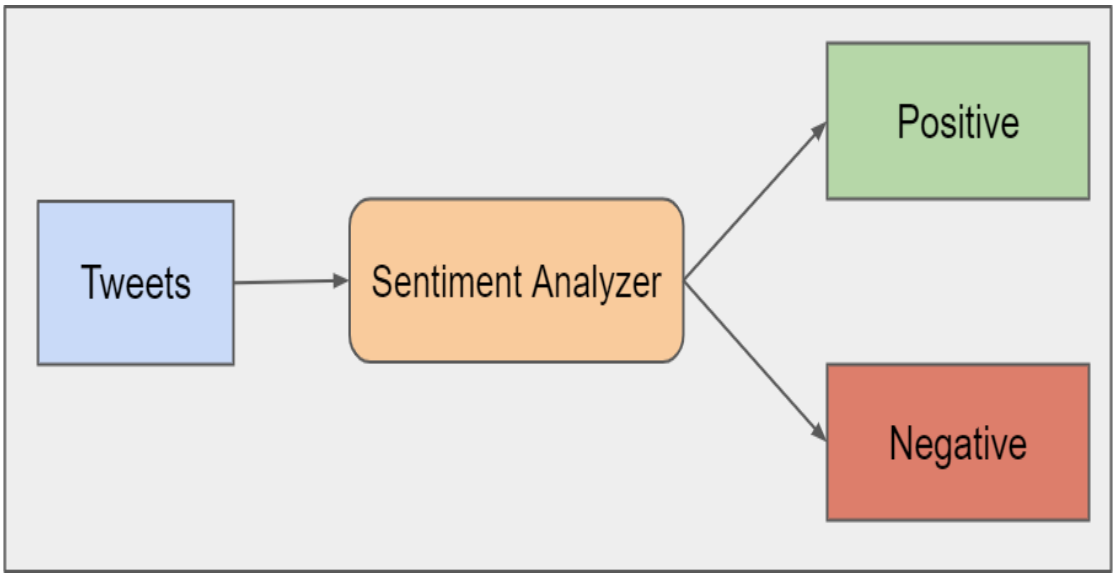

Figure 4.2.2: Diagram for sentiment analysis

The processed text is fed to the analyzing algorithm that shows the polarity of the text.

