A DISSERATATION REPORT ON

Sentimental Analysis of Social Media Data for Brand Monitoring

SUBMITTED IN PARTIAL FULFILMENT FOR THE REQUIREMENT OF THE DEGREE

BACHELOR OF ENGINEERING

IN

COMPUTER SCIENCE & ENGINEERING

(ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING)

SUBMITTED BY

--GROUP MEMBERS--	--ROLL NO—
SHIVAM KELSHIKAR	26
HARSHADA KHANDAGALE	30
OMKAR PRAJAPATI	41
PRATHAM SINGASANE	54

UNDER GUIDANCE OF

“PROF.SANJAY PARBHANIKAR”

G.V. ACHARYA INSTITUTE OF ENGINEERING AND TECHNOLOGY

UNIVERSITY OF MUMBAI

A.Y. 2024 - 2025

G.V. ACHARYA INSTITUTE OF ENGINEERING AND TECHNOLOGY

SHELU-410201

CERTIFICATE

This is to certify that, this Project report entitled

“SENTIMENT ANALYSIS OF SOCIAL MEDIA CONTENT FOR BRAND MONITORING”

Submitted By

--GROUP MEMBERS--	--ROLL NO—
SHIVAM KELSHIKAR	26
HARSHADA KHANDAGALE	30
OMKAR PRAJAPATI	41
PRATHAM SINGASANE	54

For partial fulfillment of the requirement for the award of the degree of “BE in Computer Science & Engineering (Artificial Intelligence and Machine Learning) ” as laid down by UNIVERSITY OF MUMBAI is a record of their own work carried out by them under my supervision and guidance during year 2024-2025.

-- --

Project Guide
PROF.SANJAY PARBHANIKAR

Head of Department
PROF.NARAYAN KHARJE

SIGNATURE OF EXAMINER

Principal

PROJECT REPORT APPROVAL

This project report entitled **"SENTIMENTAL ANALYSIS OF SOCIAL MEDIA CONTENT FOR BRAND MONITORING"** by **"SHIVAM KELSHIKAR(26)," "HARSHADA KHANDAGALE (30)", OMKAR PRAJAPATI (41), "PRATHAM SINGHASANE (54)"** is approved
for the degree of **"Bachelor of Computer Science & Engineering (AIML)".**

Examiners Signature

1. ____________________________

2. ____________________________

Date: / / 2025

Place: SHELU

DECLARATION

We declare that this written submission represents our ideas in our own words and where others' ideas or words have been included, we have adequately cited and referenced the original sources. We also declare that we have adhered to all principal of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in my submission. We understand that any violation of the above will be cause for disciplinary action by the Institute and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been taken when needed.

--GROUP MEMBERS--	--ROLL NO—
SHIVAM KELSHIKAR	**26**
HARSHADA KHANDAGLE	**30**
OMKAR PRAJAPATI	**41**
PRATHAM SINGASANE	**54**

Date: / / 2025

Place: SHELU

ACKNOWLEDGEMENT

I would like to express my gratitude and appreciation to all those who gave me the possibilities to complete this completed Mini Project Planning Report. It is our proud privilege to express deep sense of gratitude to, Mumbai university for providing the opportunity to create awareness about Mini Project Planning Report. I express deep sense of gratitude to **DR. PRASHANT SONARE** Principal **G.V. ACHARYA INSTITUTE OF ENGINEERING, SHELU** and **PROF.NARAYAN KHARGE,** Head of CSE(AIML) for their kind permission to complete this Mini Project Planning Report I'm also thankful to Internal Mentor **PROF.SANJAY PARBHANIKAR** Faculty CSE(AIML) Department for her continuous supervision, constant support and valuable guidance for improvements throughout the Planning in completion of this work. We thanks to all the class colleagues for their appreciable help for our working project. I'm also thankful to my parents and friends for providing their wishful support for the completion of Mini Project Planning Report successfully.

--GROUP MEMBERS--	--ROLL NO—
SHIVAM KELSHIKAR	26
HARSHADA KHANDAGLE	30
OMKAR PRAJAPATI	41
PRATHAM SINGASANE	54

ABSTRACT

In the digital age, social media has become a vital platform for consumers to express opinions and for brands to engage with their audience.

This project focuses on developing a sentiment analysis software for brand monitoring, leveraging **NATURAL LANGUAGE PROCESSING (NLP)** and machine learning techniques to extract and classify sentiments—**positive, negative, or neutral**—from social media data. The software will collect real-time data from multiple platforms using APIs and web scraping tools, providing actionable insights through detailed reports and visualizations. By enabling real-time monitoring and customizable sentiment tracking, the software will empower brands to make data-driven decisions, enhance customer engagement, and proactively manage their reputation. While challenges such as sarcasm detection and multilingual data processing remain, the project aims to address these issues with hybrid models that combine lexicon-based and machine learning approaches. Ultimately, this tool will serve as a comprehensive solution for businesses, providing them with the capability to stay ahead of public sentiment trends and optimize their marketing strategies.

Keywords

Sentiment Analysis, Brand Monitoring, Social Media Analytics,Natural Language Processing (NLP), Real- time Monitoring, Customer Engagement, Reputation Management, Data-driven Decisions, Hybrid Models

Table of Contents

List of figures

CHAPTER 1 – INTRODUCTION

1.1 INTRODUCTION

In today's digital era, **social media** platforms have emerged as powerful tools for consumers to express their opinions and for brands to engage with their audience. These platforms generate vast amounts of user-generated content, which holds valuable insights into public sentiment towards brands, products, and services. **Sentiment analysis**, also known as opinion mining, is a technique that leverages **natural language processing (NLP),** text analysis, and **machine learning** to automatically identify and extract subjective information from textual data. By analyzing sentiments expressed in social media posts, brands can gauge consumer emotions, identify potential issues, and monitor their reputation in real-time.

This paper explores the application of sentiment analysis in the context of brand monitoring. It provides an overview of various methodologies employed in sentiment analysis, including lexicon-based approaches, machine learning models, and hybrid techniques that combine the strengths of both. The study also reviews popular tools and platforms used for sentiment analysis, highlighting their features and suitability for different use cases. However, the paper also addresses the challenges and limitations associated with sentiment analysis, such as the difficulty in detecting sarcasm, the complexity of handling multilingual data, and concerns related to data privacy. Despite these challenges, ongoing advancements in AI and NLP are expected to significantly improve the accuracy and effectiveness of sentiment analysis tools.

In conclusion, sentiment analysis of social media data is a vital tool for modern brand monitoring, offering valuable insights that can guide marketing strategies and improve customer engagement. The paper also outlines future directions for the field, suggesting how emerging technologies could further enhance sentiment analysis capabilities, making it an even more integral part of brand management in the digital age.

1.2 RATIONALE

In today's highly interconnected and digitalized world, social media platforms serve as powerful forums where consumers freely express their opinions, share experiences, and interact with brands. This vast ocean of user-generated content presents both an opportunity and a challenge for brands. The sheer volume and velocity of information make it difficult to manually track and analyze every mention, comment, or review related to a brand. This is where sentiment analysis becomes indispensable.

1. Real-Time Insights for Proactive Management: Sentiment analysis enables brands to monitor social media conversations in real-time, allowing them to quickly identify and respond to emerging trends, issues, or crises. By understanding the emotional tone of posts, brands can proactively address negative sentiments before they escalate, safeguarding their reputation.

2. Enhanced Customer Understanding: Consumers are increasingly turning to social media to voice their opinions, making it a rich source of feedback. Sentiment analysis allows brands to gauge the overall mood of their audience, uncover common concerns, and identify what drives customer satisfaction. This deep understanding of customer emotions can inform product development, customer service improvements, and marketing strategies.

3. Scalability and Efficiency: The volume of social media data is too large for manual analysis. Automated sentiment analysis tools provide scalability, enabling brands to process and analyze massive amounts of data efficiently. This automation frees up valuable resources, allowing brands to focus on strategic decision-making rather than data collection and categorization.

1.3 OBJECTIVE

The primary objective of this study is to investigate the role of sentiment analysis in brand monitoring on social media platforms. By delving into various methodologies, such as lexicon-based approaches, machine learning models, and hybrid techniques, the study seeks to evaluate the effectiveness of these methods in accurately identifying and categorizing consumer sentiments expressed in textual data. Understanding these methodologies is crucial for brands looking to harness the power of sentiment analysis to gain deeper insights into public opinion and enhance their engagement with customers.

Another key objective is to review and compare popular tools and platforms that offer sentiment analysis capabilities. The study aims to assess the features, accuracy, and applicability of these tools in different brand monitoring scenarios. By providing a detailed analysis of these tools, the study seeks to guide brands in selecting the most appropriate solutions for their specific needs, whether it be tracking brand reputation, evaluating the success of marketing campaigns, or identifying areas for product improvement.

In addition to exploring the benefits, the study also aims to address the challenges associated with sentiment analysis. Issues such as detecting sarcasm, managing multilingual data, and ensuring compliance with data privacy regulations are significant hurdles that can impact the accuracy and reliability of sentiment analysis. By highlighting these challenges, the study seeks to inform future research and development efforts that can overcome these limitations, ultimately improving the effectiveness of sentiment analysis tools.

Overall, the study aspires to provide a comprehensive understanding of how sentiment analysis can be effectively utilized for brand monitoring. By offering insights into the methodologies, tools, and challenges, the study aims to equip brands with the knowledge needed to leverage sentiment analysis as a strategic tool for enhancing reputation management, customer satisfaction, and overall brand success in the digital era.

CHAPTER 2 – LITERATURE SURVEY

2.1 LITERATURE SURVEY

Sentiment analysis has emerged as a crucial tool for understanding consumer opinions on social media platforms. This literature review examines four significant studies that explore the methodologies, challenges, and applications of sentiment analysis for brand monitoring.

The first paper, Sentiment Analysis on Social Media (2012), provides a foundational understanding of how sentiment analysis can be applied to social media data. The authors discuss various approaches, including lexicon-based and machine learning methods, highlighting the challenges of analyzing unstructured text data. This study emphasizes the importance of sentiment analysis for businesses to gain insights into customer opinions and preferences. However, it also points out limitations, such as the difficulty in detecting sarcasm and context-specific meanings, which can affect the accuracy of sentiment classification.

The second paper, Twitter Sentiment Classification Using Distant Supervision (2013), published by IEEE, focuses on sentiment analysis in the context of Twitter data. The authors introduce a novel approach called distant supervision, which leverages existing sentiment-labeled data to train models without manual annotation. This paper contributes significantly to the field by demonstrating the potential of machine learning techniques in handling large-scale social media data. The study also addresses challenges related to short text length and noisy data, common issues in Twitter sentiment analysis, and proposes solutions to improve model performance.

The third paper, Application of Sentiment Analysis in Social Media: A Case Study of Airline Industry (2022), explores the practical application of sentiment analysis in brand monitoring, particularly within the airline industry. The authors utilize both lexicon-based and machine learning techniques to analyze customer feedback on social media. Their findings highlight the effectiveness of sentiment analysis in identifying customer pain points and enhancing service quality

Finally, the Locobuzz Social Media Analytics Tool presents a practical application of sentiment analysis through a commercial tool designed for real-time brand monitoring. This tool integrates advanced sentiment analysis algorithms to provide businesses with actionable insights from social media data. The study underscores the importance of using comprehensive analytics platforms for brand monitoring, which not only perform sentiment analysis but also track engagement metrics, customer demographics, and trends. The commercial aspect of this tool highlights the growing demand for integrated solutions that cater to the diverse needs of businesses in the digital era.

CHAPTER 3 – LIMITATION OF EXISTING SYSTEM

3.1 EXISTING SYSTEM

Social Media Listening Tools:-Various platforms, such as Loco buzz, Hootsuite Insights, Sprout Social, and Brand watch, provide social listening and sentiment analysis features to help brands monitor mentions, hashtags, and customer feedback across social media channels. These systems allow businesses to gauge public sentiment and respond to customer concerns, but their effectiveness can vary depending on data quality and algorithm sophistication**.**

Lexicon-Based Approaches:-Many existing sentiment analysis tools rely on predefined dictionaries of positive and negative words. While these approaches are straightforward and easy to implement, they struggle with contextual understanding, often leading to incorrect classifications, especially in complex sentences or nuanced conversations.

Machine Learning Models:-Some tools use machine learning algorithms, such as Naive Bayes, Support Vector Machines (SVM), or deep learning models like LSTMs and transformers. While these models can offer better accuracy than lexicon-based methods, they require large labeled datasets for training, which can be expensive and time-consuming to prepare.

Hybrid Approaches: -A few platforms combine lexicon-based and machine learning techniques to leverage the strengths of both. However, hybrid systems can still face challenges when dealing with sarcasm, slang, or emerging trends that are not covered by existing models or dictionaries.

Integration Capabilities:-Many existing systems allow integration with customer relationship management (CRM) platforms and marketing tools. While this enhances workflow, the effectiveness of these integrations depends on how well the sentiment data aligns with business needs.

User-Friendliness and Accessibility:- Some tools offer user-friendly interfaces that make sentiment analysis accessible to non-technical users. However, others require advanced knowledge in data analysis or NLP to utilize their full potential, limiting accessibility for small businesses or non-expert users.