Low-level design

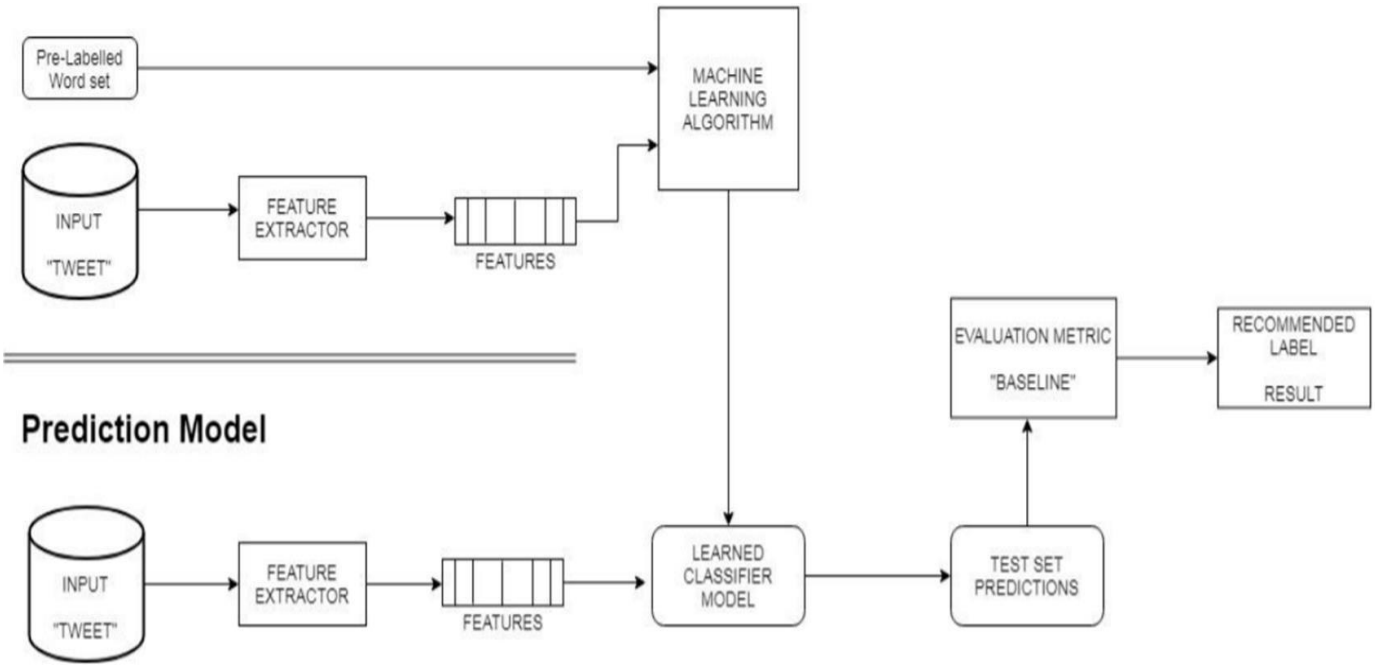

Figure 4.3: Detailed Diagram for sentiment analysis

Functional Module / Framework

Following are functional modules and frameworks used in the project:

- NumPy - NumPy is a library for the Python programming language, adding support for large, multi-dimensional arrays and matrices, along with a large collection of high-level mathematical functions to operate on these arrays.
- Matplotlib - Matplotlib is a plotting library for the Python programming language and its numerical mathematics extension NumPy. It provides an object-oriented API for embedding plots into applications using general-purpose GUI toolkits like Tkinter, wxPython, Qt, or GTK.
- Seaborn - Seaborn is a Python data visualization library based on matplotlib. It provides a high-level interface for drawing attractive and informative statistical graphics.
- Pandas – It is a Python library for data analysis. Pandas is built on top of 2 core Python libraries – matplotlib for data visualization and NumPy for mathematical operations. Pandas acts as a wrapper over these libraries, allowing us to access many of matplotlib's and NumPy's methods woth less code.
- Streamlit - Streamlit is an open-source python framework for building web apps for Machine Learning and Data Science. We can instantly develop web apps and deploy them easily using Streamlit. Streamlit allows you to write an app the same way you write a python code. Streamlit makes it seamless to work on the interactive loop of coding and viewing results in the web app.
- Sqlite3 - SQLite is a C library that provides a lightweight disk-based database that doesn't require a separate server process and allows accessing the database using a nonstandard variant of the SQL query language. It's also possible to prototype an application using SQLite and then port the code to a larger database such as PostgreSQL or Oracle. SQLite3 can be integrated with Python using sqlite3 module, which was written by Gerhard Haring.
- Nltk – Natural Language Toolkit is a suite of libraries and programs for symbolic and statistical natural language processing (NLP) for Python Language. NLTK includes graphical demonstrations and sample data.

Chapter 5

IMPLEMENTATION DETAILS

1. Exploratory data analysis

```
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
from jupyterthemes import jtplot
jtplot.style(theme='monokai', context='notebook', ticks=True, grid=False)
```

Figure 5.1.1: importing libraries and datasets

```
In [5]: tweets_df.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 31962 entries, 0 to 31961
Data columns (total 3 columns):
 #   Column  Non-Null Count  Dtype
---  ------  --------------  -----
 0   id      31962 non-null  int64
 1   label   31962 non-null  int64
 2   tweet   31962 non-null  object
dtypes: int64(2), object(1)
memory usage: 749.2+ KB
```

Figure 5.1.2: Information about the data set

```
tweets_df.describe()
```

	id	label
count	31962.000000	31962.000000
mean	15981.500000	0.070146
std	9226.778988	0.255397
min	1.000000	0.000000
25%	7991.250000	0.000000
50%	15981.500000	0.000000
75%	23971.750000	0.000000
max	31962.000000	1.000000

Figure 5.1.3: Describing the data set

2. Data pre-processing and data cleaning

- Convert to lowercase
- Replace @user with at_user
- Replace http://www.url.com with URL
- Remove special characters
- Remove words that doesn't affect sentiment like pronouns, conjunctions

Raw text: "Hello, I loveeee www.wikipedia .com because it has so much information which is available for free !!!"

Pre-processed text: "Hello loveeee URL has so much information available for free"

This process of removal of pre-processing is called stop word removal.

3. Creating feature vector

1. Tokenization
2. Replace two or more with two occurrences
3. Remove the word that doesn't start with the alphabet
4. Remove all stop word which doesn't have any sentiment
5. Attach "not_" to words which have negation impact.
6. Remove negation words.
7. Remove the remaining punctuations.

Processing above text

Pre-processed text : "Hello loveeee URL has so much information available for free"

FeatureVector = "love, information, available, free"

4. Create a feature list

Featurelist = featureList + FeatureVector

5. Append sentiment and feature vector

Tweets = Tweets + ["positive, information, available, free"]

Repeat the above step for each tweet to get a complete list of processed tweets

5. Apply features to get training set

```
X.shape

(31962, 47386)

y.shape

(31962,)

from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2)

from sklearn.naive_bayes import MultinomialNB

NB_classifier = MultinomialNB()
NB_classifier.fit(X_train, y_train)

MultinomialNB(alpha=1.0, class_prior=None, fit_prior=True)
```

Figure 5.1.4: Getting training set

Algorithms like Naïve Bayes and Logistic regression take the training data and predict the results for the test data. The algorithm that gives highest accuracy is take for the classification.

6. Trained classifier is stored in Pickle

The trained classifier is stored in the system to avoid training classification every time. Whenever it is required to classify the tweet or sentence then the classifier is loaded from the pickle and used to give the output without training once again.

7. User input sentence to classify

Now comes the user part. User enters a sentence or a paragraph and the pre-processing is done to create features. Then the features are taken to do further processing and classification to give the sentiment as output. The output may vary depending on the labels given in the dataset or the algorithm used to classify.

8. Storing the output in database

The output and the classification information is stored in the database.

Chapter 6

TESTING

Testing is done by finding the maximum posterior probability of the test data and then finding its accuracy.

$$P(A|B) = \frac{P(B|A) * P(A)}{P(B)}$$

Where,

P(A|B) – Posterior Probability

P(A) – Prior Probability

P(B|A) – Maximum Likelihood

P(B) – Marginal Probability

Accuracy = [Summation (True positive) + Summation (True Negative)] / (Summation (Total Population)]

The purpose of testing is to discover errors. Testing is the process of trying to discover every conceivable fault or weakness in a work product. It provides a way to check functionality of components, sub-assemblies, and/or a finished product it is the process of exercising software with the intent of ensuring that the software system meets its requirements and user expectations and does not fail in an unacceptable manner.

Types of Testing

- **Unit Testing**: A unit test is a way of testing a unit - the smallest piece of code that can be logically isolated in a system. In most programming languages, that is a function, a subroutine, a method, or property. The isolated part of the definition is important.
- **Integration Testing**: Integration testing is the phase in software testing in which individual software modules are combined and tested as a group. Integration testing is conducted to evaluate the compliance of a system or component with specified functional requirements.
- **Validation Testing**: The process of evaluating software during the development process or at the end of the development process to determine whether it satisfies specified business requirements. Validation Testing ensures that the product actually meets the client's needs.

- **System Testing**: System Testing is a type of software testing that is performed on a complete integrated system to evaluate the compliance of the system with the corresponding requirements. In system testing, integration testing passed components are taken as input.

Testing User Interface

Number	Description	Action	output
1.	Input no text in the text box	Press Submit	Neutral sentiment is displayed
2.	Input random characters in the text box	Press Submit	Neutral Sentiment is displayed
3.	Input “I love rainy season”	Press Submit	Love is displayed as sentiment
4.	Input “I hate standing in a queue”	Press Submit	Hate is displayed as sentiment
5.	-	Select Home from drop-down menu	Home page is displayed
6.	-	Select Monitor from drop-down menu	Montior page is displayed
7.	-	Press page metrics	Page metrics is displayed
8.	-	Press emotion classifier metrics	Emotion classifier metrics is displayed
9.	-	Select about from drop-down menu	Empty about page is displayed

Table 6.1: Testing UI

.