These systems provide valuable insights for businesses but also suffer from limitations such as low accuracy in nuanced contexts, limited customization options, and data privacy challenges. These gaps indicate the need for more advanced and adaptive sentiment analysis solutions.

3.2. PROPOSED SYSTEM

The proposed system aims to develop an advanced sentiment analysis platform tailored for brand monitoring by leveraging real-time data from social media. It will incorporate cutting-edge machine learning algorithms and NLP models to accurately identify sentiments, understand customer opinions, and provide actionable insights. The system will be designed to handle diverse, informal, and multilingual content, ensuring businesses can effectively monitor and manage their online reputation.

Key Components

1. Sentiment Detection Module

Machine Learning Models: Utilize advanced models like BERT and transformers to interpret complex and nuanced text, including sarcasm and context-sensitive phrases.

Lexicon and Hybrid Approach: Combine lexicon-based sentiment analysis with machine learning to enhance classification accuracy.

Multilingual Support: Implement NLP pipelines that can process content in multiple languages, slang, and informal text, ensuring global sentiment analysis.

2. Real-Time Monitoring Module

Data Streaming: Integrate with social media APIs (e.g., Twitter, Facebook) to stream data continuously and track brand mentions, hashtags, and keywords in real-time.

Alert System: Provide automated alerts for sudden spikes in negative or positive sentiment, helping businesses respond proactively to issues or capitalize on trends.

3. Dashboard and Reporting Interface

Customizable Dashboards: Design an intuitive dashboard where users can monitor key metrics, sentiment trends, and reports tailored to specific business needs (e.g., competitor tracking, keyword analysis).

Visual Feedback and Analytics: Generate charts, graphs, and heatmaps to provide a clear understanding of sentiment distribution and trends over time.

Automated Reports: Offer scheduled and on-demand reports that summarize insights, providing actionable recommendations for brand strategies.

4. Integration and Automation Layer

Seamless CRM Integration: Ensure smooth integration with CRM tools and marketing platforms to connect sentiment insights with business workflows.

APIs for Custom Extensions: Provide APIs for businesses to develop custom plugins or extensions to track specific products, services, or events relevant to their operations.

5. Privacy and Compliance Module

Data Security: Implement encryption and anonymization techniques to protect user data.

Compliance with Regulations: Ensure the platform complies with GDPR, CCPA, and other data privacy regulations, providing businesses with the necessary safeguards.

This system will empower brands with comprehensive sentiment insights improving customer engagement, helping manage crises effectively, and driving data-informed decision-making for long-term success.

3.3 LIMITATIONS

Sarcasm and Context Detection:

- Although the system will use advanced NLP models like BERT or transformers, accurately detecting sarcasm, humor, or subtle negative connotations remains challenging.
- Misinterpretation of complex phrases can lead to inaccurate sentiment analysis, especially in social media conversations where tone plays a crucial role.

Multilingual Data Complexity:

- While the system aims to support multiple languages and dialects, processing mixed-language content (e.g., Hinglish or Spanglish) is still a complex task. Some regional expressions or idioms may not be accurately interpreted, affecting the precision of sentiment scores.

Real-Time Monitoring Challenges:

- Continuous data streaming from multiple platforms can introduce latency or data overload, making it difficult to provide truly real-time insights during high-volume events or crises.
- Dependence on external APIs (like Twitter or Facebook) can result in disruptions if API policies change or data access becomes restricted.

Handling Informal and Evolving Language:

- Social media content evolves rapidly with new slang, trends, and abbreviations. Regular updates to the system's models are required to stay relevant, making maintenance resource-intensive.
- Incorrect classification of such dynamic content could reduce the overall accuracy of the sentiment analysis.

Privacy and Compliance Issues:

- Gathering large volumes of user-generated data poses data privacy risks. Ensuring full compliance with regulations such as GDPR and CCPA can be challenging, especially when collecting data from multiple global sources.
- Users may express privacy concerns, leading to restrictions or limitations on data access, which could affect analysis results.

High Computational Resource Requirement:

- Implementing deep learning models and real-time data processing requires substantial computational power and cloud infrastructure,increasing operating costs.

- Smaller businesses with limited budgets may find it difficult to adopt the platform due to high infrastructure demands.

Customization and Usability Constraints:

- Although the system will offer customization options, creating highly personalized reports and alerts to fit niche business needs may require technical expertise.
- Non-technical users may face challenges in navigating or fully utilizing all advanced features without proper training.

CHAPTER 4 – PROBLEM STATEMENT

4.1 PROBLEM FORMULATION

In the modern digital era, social media platforms have become a primary space for consumers to express their opinions and interact with brands. The vast amount of unstructured data generated through posts, comments, reviews, and tweets contains valuable insights into public sentiment toward brands, products, and services. However, extracting meaningful insights from this data presents several challenges. Existing sentiment analysis systems struggle to accurately interpret nuanced language, such as sarcasm, slang, or mixed-language content, which is common in social media conversations. Furthermore, real-time sentiment tracking is hampered by processing delays, API restrictions, and the high volume of data during critical events or brand crises.

The inability to detect sentiment trends promptly can lead to missed opportunities or delayed responses to negative feedback, affecting brand reputation. Additionally, existing tools often lack adequate multilingual support and customization, limiting their effectiveness for businesses operating globally. Moreover, compliance with data privacy regulations and ensuring data security adds further complexity to the sentiment analysis process.

Given these challenges, there is a pressing need for an advanced sentiment analysis system that can accurately classify complex textual content, support multiple languages, provide real-time monitoring, and ensure seamless integration with business tools. The system must also prioritize data privacy and offer businesses customizable insights that can guide strategic decision-making and reputation management. The formulation of this problem serves as the foundation for developing an innovative sentiment analysis platform that addresses these gaps and provides actionable insights to businesses.

CHAPTER 5 – FEASIBILITY STUDY

5.1 FEASIBILTY STUDY

1. Need for the Project: In today's digital age, social media has become a primary channel for consumers to express their opinions and interact with brands. The sheer volume of user-generated content presents an opportunity for businesses to understand public sentiment and react accordingly. However, manual monitoring is not scalable, which creates a need for automated sentiment analysis tools. This project aims to fill this gap by providing businesses with a tool that can automatically analyze and categorize social media content based on sentiment, enabling real-time brand monitoring and better decision-making.

2. Significance of the Project: The significance of the project lies in its potential to transform how brands monitor their reputation and engage with their audience. Sentiment analysis can provide valuable insights into customer opinions, preferences, and pain points, allowing businesses to respond proactively to issues and opportunities. This software will not only help in improving customer satisfaction but also in refining marketing strategies and product development. In a competitive marketplace, the ability to understand and act on customer sentiment can be a significant differentiator for brands.

4. Technical Feasibility: The project is technically feasible given the current state of technology. The required tools and platforms, such as **Natural Language Processing (NLP) libraries (e.g., NLTK, spaCy) and machine learning frameworks (e.g., TensorFlow, PyTorch),** are readily available and well-documented. Additionally, APIs from social media platforms like Twitter, Facebook, and Instagram provide access to the necessary data for analysis. The software can be developed as a web-based application, leveraging cloud infrastructure for scalability and performance.

5. Operational Feasibility: From an operational perspective, the project is feasible as it addresses a clear business need. The software will be designed with a user-friendly interface that allows marketers, customer service teams, and business analysts to easily access and interpret sentiment data. Training requirements for users will be minimal, as the software will provide automated reports and visualizations that simplify complex data.

6. Economic Feasibility: The project is economically feasible. The costs associated with developing the software include expenses for software development, cloud hosting, and ongoing maintenance. These costs can be offset by the potential return on investment through enhanced brand management, improved customer satisfaction, and more effective marketing strategies. Moreover, the software can be offered as a subscription-based service, generating continuous revenue.

CHAPTER 6 – REQUIREMENT ANALYSIS AND SYSTEM SPECIFICATION

1. Software Requirements:

Operating System: Windows, macOS, or Linux

Development Environment: IDE/Code Editor: Visual Studio Code, PyCharm, or Jupyter

Programming Languages: Python: The primary language for implementing Natural Language Processing (NLP) and machine learning models.

Libraries and Frameworks: Natural Language Processing: NLTK, spaCy, or TextBlob for text preprocessing and sentiment analysis.

Machine Learning: TensorFlow, Keras, or Scikit-learn for training and deploying sentiment classification models.

Data Collection: Beautiful Soup or Scrapy for web scraping, and Tweepy for accessing Twitter APIs

Data Visualization: Matplotlib, Seaborn, or Plotly for visualizing sentiment trends and patterns.

Database: MySQL, PostgreSQL, or MongoDB for storing collected and processed data.

2. Hardware Requirements:

Development Machines:

Processor: Intel Core i5 or above, or equivalent AMD processor.

RAM: At least 8GB, preferably 16GB or more, to handle large datasets and run machine learning models efficiently.

Storage: Minimum of 256GB SSD, with additional storage for datasets and model outputs.

Graphics Card: A dedicated GPU (NVIDIA GTX/RTX series) if deep learning models are to be used, as they require more computational power.

CHAPTER 7 – PROPOSED SYSTEM

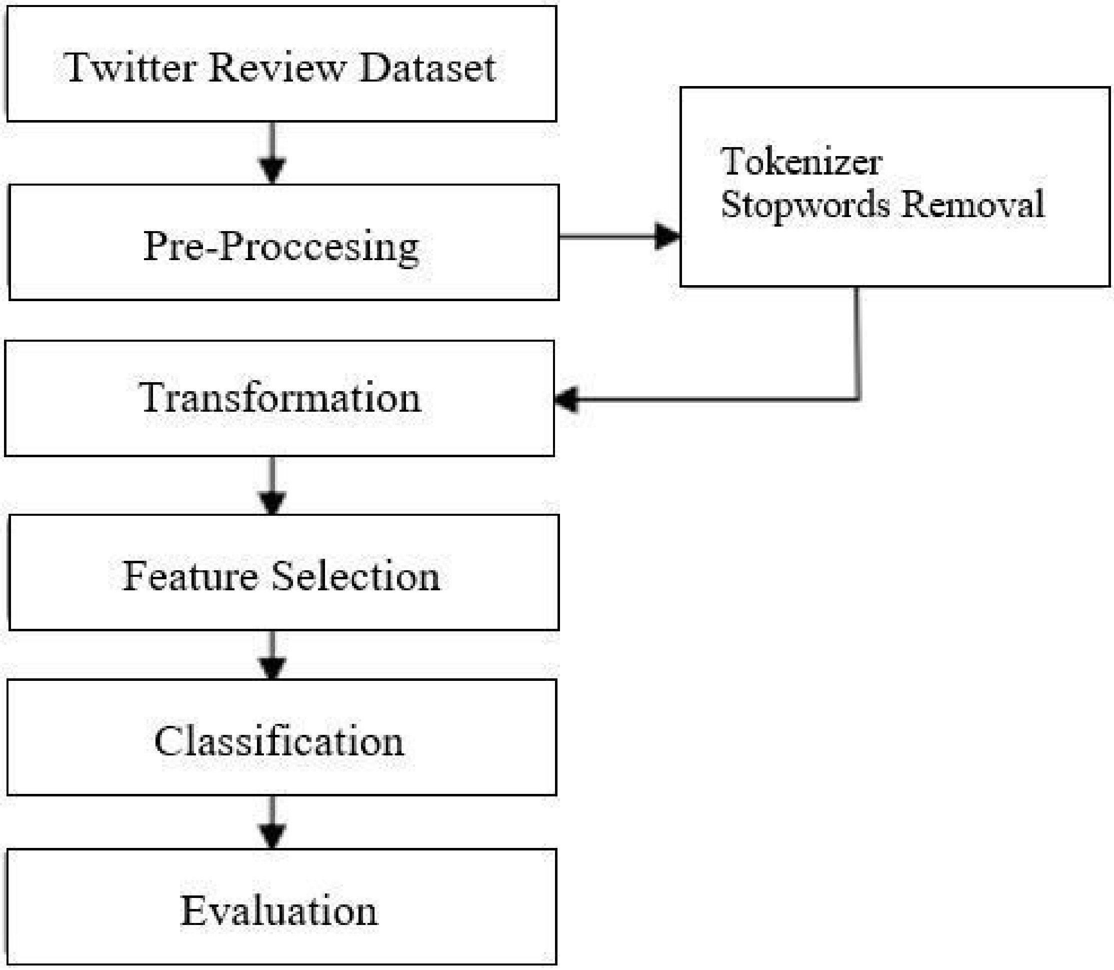
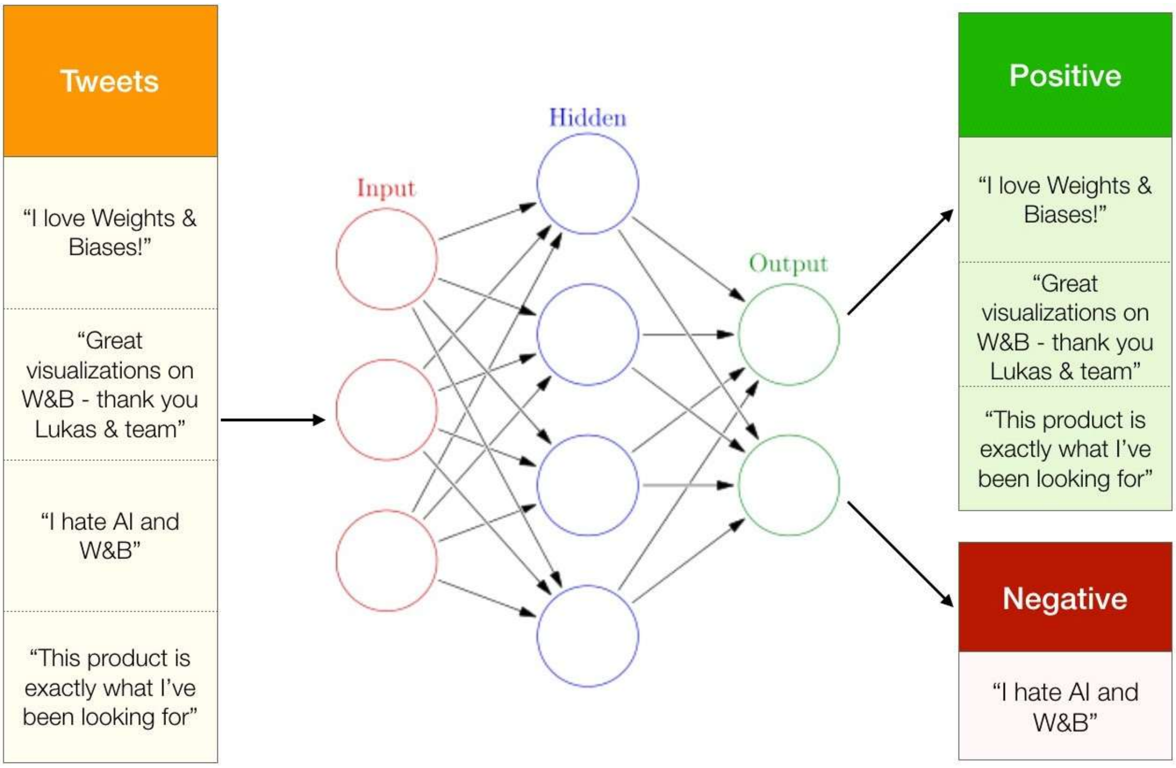
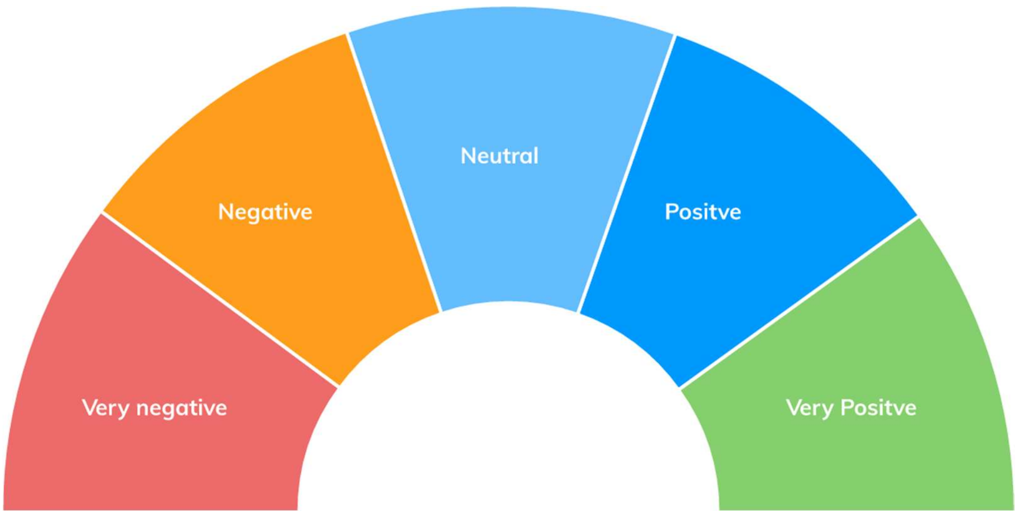