Chapter 7

RESULTS

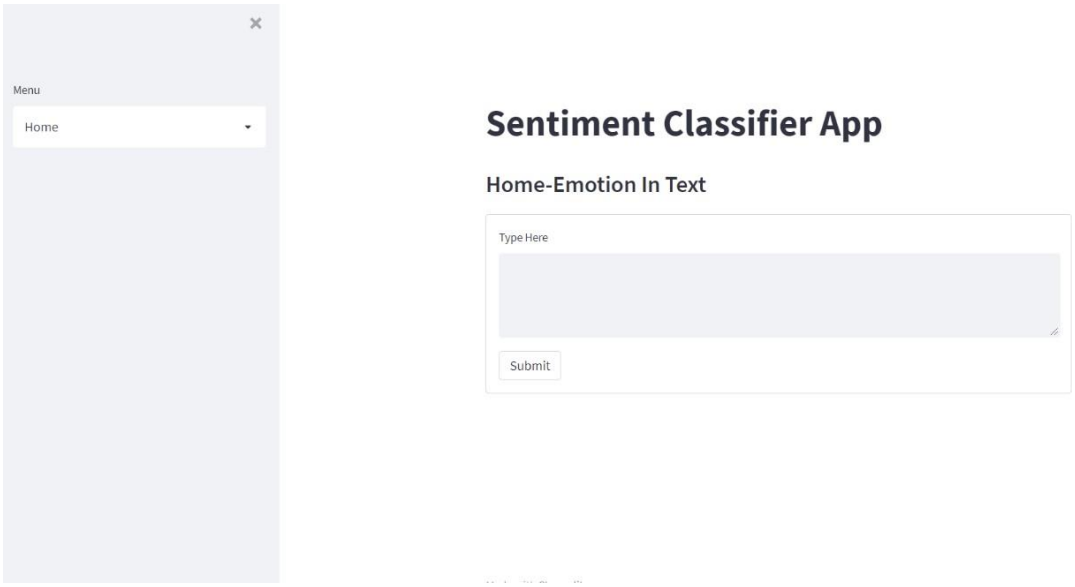

Figure 7.1: Landing page UI

The above image shows the user interface to input text.