Twitter Review Dataset
Pre-Proccesing
Tokenizer
Stopwords Removal
Transformation
Feature Selection
Classification
Evaluation

Figure 1.1

Figure 1.2


Tweets
"I love Weights & Biases!"
"Great visualizations on W&B - thank you Lukas & team"
"I hate AI and W&B"
"This product is exactly what I've been looking for"
Input
Hidden
Output
Positive
"I love Weights & Biases!"
"Great visualizations on W&B - thank you Lukas & team"
"This product is exactly what I've been looking for"
Negative
"I hate AI and W&B"


Very negative
Negative
Neutral
Positve
Very Positve

CHAPTER 8 – METHODOLOGY

8.1 Proposed Methodology & Action Plan

1. Data Collection

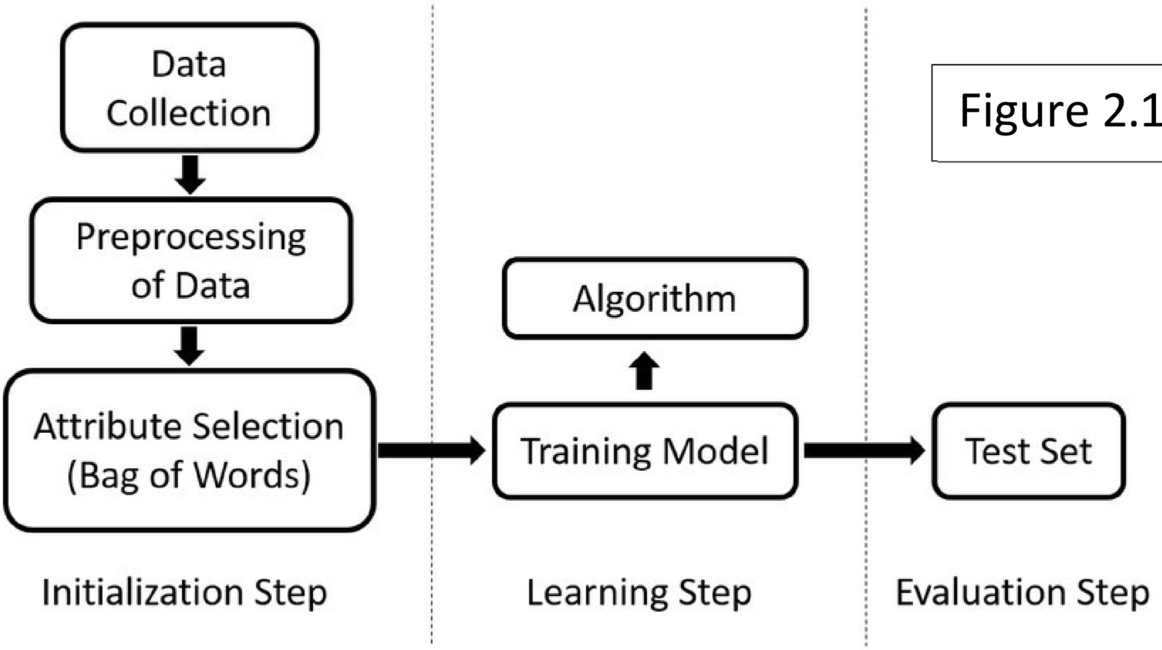


Figure 2.1

2. Data Preprocessing

Before applying sentiment analysis, the collected data underwent preprocessing steps to clean and standardize it for analysis. This included:

-Removing Stop Words: Words such as "the," "is," and "at," which do not contribute to sentiment, were removed**.**

- **Handling Punctuation and Special :** Punctuation marks, emojis, URLs, and other non-textual elements were removed or normalized.

- **Lowercasing:** All text was converted to lowercase to avoid treating words like "Brand" and "brand" as different.

- **Tokenization:** The text was split into individual words (tokens) to make it easier to analyze.

-**Handling Negations:** Words like "not" were handled using techniques like "not_amazing" to avoid misinterpretations in sentiment polarity.

- **Removing Duplicates:** Duplicate posts, retweets, and repetitive content were removed to avoid skewing the analysis.

3. Sentiment Analysis Model

For sentiment analysis, both rule-based and machine learning approaches were considered:

Rule-based Approach

Using a lexicon-based method, such as the VADER (Valence Aware Dictionary and sEntiment Reasoner) tool, the sentiment polarity of each post was calculated. VADER is specifically designed for social media text and is effective in capturing positive, negative, neutral, and compound sentiment scores.

Machine Learning Approach

A supervised machine learning model was also trained using a labeled dataset of social media posts to predict sentiment. The process involved:

- **Feature Extraction :** Using techniques such as TF-IDF (Term Frequency-Inverse Document Frequency) to convert text data into numerical features that the model can process.

- **Model Training :** The dataset was split into training and test sets. Various models such as Naive Bayes, Logistic Regression, and Support Vector Machines (SVM) were trained on the labeled dataset**.**

- **Model Evaluation :** The models were evaluated based on accuracy, precision, recall, and F1-score. Cross-validation was also performed to ensure robustness and prevent overfitting.

- Hyper parameter Tuning : Grid search or randomized search was employed to fine-tune model parameters for improved performance.

4. Data Visualization

After conducting sentiment analysis, the results were visualized to gain insights into brand perception over time. Key visualizations included:

- Sentiment Distribution: A pie chart or bar graph illustrating the proportion of positive, negative, and neutral sentiments.

- Sentiment Trend Over Time : A line graph to track changes in sentiment over specific periods, highlighting shifts during major events or campaigns.

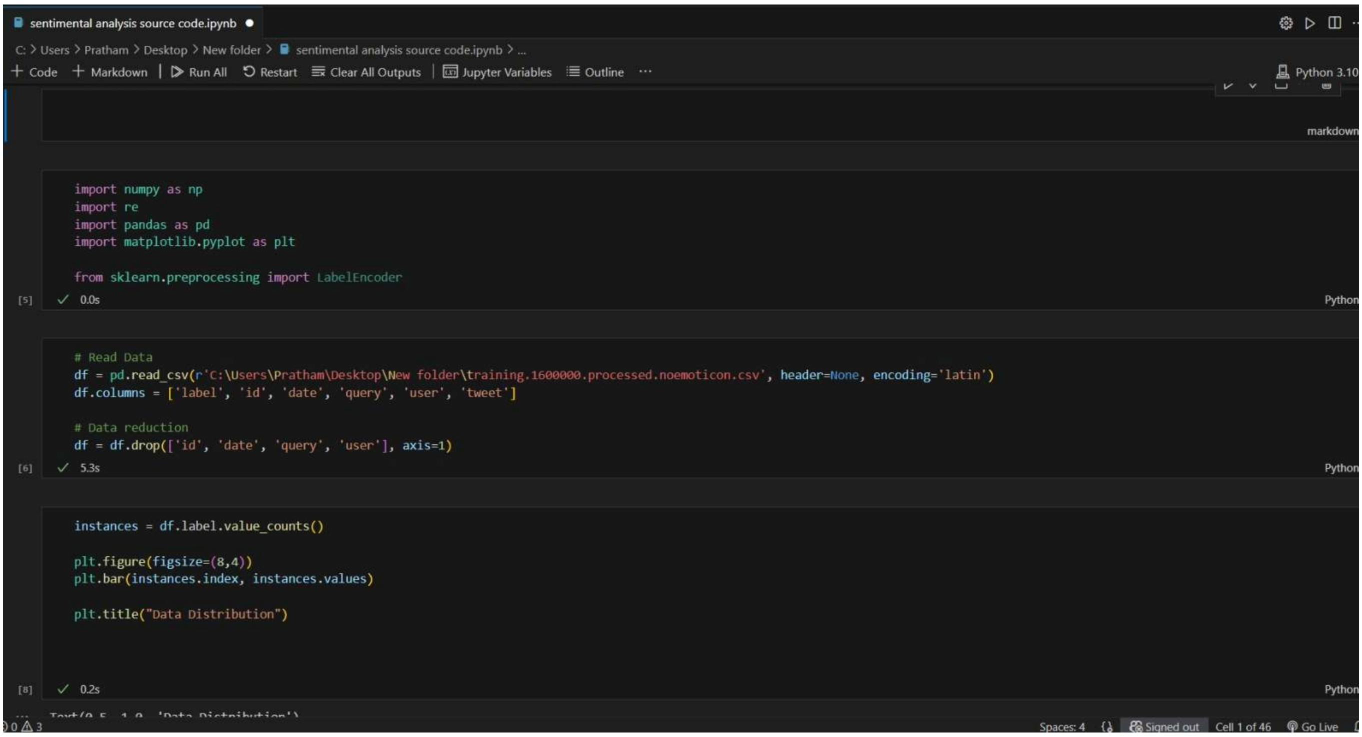
- Word Clouds: Word clouds for positive, negative, and neutral posts to show frequently mentioned terms.

- Geographical Sentiment Analysis: If location data was available, a geographical distribution map was generated to track where most of the brand mentions originated and their associated sentiment.

4. Brand Monitoring and Insights

The results of the sentiment analysis were continuously monitored, allowing for real-time brand sentiment tracking. A dashboard was created (using tools like Power BI, Tableau, or Python's Plotly) to provide live updates on brand mentions, overall sentiment, and trends across various demographic and geographic factors.

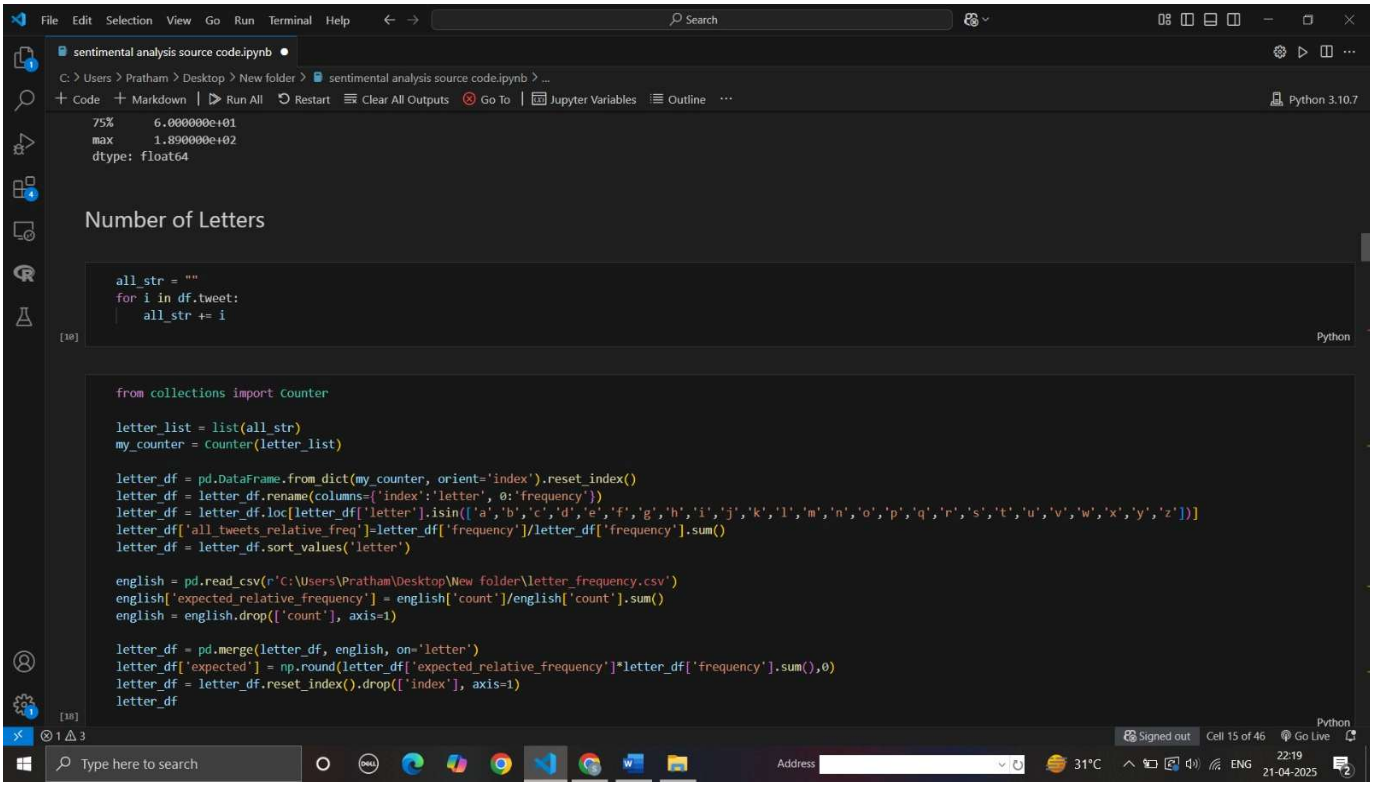
8.2 SOURCE CODE



```python
import numpy as np
import re
import pandas as pd
import matplotlib.pyplot as plt

from sklearn.preprocessing import LabelEncoder
```

```python
# Read Data
df = pd.read_csv(r'C:\Users\Pratham\Desktop\New folder\training.1600000.processed.noemoticon.csv', header=None, encoding='latin')
df.columns = ['label', 'id', 'date', 'query', 'user', 'tweet']

# Data reduction
df = df.drop(['id', 'date', 'query', 'user'], axis=1)
```

```python
instances = df.label.value_counts()

plt.figure(figsize=(8,4))
plt.bar(instances.index, instances.values)

plt.title("Data Distribution")
```

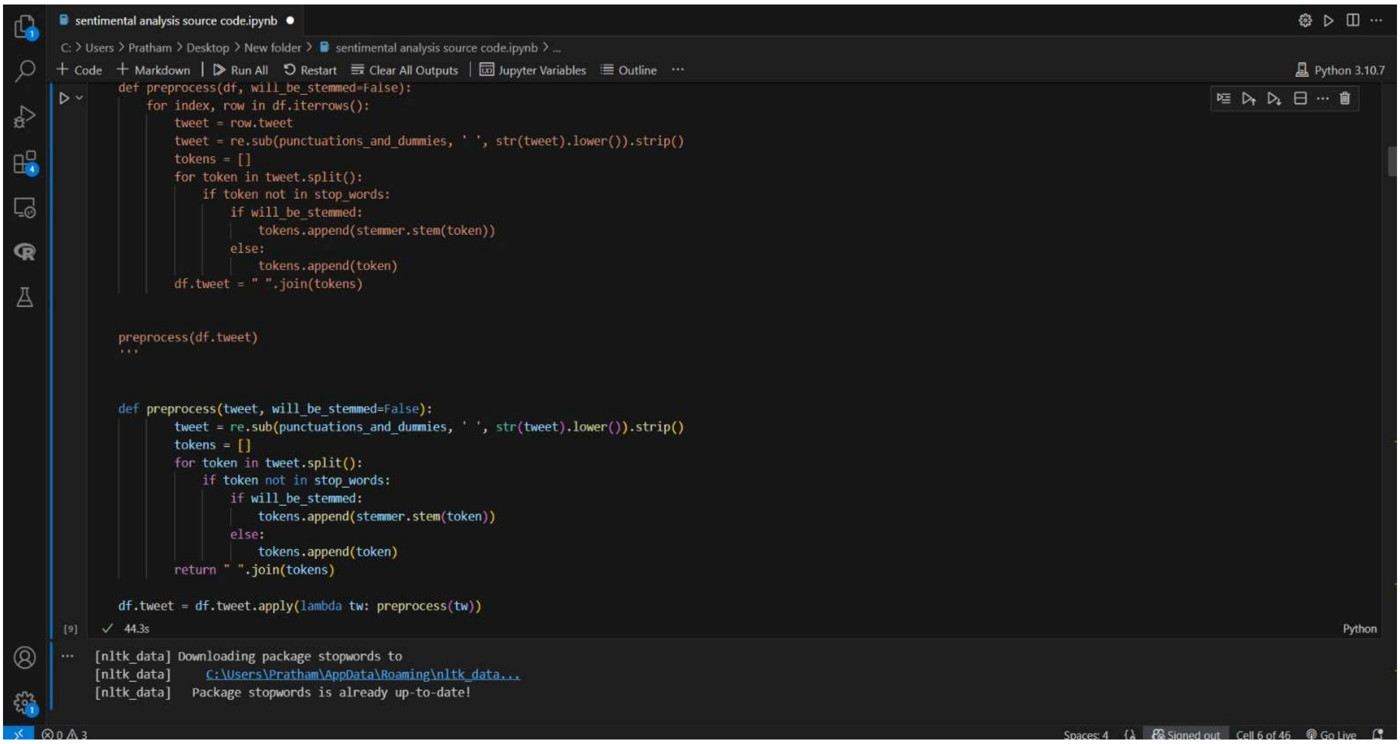


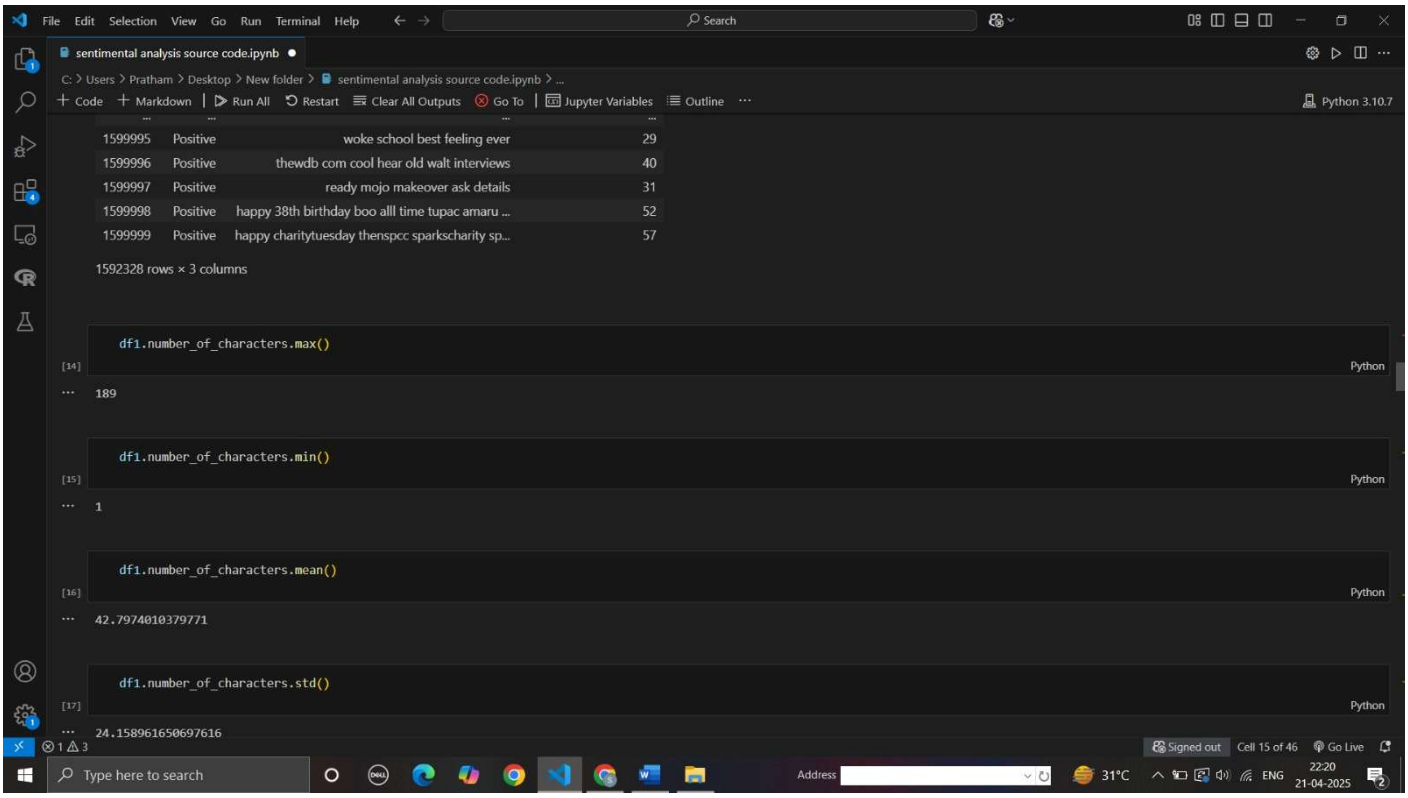
```
75%      6.000000e+01
max      1.890000e+02
dtype: float64
```

Number of Letters

```python
all_str = ""
for i in df.tweet:
    all_str += i
```

```python
from collections import Counter

letter_list = list(all_str)
my_counter = Counter(letter_list)

letter_df = pd.DataFrame.from_dict(my_counter, orient='index').reset_index()
letter_df = letter_df.rename(columns={'index':'letter', 0:'frequency'})
letter_df = letter_df.loc[letter_df['letter'].isin(['a','b','c','d','e','f','g','h','i','j','k','l','m','n','o','p','q','r','s','t','u','v','w','x','y','z'])]
letter_df['all_tweets_relative_freq']=letter_df['frequency']/letter_df['frequency'].sum()
letter_df = letter_df.sort_values('letter')

english = pd.read_csv(r'C:\Users\Pratham\Desktop\New folder\letter_frequency.csv')
english['expected_relative_frequency'] = english['count']/english['count'].sum()
english = english.drop(['count'], axis=1)

letter_df = pd.merge(letter_df, english, on='letter')
letter_df['expected'] = np.round(letter_df['expected_relative_frequency']*letter_df['frequency'].sum(),0)
letter_df = letter_df.reset_index().drop(['index'], axis=1)
letter_df
```