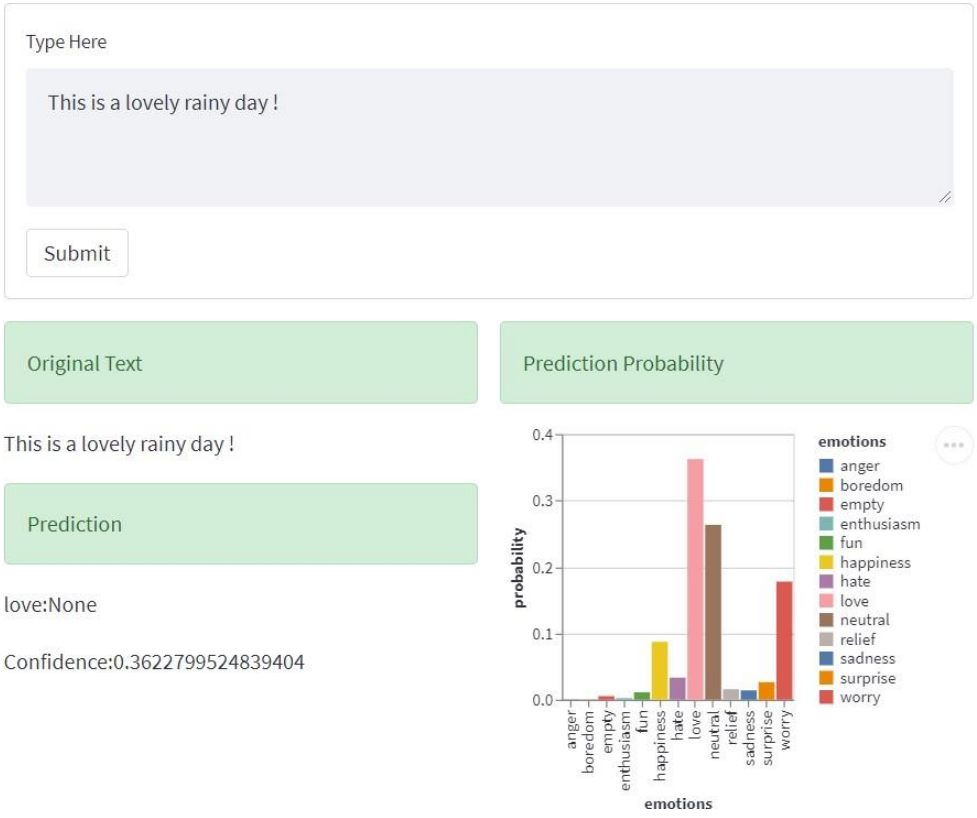

Figure 7.2: results with bar graph

Figure 7.2 shows the result after pressing the submit button

Page Metrics

	Pagename	Time_of_Visit
0	Home	2021-12-03 23:05:18.577647
1	Home	2021-12-03 23:10:16.756982
2	Home	2021-12-03 23:11:19.868790
3	Home	2021-12-03 23:11:44.673878
4	Home	2021-12-03 23:11:56.091538
5	Home	2021-12-03 23:12:03.553813
6	Home	2021-12-03 23:12:23.260775
7	Home	2021-12-03 23:12:36.889253
8	Home	2021-12-03 23:12:40.969936
9	Home	2021-12-03 23:12:51.330231

Table 7.3: page visit log table

The above table, 7.3 shows the page name visited by the user and their time and date of visit in the database.

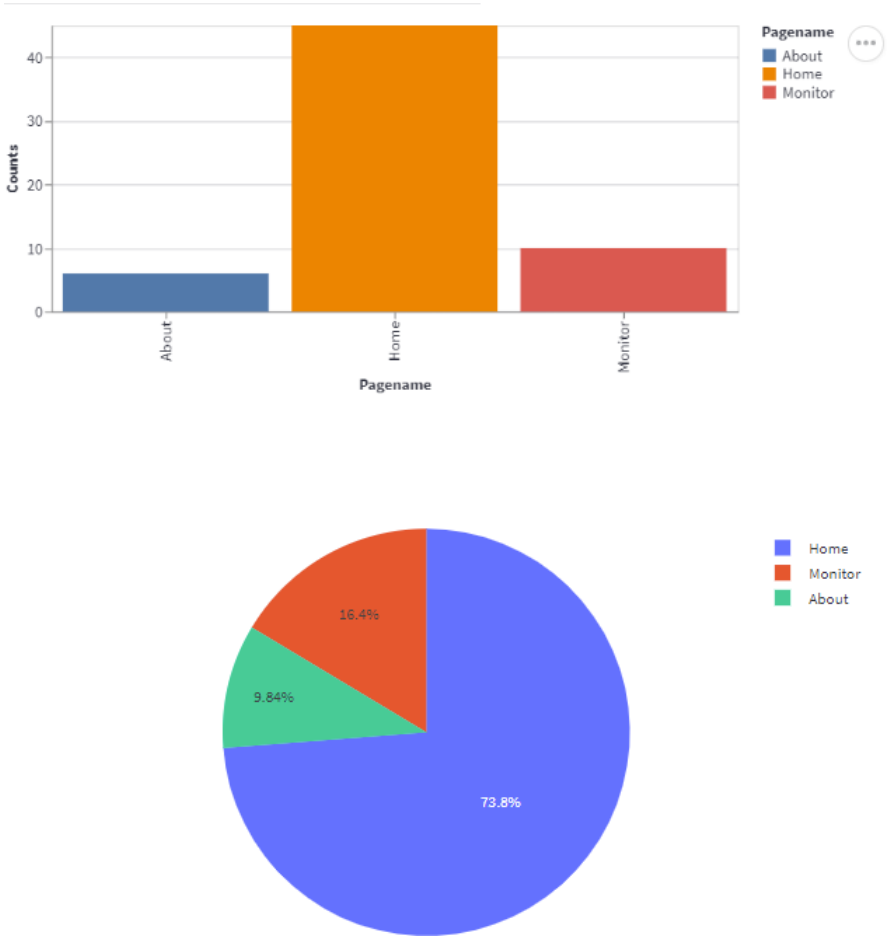

Figure 7.3: graphical representation

The above figure, 7.3 shows the page visits in graphical representation

Emotion Classifier Metrics

	Rawtext	Prediction	Probability	Time_of_Visit
0	Hi I'm Shreelakshmi Joshi !	neutral	0.5318	2021-12-03 23:11:19.883826
1	Hi I'm Shreelakshmi Joshi...	love	0.3679	2021-12-03 23:11:44.686885
2	Hi I'm Shreelakshmi Joshi...	love	0.3679	2021-12-03 23:12:03.565817
3	I'm Shreelakshmi Joshi a...	love	0.6386	2021-12-03 23:12:23.270777
4	I'm Shreelakshmi Joshi a...	neutral	0.4146	2021-12-03 23:12:40.979942
5	I'm Shreelakshmi Joshi a...	neutral	0.2535	2021-12-03 23:13:09.171008
6	I'm Shreelakshmi Joshi a...	love	0.3090	2021-12-03 23:13:26.255607
7	I hate doing that	hate	0.3902	2021-12-03 23:13:52.512283
8	I hate that thing	hate	0.5838	2021-12-03 23:14:32.431577
9	What a nice day !	surprise	0.2471	2021-12-03 23:22:00.055347