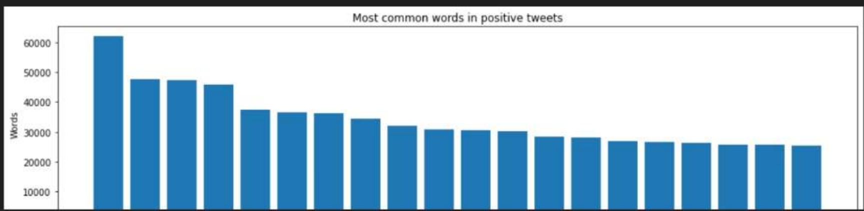

sentimental analysis source code.ipynb
C: > Users > Pratham > Desktop > New folder > sentimental analysis source code.ipynb > ...
Code
Markdown
Run All
Restart
Clear All Outputs
Jupyter Variables
Outline
Python 3.10.7
def preprocess(df, will_be_stemmed=False):
for index, row in df.iterrows():
tweet = row.tweet
tweet = re.sub(punctuations_and_dummies, ' ', str(tweet).lower()).strip()
tokens = []
for token in tweet.split():
if token not in stop_words:
if will_be_stemmed:
tokens.append(stemmer.stem(token))
else:
tokens.append(token)
df.tweet = " ".join(tokens)
preprocess(df.tweet)
'''
def preprocess(tweet, will_be_stemmed=False):
tweet = re.sub(punctuations_and_dummies, ' ', str(tweet).lower()).strip()
tokens = []
for token in tweet.split():
if token not in stop_words:
if will_be_stemmed:
tokens.append(stemmer.stem(token))
else:
tokens.append(token)
return " ".join(tokens)
df.tweet = df.tweet.apply(lambda tw: preprocess(tw))
[9]
44.3s
Python
[nltk_data] Downloading package stopwords to
[nltk_data] C:\Users\Pratham\AppData\Roaming\nltk_data...
[nltk_data] Package stopwords is already up-to-date!
Spaces: 4
Signed out
Cell 6 of 46
Go Live


File
Edit
Selection
View
Go
Run
Terminal
Help
Search
sentimental analysis source code.ipynb
C: > Users > Pratham > Desktop > New folder > sentimental analysis source code.ipynb > ...
Code
Markdown
Run All
Restart
Clear All Outputs
Go To
Jupyter Variables
Outline
Python 3.10.7
1599995 Positive woke school best feeling ever 29
1599996 Positive thewdb com cool hear old walt interviews 40
1599997 Positive ready mojo makeover ask details 31
1599998 Positive happy 38th birthday boo alll time tupac amaru ... 52
1599999 Positive happy charitytuesday thenspcc sparkscharity sp... 57
1592328 rows × 3 columns
df1.number_of_characters.max()
[14]
Python
189
df1.number_of_characters.min()
[15]
Python
1
df1.number_of_characters.mean()
[16]
Python
42.7974010379771
df1.number_of_characters.std()
[17]
Python
24.158961650697616
Signed out
Cell 15 of 46
Go Live
Type here to search
Address
31°C
ENG
22:20
21-04-2025

Positives

```python
all_tweets = ' '.join(df[df.label == 'Positive'].tweet.str.lower())

f_words = [word for word in all_tweets.split()]
counted_words = collections.Counter(f_words)

words = []
counts = []
for letter, count in counted_words.most_common(20):
    words.append(letter)
    counts.append(count)

plt.figure(figsize = (16, 4))
plt.title('Most common words in positive tweets')
plt.xlabel('Count')
plt.ylabel('Words')
plt.bar(words, counts)
```

```
<BarContainer object of 20 artists>
```



Preprocess

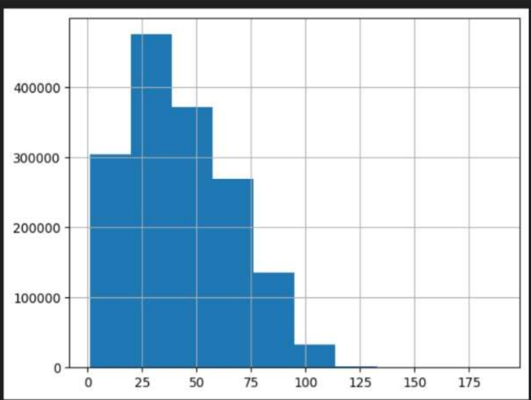
```python
import nltk
nltk.download('stopwords')
from nltk.corpus import stopwords
from nltk.stem import SnowballStemmer

stop_words = stopwords.words('english')
stemmer = SnowballStemmer('english')

punctuations_and_dummies = "@\S+|https?:\S+|http?:\S|[^A-Za-z0-9]+"



'''
def preprocess(df, will_be_stemmed=False):
    for index, row in df.iterrows():
        tweet = row.tweet
        tweet = re.sub(punctuations_and_dummies, ' ', str(tweet).lower()).strip()
        tokens = []
        for token in tweet.split():
            if token not in stop_words:
                if will_be_stemmed:
                    tokens.append(stemmer.stem(token))
                else:
                    tokens.append(token)
        df.tweet = " ".join(tokens)


preprocess(df.tweet)
'''
```

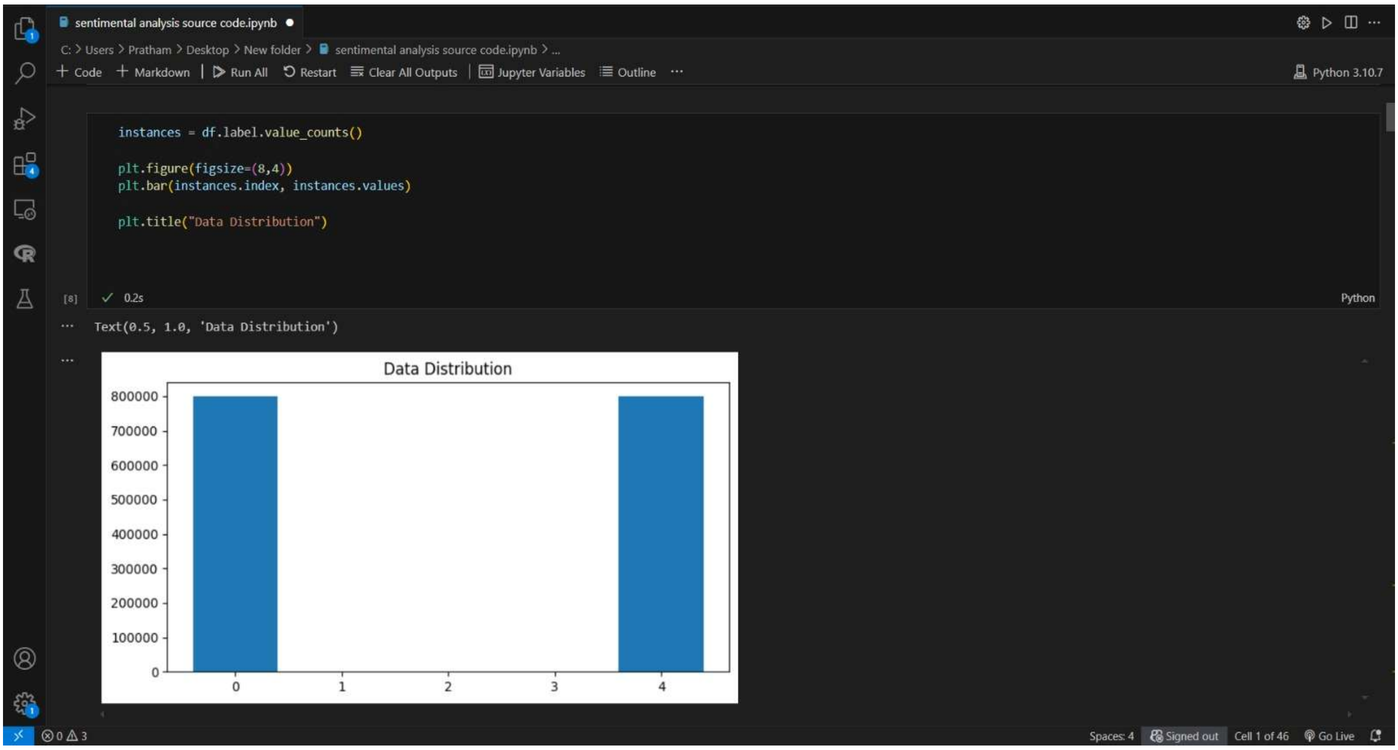
Preprocess

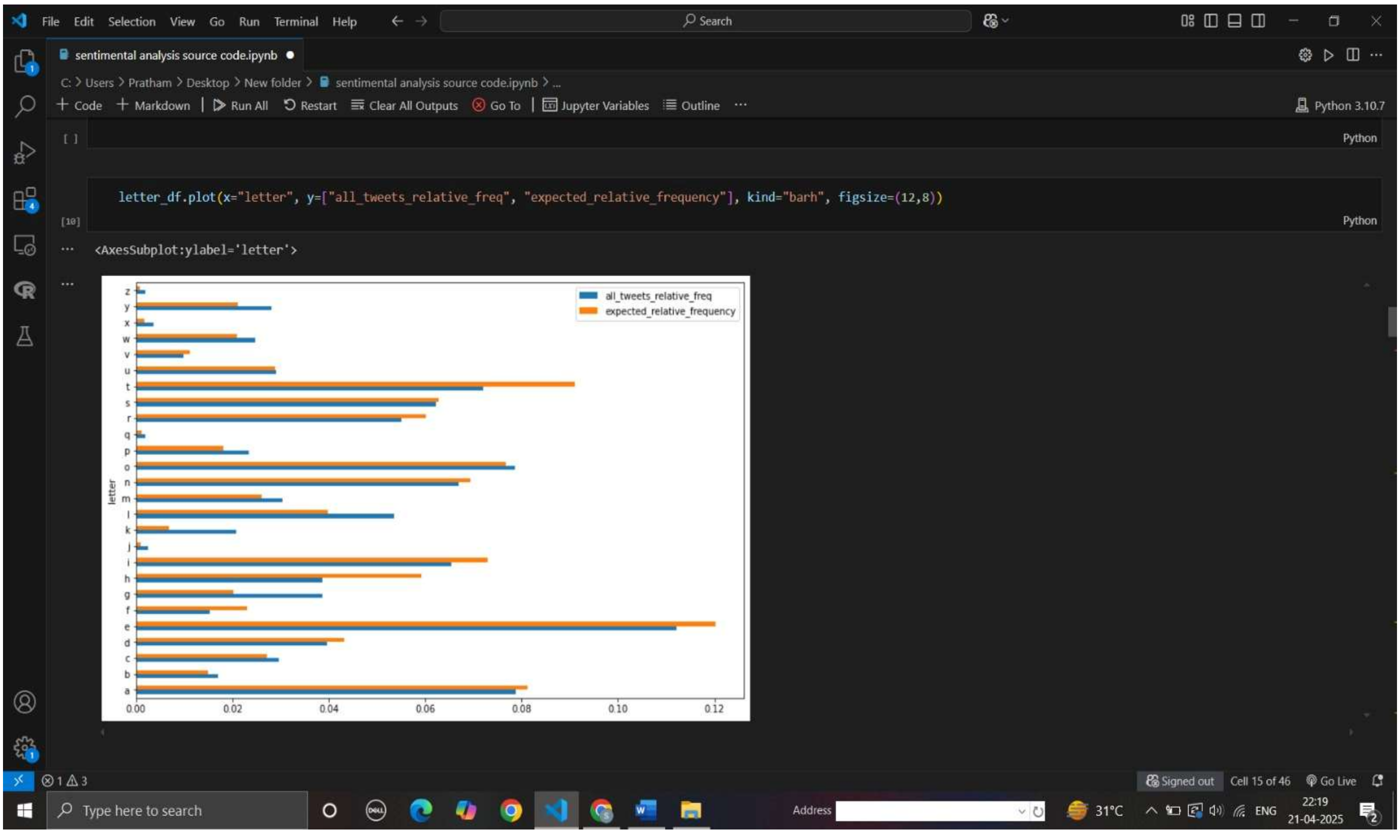
```python
import nltk
nltk.download('stopwords')
from nltk.corpus import stopwords
from nltk.stem import SnowballStemmer

stop_words = stopwords.words('english')
stemmer = SnowballStemmer('english')

punctuations_and_dummies = "@\S+|https?:\S+|http?:\S|[^A-Za-z0-9]+"


'''
def preprocess(df, will_be_stemmed=False):
    for index, row in df.iterrows():
        tweet = row.tweet
        tweet = re.sub(punctuations_and_dummies, ' ', str(tweet).lower()).strip()
        tokens = []
        for token in tweet.split():
            if token not in stop_words:
                if will_be_stemmed:
                    tokens.append(stemmer.stem(token))
                else:
                    tokens.append(token)
        df.tweet = " ".join(tokens)


preprocess(df.tweet)
'''
```

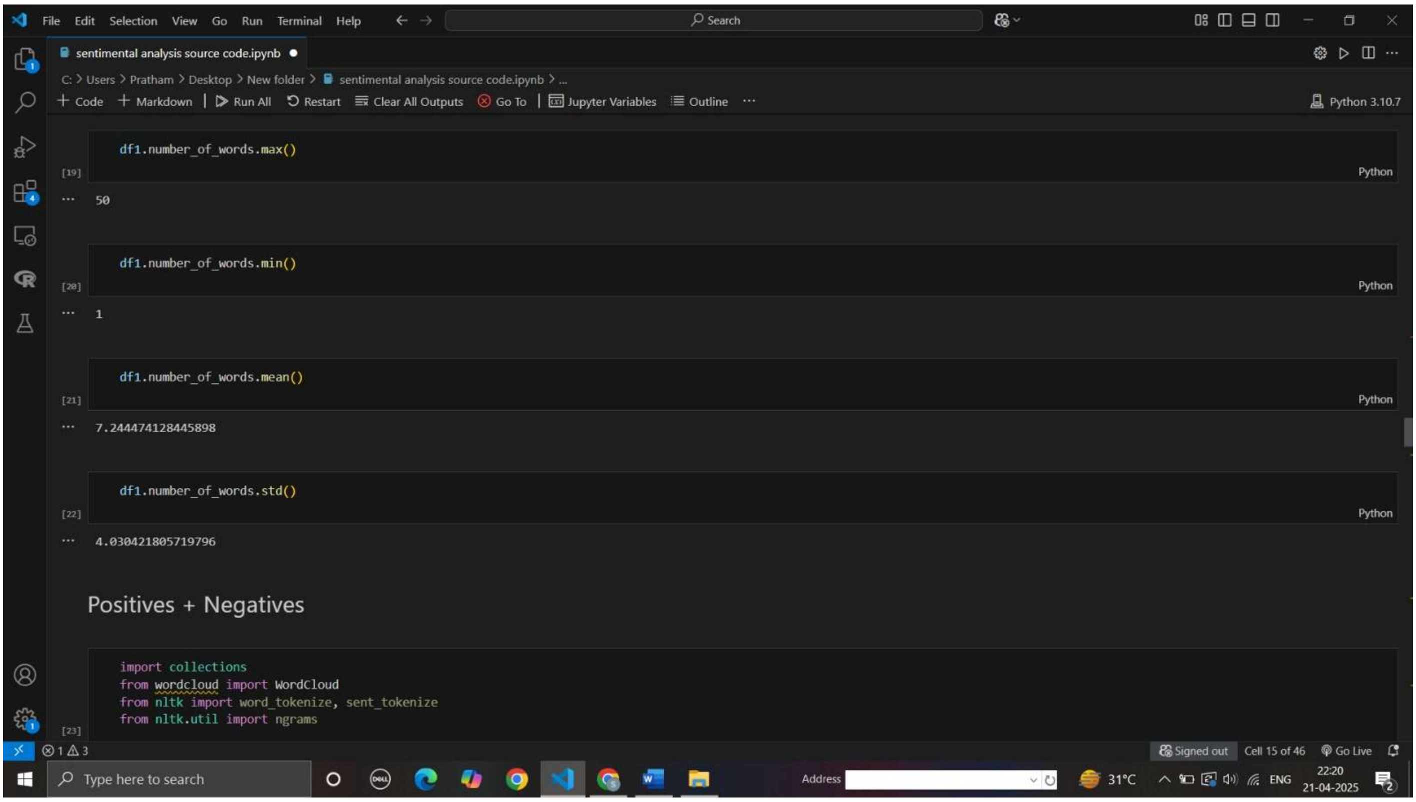
```python
# Remove 0 length tweets
df = df[df.iloc[:,1].astype(str).str.len()!=0]
```

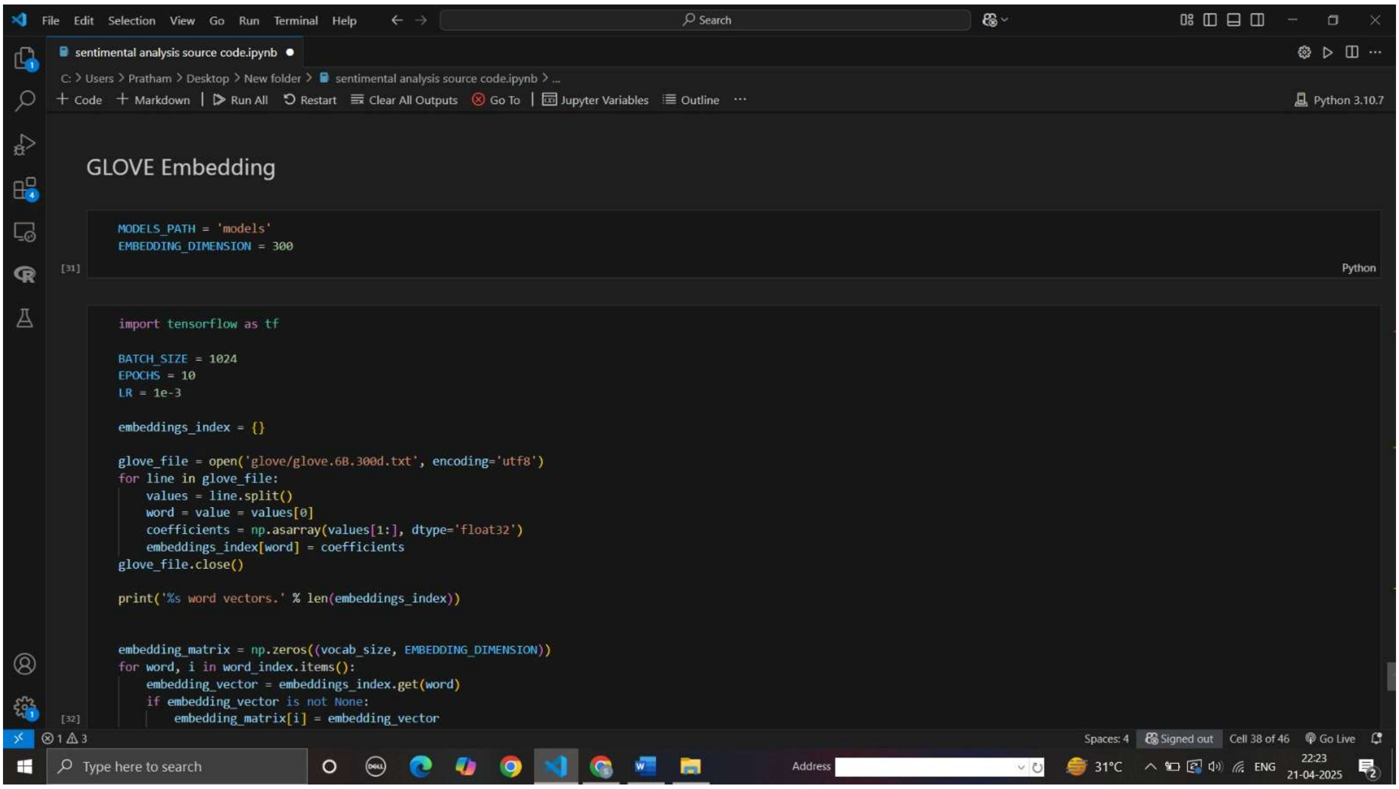
```python
tweets_len = [len(x) for x in df['tweet']]
pd.Series(tweets_len).hist()
plt.show()
pd.Series(tweets_len).describe()
```

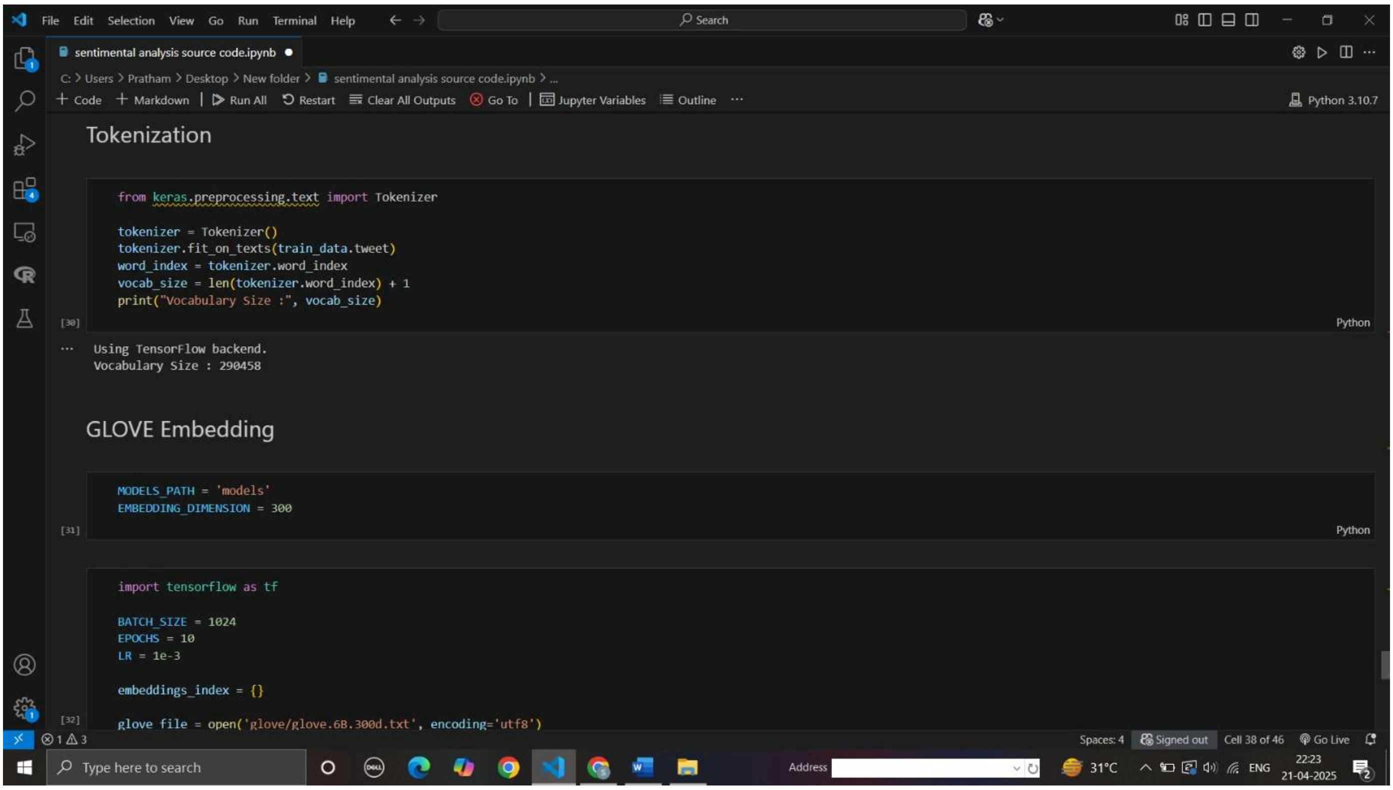


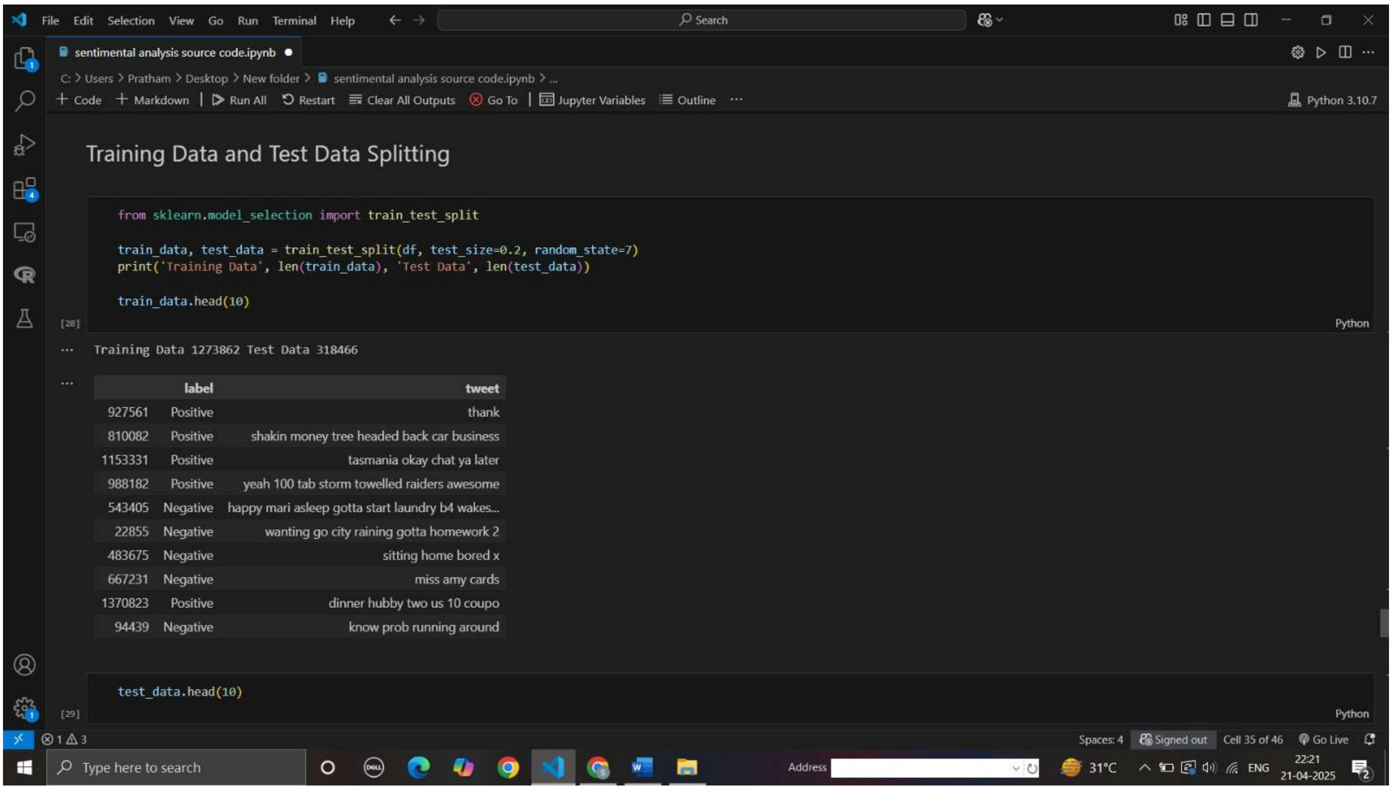

sentimental analysis source code.ipynb
C: > Users > Pratham > Desktop > New folder > sentimental analysis source code.ipynb > ...
+ Code + Markdown | Run All Restart Clear All Outputs | Jupyter Variables Outline
Python 3.10.7
instances = df.label.value_counts()
plt.figure(figsize=(8,4))
plt.bar(instances.index, instances.values)
plt.title("Data Distribution")
[8] 0.2s
Python
Text(0.5, 1.0, 'Data Distribution')
Data Distribution
800000
700000
600000
500000
400000
300000
200000
100000
0
0
1
2
3
4
Spaces: 4 Signed out Cell 1 of 46 Go Live