Table 7.4: Text entry and result log table

The results from the user are stored in the database table is showed in the above table 7.4

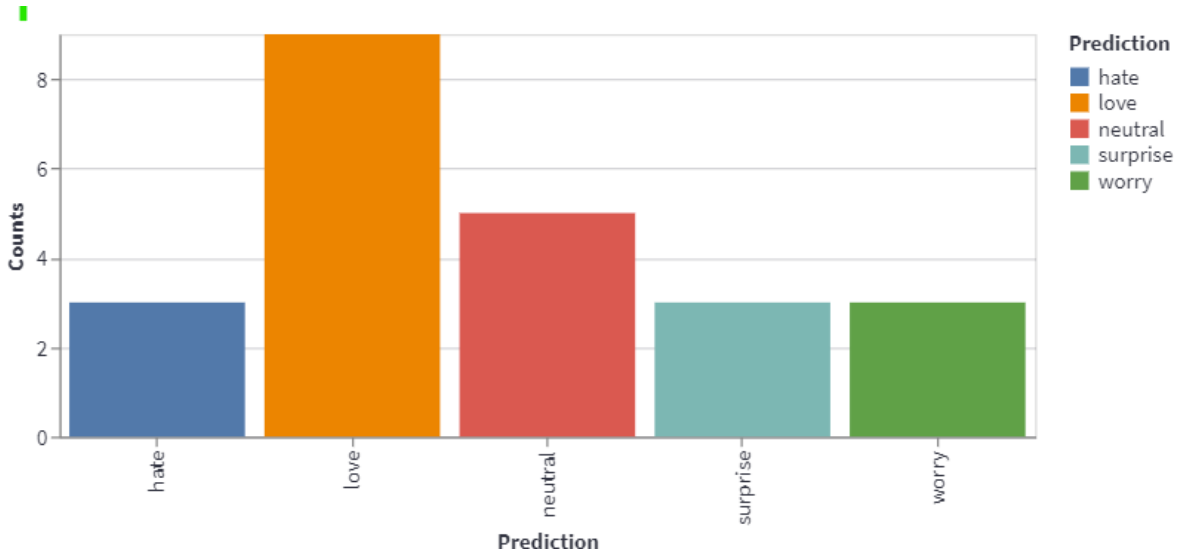

Figure 7.4: graphical representation of test results

In the above figure 7.4, graphical representation of the table 7.4 is shown.

Chapter 8

CONCLUSION AND FUTURE ENHANCEMENT

Conclusion

This project was an attempt to make the structure and working of a Twitter Sentiment Analysis simpler and user-friendly. This was an attempt to make it similar to the real-world implementation. In this scenario, all the undertakings of the Twitter Sentiment Analysis were achieved in a constructive manner. Given the right guidance and support its applications and availability can be enhanced. The sentiment analysis model can also be applied to the stock market, customer feedback services for a particular product, etc.

Future Enhancements

1. Finding Sarcasm in the text and predicting emotion accordingly.
2. Sentiment analysis can be done on live tweets on a large-scale using Twitter API
3. Future Enhancement can be done for Sentiment analysis for Dravidian Languages
4. Many analyzers don't perform well when the number of classes is increased. Also, it's still not tested that how accurate the model will be for topics other than the one in consideration. Hence sentiment analysis has a very bright scope of development in the future.
5. Machine Learning algorithm's accuracy to be increased.
6. Database is limited for processing structured data and has a limitation when dealing with a large amount of data.

Chapter 9

REFERENCES

[1] https://www.kaggle.com for Twitter sentiment dataset

[2] https://ieeexplore.ieee.org/document/7877424

[3] https://ieeexplore.ieee.org/document/9537243

[4] https://ieeexplore.ieee.org/document/8474783

[5] https://ieeexplore.ieee.org/document/7912034

[6] https://ieeexplore.ieee.org/document/8252118

[7] https://ieeexplore.ieee.org/document/8515844