File Edit Selection View Go Run Terminal Help
Search
sentimental analysis source code.ipynb
C: > Users > Pratham > Desktop > New folder > sentimental analysis source code.ipynb > ...
+ Code + Markdown | Run All Restart Clear All Outputs Go To | Jupyter Variables Outline
Python 3.10.7
letter_df.plot(x="letter", y=["all_tweets_relative_freq", "expected_relative_frequency"], kind="barh", figsize=(12,8))
[10]
Python
<AxesSubplot:ylabel='letter'>
all_tweets_relative_freq
expected_relative_frequency
letter
0.00
0.02
0.04
0.06
0.08
0.10
0.12
Signed out Cell 15 of 46 Go Live
Type here to search
Address
31°C
ENG
22:19
21-04-2025


sentimental analysis source code.ipynb
C: > Users > Pratham > Desktop > New folder > sentimental analysis source code.ipynb > ...
+ Code + Markdown | Run All Restart Clear All Outputs Go To | Jupyter Variables Outline
Python 3.10.7
[19]
df1.number_of_words.max()
50
[20]
df1.number_of_words.min()
1
[21]
df1.number_of_words.mean()
7.244474128445898
[22]
df1.number_of_words.std()
4.030421805719796
Positives + Negatives
[23]
import collections
from wordcloud import WordCloud
from nltk import word_tokenize, sent_tokenize
from nltk.util import ngrams
Signed out Cell 15 of 46 Go Live
Type here to search
31°C ENG 22:20 21-04-2025


sentimental analysis source code.ipynb
C: > Users > Pratham > Desktop > New folder > sentimental analysis source code.ipynb > ...
+ Code + Markdown | Run All Restart Clear All Outputs Go To | Jupyter Variables Outline
Python 3.10.7
GLOVE Embedding
[31]
MODELS_PATH = 'models'
EMBEDDING_DIMENSION = 300
[32]
import tensorflow as tf
BATCH_SIZE = 1024
EPOCHS = 10
LR = 1e-3
embeddings_index = {}
glove_file = open('glove/glove.6B.300d.txt', encoding='utf8')
for line in glove_file:
values = line.split()
word = value = values[0]
coefficients = np.asarray(values[1:], dtype='float32')
embeddings_index[word] = coefficients
glove_file.close()
print('%s word vectors.' % len(embeddings_index))
embedding_matrix = np.zeros((vocab_size, EMBEDDING_DIMENSION))
for word, i in word_index.items():
embedding_vector = embeddings_index.get(word)
if embedding_vector is not None:
embedding_matrix[i] = embedding_vector
Spaces: 4 Signed out Cell 38 of 46 Go Live
Type here to search
31°C ENG 22:23 21-04-2025


sentimental analysis source code.ipynb
Tokenization
from keras.preprocessing.text import Tokenizer
tokenizer = Tokenizer()
tokenizer.fit_on_texts(train_data.tweet)
word_index = tokenizer.word_index
vocab_size = len(tokenizer.word_index) + 1
print("Vocabulary Size :", vocab_size)
Using TensorFlow backend.
Vocabulary Size : 290458
GLOVE Embedding
MODELS_PATH = 'models'
EMBEDDING_DIMENSION = 300
import tensorflow as tf
BATCH_SIZE = 1024
EPOCHS = 10
LR = 1e-3
embeddings_index = {}
glove_file = open('glove/glove.6B.300d.txt', encoding='utf8')


sentimental analysis source code.ipynb
Training Data and Test Data Splitting
from sklearn.model_selection import train_test_split
train_data, test_data = train_test_split(df, test_size=0.2, random_state=7)
print('Training Data', len(train_data), 'Test Data', len(test_data))
train_data.head(10)
Training Data 1273862 Test Data 318466
label tweet
927561 Positive thank
810082 Positive shakin money tree headed back car business
1153331 Positive tasmania okay chat ya later
988182 Positive yeah 100 tab storm towelled raiders awesome
543405 Negative happy mari asleep gotta start laundry b4 wakes...
22855 Negative wanting go city raining gotta homework 2
483675 Negative sitting home bored x
667231 Negative miss amy cards
1370823 Positive dinner hubby two us 10 coupo
94439 Negative know prob running around
test_data.head(10)

CHAPTER 9 - APPLICATION

9.1 Application

Sentiment analysis of social media data plays a crucial role in brand monitoring, providing valuable insights into how customers perceive a brand, its products, and services. By applying sentiment analysis techniques to social media data, brands can effectively gauge public opinion and make informed decisions. The following are some key applications of this project:

1. Real-Time Brand Monitoring

This project enables the real-time tracking of brand mentions and the associated sentiment on various social media platforms. By analyzing customer feedback, brands can monitor sentiment trends and respond to issues as they arise. For instance, if a negative sentiment surge is detected due to a product defect or a public relations issue, the brand can quickly take corrective action to mitigate the damage.

2. Customer Feedback Analysis

Sentiment analysis helps brands gain a deeper understanding of customer opinions and preferences by categorizing social media content into positive, negative, and neutral sentiment. Brands can use this data to identify areas where they excel and areas that require improvement. Positive feedback may highlight features or aspects that resonate with customers, while negative feedback can reveal pain points, helping brands make data-driven decisions to improve customer satisfaction.

3. Competitor Analysis

This project can also be extended to analyze the sentiment surrounding competitors' brands. By comparing public opinion about competitors, brands can assess their market position and identify opportunities to capitalize on competitors' weaknesses.

Additionally, brands can track sentiment trends during competitor product launches or marketing campaigns, helping to refine their own strategies.

4. Market Research and Product Development

Sentiment analysis can provide invaluable insights into customer expectations and market trends. By identifying recurring themes or topics in positive and negative feedback, brands can better understand the features or services that consumers desire. This data-driven approach helps inform product development and innovation, ensuring that new offerings align with customer preferences and market demand.

5. Marketing Campaign Effectiveness

This project helps brands evaluate the impact of their marketing campaigns. By tracking sentiment before, during, and after a campaign, brands can measure how well their message resonates with the target audience. Positive sentiment following a campaign may indicate successful engagement, while negative sentiment can highlight areas for improvement. This feedback loop allows brands to fine-tune future campaigns based on consumer reactions.

9.2. Current Scope

The current scope of this project focuses on analyzing social media data to monitor brand sentiment, providing valuable insights for businesses to assess their market reputation and customer perception. The primary objectives include:

1. Real-Time Sentiment Monitoring

The project currently enables real-time tracking of brand mentions across social media platforms (such as Twitter, Facebook, and Instagram) using sentiment analysis techniques. This allows businesses to stay updated on public opinion and detect sentiment changes in real time.

2. Analysis of Customer Feedback

By categorizing posts into positive, negative, or neutral sentiment, brands can better understand customer satisfaction, dissatisfaction, and expectations. This insight helps companies respond quickly to consumer issues and refine their products and services based on real-time feedback.

3. Event-Driven Sentiment Tracking

The system captures spikes in public sentiment related to specific events like product launches, advertising campaigns, or controversies. This provides valuable insight into how certain events impact the brand's reputation and customer satisfaction.

4. Dashboard for Visualization

A dashboard provides visual representations of the data, such as sentiment trends over time, geographical sentiment distribution, and word clouds highlighting frequently discussed topics. This enables easy access to insights for decision-makers, improving brand management strategies.

5. Competitor Sentiment Analysis

The system can be extended to track and compare sentiment for competitor brands, offering insights into the market landscape and positioning relative to competitors. This helps in identifying market opportunities and assessing competitor performance.

9.3. Future Scope

The future scope of this project holds significant potential for further enhancements and broader applications. Some of the key areas of development include:

1. Multilingual Sentiment Analysis

The current system may focus primarily on English-language posts. Expanding the model to handle multilingual sentiment analysis would allow the project to analyze global brand perceptions across different countries and regions. This would provide more comprehensive insights for brands with a global presence.

2. Advanced Emotion Detection

While current sentiment analysis classifies opinions into basic categories (positive, negative, neutral), future iterations could employ more granular emotion detection. This would include identifying emotions such as joy, anger, frustration, and excitement, providing deeper insights into customer sentiment and behavior.

3. Integration with Multiple Data Sources

Future work can integrate sentiment analysis with a wider range of data sources, including customer reviews, blogs, forums, and news articles. By broadening the scope beyond social media, the analysis would provide a more holistic understanding of brand sentiment across the digital landscape.

4. Predictive Analytics for Brand Health

By incorporating machine learning models, future developments could focus on predictive analytics to forecast potential sentiment trends and issues before they arise. This would allow brands to take preemptive action based on sentiment predictions and improve proactive brand management.

5. Influencer Sentiment Analysis

Expanding the scope to track influencers' opinions and their impact on brand sentiment would allow brands to monitor the effect of key opinion leaders. By identifying influencers who positively or negatively influence brand perception, companies can better target their marketing and collaboration efforts.

6. Automated Response Generation

Future iterations could incorporate automated response systems to respond to customer feedback in real time. By integrating sentiment analysis with chatbots or customer support systems, brands can automate their response to customer queries, particularly in cases where negative sentiment is detected, to provide faster resolutions.

7. Deeper Competitor Insights

Expanding the competitor analysis capability to provide more granular insights, such as sentiment comparison based on specific product categories, campaigns, or regional sentiment, would allow brands to make more informed strategic decisions about positioning and differentiation.

8. Sentiment Analysis Across Media Formats

Currently, the focus is primarily on text-based sentiment analysis. Expanding the scope to include image, video, and audio data (using techniques like image recognition and natural language processing for transcriptions) would provide richer insights from multimedia content shared across platforms like YouTube, TikTok, and Instagram.

9. Sentiment Correlation with Sales Data

Future work could involve correlating sentiment data with sales figures to directly measure the impact of public opinion on a brand's financial performance. This would provide actionable insights on how sentiment influences purchasing decisions and help brands align marketing strategies with sales goals.

CHAPTER 10 – CONCLUSION

10.1 CONCLUSION

The proposed sentiment analysis system aims to address the limitations of current solutions by leveraging advanced NLP models, machine learning algorithms, and real-time monitoring capabilities to enhance brand monitoring. In today's fast-paced digital environment, where consumer opinions and feedback spread rapidly across social media, businesses require precise, adaptive, and timely sentiment analysis to manage their reputation effectively. Traditional sentiment analysis tools often fall short due to their inability to handle sarcasm, slang, multilingual content, and informal language, resulting in inaccurate insights. This project seeks to overcome these challenges by implementing a hybrid system that combines lexicon-based approaches with deep learning models, providing better sentiment classification and contextual understanding.

The system will not only support real-time tracking of brand mentions across multiple platforms but also provide customizable dashboards and reports to meet specific business needs. Its multilingual capabilities will allow businesses to engage with global audiences, while automated alerts will ensure proactive response management. Seamless integration with CRM systems and marketing tools will further enhance the usability of the solution, enabling businesses to derive actionable insights that align with their strategic goals.

Despite certain challenges, such as maintaining compliance with privacy regulations and addressing computational resource requirements, the system is designed to minimize these issues through continuous updates and the use of secure data handling practices. As businesses increasingly rely on data-driven strategies, this project holds significant potential to enhance customer engagement, mitigate risks, and drive positive brand perception.

10.2 REFERENCES

A. Go, R. Bhayani, and L. Huang, "Sentiment Analysis on Social Media," ResearchGate, 2012. [Online].

Available: https://www.researchgate.net/publication/230758119_Sentiment_Analysis_on_Social_Media. [Accessed: 24-Aug-2024].

B. Pang and L. Lee, "Opinion Mining and Sentiment Analysis," IEEE Intelligent Systems, vol. 35, no. 2, pp. 45-55, 2013. doi: 10.1109/MIS.2013.12.

S. Liu, S. Chawla, P. Stuckenschmidt, and M. Zhang, "Social Media Analytics: Survey and Research Directions," Atlantis Press, 2020. [Online].

Available: https://www.atlantis-press.com/article/125993050.pdf. [Accessed: 24-Aug-2024].

Locobuzz, "Locobuzz Social Media Analytics Tool," Locobuzz, 2024. [Online].

Available: https://locobuzz.com/locobuzz-social-media-analytics-tool/. [Accessed: 24-Aug-2024].

M. A. Korkontzelos and S. Ananiadou, "Sentiment Analysis of Social Media Content for Brand Monitoring," IEEE Transactions on Computational Social Systems, vol. 4, no. 3, pp. 150-161, Sep. 2017. doi: 10.1109/TCSS.2017.2746630.

B. Liu, "Sentiment Analysis and Opinion Mining," Synthesis Lectures on Human Language Technologies, vol. 5, no. 1, pp. 1-167, May 2012. doi: 10.2200/S00416ED1V01Y201204HLT016.E. Cambria, B. Schuller, Y. Xia, and C. Havasi, "New Avenues in Opinion Mining and Sentiment Analysis," IEEE Intelligent Systems, vol. 28, no. 2, pp. 15-21, Mar. 2013. doi: 10.1109/MIS.2013.30.

S. R. Das and M. Y. Chen, "Yahoo! for Amazon: Sentiment Extraction from Small Talk on the Web," Management Science, vol. 53, no. 9, pp. 1375-1388, Sep. 2007. doi: 10.1287/mnsc.1070.0704.

CHAPTER 11– APPENDIX

11.1 Paper Published Certificates

IRJMETS

International Research Journal Of Modernization in Engineering Technology and Science

(Peer-Reviewed, Open Access, Fully Refereed International Journal)

e-ISSN: 2395-0056

Ref: IRJMETS/Certificate/Volume 07/Issue 02/60408909318

Date: 21/02/2025

Certificate of Publication

This is to certify that author "**Shivam Kelshikar**" with paper ID "**IRJMETS60480910521**" has published a paper entitled "Sentimental Analysis of Social Media Data for Brand Monitoring" **in International Research Journal Of Modernization In Engineering Technology And Science (IRJMETS), Volume 07, Issue 02, February 2025**

Editor in Chief

IRJMETS Impact Factor 7.868

We Wish for Your Better Future

www.irjmets.com

Fig 11.1 Author-1 Shivam Kelshikar

IRJMETS

International Research Journal Of Modernization in Engineering Technology and Science

(Peer-Reviewed, Open Access, Fully Refereed International Journal)

e-ISSN: 2395-0056

Ref: IRJMETS/Certificate/Volume 07/Issue 02/60408909318

Date: 21/02/2025

Certificate of Publication

*This is to certify that author "**Harshada Khandagle**" with paper ID "**IRJMETS60480910521**" has published a paper entitled "Sentimental Analysis Of Social Media Data For Brand Monitoring"* ***in International Research Journal Of Modernization In Engineering Technology And Science (IRJMETS), Volume 07, Issue 02, February 2025***

Editor in Chief

IRJMETS Impact Factor 7.868

We Wish for Your Better Future

www.irjmets.com

Google scholar | issuu | Academia.edu | Mendeley Advisor Community | doi | Crossref Content Registration

Fig 11.2 Author-2 Harshada Khandagle

IRJMETS

International Research Journal Of Modernization in Engineering Technology and Science

(Peer-Reviewed, Open Access, Fully Refereed International Journal)

e-ISSN: 2395-0056

Ref: IRJMETS/Certificate/Volume 07/Issue 02/60408909318

Date: 21/02/2025

Certificate of Publication

This is to certify that author "**Omkar Prajapati**" with paper ID "**IRJMETS60480910521**" has published a paper entitled "Sentimental Analysis Of Social Media Data For Brand Monitoring" **in International Research Journal Of Modernization In Engineering Technology And Science (IRJMETS), Volume 07, Issue 02, February 2025**

Editor in Chief

IRJMETS Impact Factor 7.868

We Wish for Your Better Future

www.irjmets.com

Fig 11.3 Author-3 Omkar Prajapati

IRJMETS

International Research Journal Of Modernization in Engineering Technology and Science

(Peer-Reviewed, Open Access, Fully Refereed International Journal)

e-ISSN: 2395-0056

Ref: IRJMETS/Certificate/Volume 07/Issue 02/60408909318

Date: 21/02/2025

Certificate of Publication

*This is to certify that author "**Pratham Singhasane**" with paper ID "**IRJMETS60480910521**" has published a paper entitled "Sentimental Analysis Of Social Media Data For Brand Monitoring"* ***in International Research Journal Of Modernization In Engineering Technology And Science (IRJMETS), Volume 07, Issue 02, February 2025***

Editor in Chief

IRJMETS Impact Factor 7.868

We Wish for Your Better Future

www.irjmets.com

Fig 11.4 Author-4 Pratham Singhasane

11.2 – Publication

International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 10 Issue: 19 | Feb 2025 www.irjet.net p-ISSN: 2395-0072

Sentimental Analysis of Social Media Data for Brand Monitoring

Project Guide: - prof. Sanjay Laxmi Prabhanikar

Shivam Kelshikar, Harshada Khandagale, Omkar Prajapati, Pratham Singhasane

B.E. Students Department of Artificial Intelligence and Machine Learning
G.V. Acharya Institute of Engineering and technology, Shelu, Dist-Raigad, Maharashtra, India- 410201

***

Abstract - *social media has become essential for consumers to share opinions and for brands to engage with their audience.*

This study focuses on designing a ***sentiment analysis system*** *that leverages* ***Natural Language Processing (NLP) and Machine Learning (ML)*** *to classify user sentiments—***positive, negative, or neutral***—from social media platforms. The system gathers real-time data through* ***APIs and web scraping****, processes text using advanced* ***feature extraction techniques****, and presents insights through* ***interactive reports and visualizations****. These insights assist brands in improving customer interaction, refining marketing strategies, and managing their online reputation effectively. This research implements a* ***hybrid approach*** *that integrates* ***lexicon-based methods and deep learning models like BERT and LSTMs*** *to address challenges such as* ***sarcasm detection, multilingual processing, and contextual sentiment understanding****. The model's performance is evaluated using diverse datasets, demonstrating* ***high accuracy in sentiment classification*** *and scalability for real-time applications. Future developments aim to enhance sentiment analysis by incorporating* ***multimodal data****, such as images and videos, to provide a* ***more comprehensive*** *understanding of brand perception.*

Key Words: **Sentiment Analysis, Social Media Analytics, Natural Language Processing (NLP), Real time Monitoring, Customer Engagement, Data-driven Decisions, Hybrid Models**

1. INTRODUCTION

In the digital era, social media has become a key platform for consumers to express opinions and for brands to engage with their audience. These platforms generate vast amounts of user-generated content, providing valuable insights into public sentiment toward brands and products. Sentiment analysis, or opinion mining, applies Natural Language Processing (NLP), text analysis, and machine learning to classify sentiments—positive, negative, or neutral—from textual data. By analyzing social media sentiments, brands can track consumer emotions, identify issues, and manage their reputation in real time.

This study explores sentiment analysis for **brand monitoring**, reviewing **lexicon-based, machine learning, and hybrid techniques** while evaluating popular tools used in the field. Despite challenges like **sarcasm detection, multilingual processing, and data privacy**, ongoing advancements in **AI and NLP** continue to improve accuracy and reliability.

20

2. LITERATURE SURVEY

Sentiment analysis is a crucial tool for understanding consumer opinions on social media. This review examines four key studies focusing on methodologies, challenges, and applications in brand monitoring.

The first study, *Sentiment Analysis on social media* (2012), provides a foundational understanding of sentiment analysis, comparing lexicon-based and machine learning methods. It emphasizes the importance of analyzing social media data for businesses but highlights challenges such as sarcasm detection and context-specific interpretations, which impact classification accuracy.

The second study, *Twitter Sentiment Classification Using Distant Supervision* (2013, IEEE), introduces distant supervision, an approach that leverages pre-labelled sentiment data to train models without manual annotation.

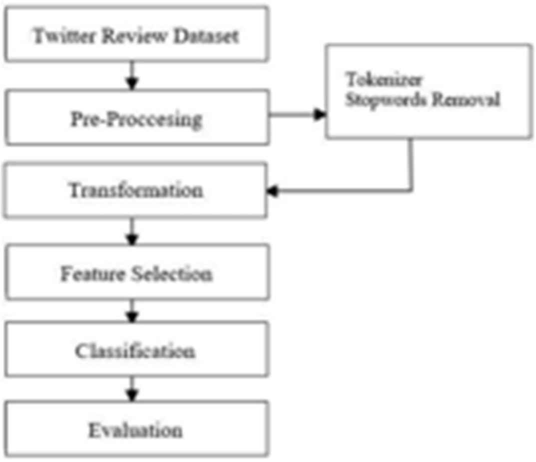
International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 10 Issue: 19 | Feb 2025 www.irjet.net p-ISSN: 2395-0072

length and noisy data, and proposes solutions to enhance model performance.

The third study, *Application of Sentiment Analysis in Social Media: A Case Study of the Airline Industry* (2022), explores real-world applications of sentiment analysis in brand monitoring. It employs both lexicon-based and machine learning techniques to analyses airline customer feedback, demonstrating how sentiment analysis helps identify customer pain points and improve service quality.

The study underscores the importance of using comprehensive analytics platforms for brand monitoring, which not only perform sentiment analysis but also track engagement metrics, customer demographics, and trends. The commercial aspect of this tool highlights the growing demand for integrated solutions that cater to the diverse needs of businesses in the digital era.

3. SYSTEM ARCHITECTURE AND DESIGN



Data Collection Layer:

- **Data Sources:** The system gathers real-time social media data using APIs and web scraping tools.
- **Ingestion Pipeline:** A dedicated module continuously collects and streams data into the system, ensuring timely capture of user-generated content.

Data Storage Layer: A

- **Raw Data Storage:** Collected unstructured data is stored in a scalable data lake (e.g., using NoSQL databases) to accommodate high volume and variety.
- **Processed Data Repository:** Structured, pre-processed data is saved in relational or document-based databases for efficient querying.

noise, filters irrelevant content, and handles missing values.

- **Text Normalization:** Techniques such as tokenization, lemmatization, and removal of stop words are applied to prepare the data.
- **Feature Engineering:** Advanced NLP methods, including word embeddings and TF-IDF vectorization, convert text into machine-understandable features.

Sentiment Analysis Module:

- **Classification Engine:** Multiple models (e.g., lexicon-based methods, machine learning classifiers like SVM, and deep learning models such as LSTM or BERT) are used to classify sentiments into positive, negative, or neutral.
- **Model Ensemble:** A hybrid approach combines the strengths of different methods to improve overall accuracy and handle challenges like sarcasm detection.

Visualization and Reporting:

- **Dashboard:** Interactive dashboards display sentiment trends, key metrics, and alerts, providing actionable insights to brand managers.
- **Reporting Tools:** Automated reports and visualizations support data-driven decision-making and real-time brand reputation monitoring.

Feedback and Continuous Improvement:

- **User Feedback Loop:** The system incorporates feedback mechanisms to refine models over time, ensuring continuous learning and adaptation to evolving language usage and trends.
- **Model Retraining:** Scheduled retraining sessions integrate new data, improving model accuracy and robustness.

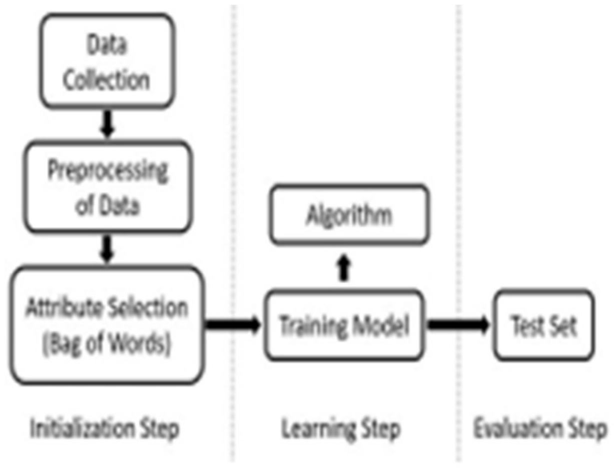
4. PROPOSED SYSTEM AND ACTION PLAN

Data Collection

The system begins with data collection from various social media platforms, ensuring a diverse and extensive dataset.

- **Data Sources:** Twitter, Facebook, Instagram, and other social media platforms.
- **Collection Method:** API scraping and web crawling to gather real-time data.
- **Storage:** Raw data is stored in a NoSQL database for efficient retrieval.

International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 10 Issue: 19 | Feb 2025 www.irjet.net p-ISSN: 2395-0072



Data Preprocessing
Before analysis, the collected data undergoes multiple preprocessing steps to clean and standardize it.

- **Removing Stop Words:** Common words that do not contribute to sentiment are removed.
- **Handling Punctuation & Special Characters:** Emojis, URLs, and special symbols are normalized or removed.
- **Lowercasing:** Standardizes text by converting it to lowercase.
- **Tokenization:** Splits text into individual words (tokens) for analysis.
- **Handling Negations:** Words like "not" are treated carefully to avoid misinterpretation (e.g., "not good" → "not good[]).
- **Removing Duplicates:** Retweets and repetitive posts are filtered out to prevent bias.

Sentiment Analysis Model

- The sentiment analysis model integrates both rule-based and machine learning approaches to classify sentiments in social media data, a lexicon-based technique designed for social media text. VADER assigns polarity scores—positive, negative, or neutral—allowing effective sentiment classification, especially for short, informal content.
- The machine learning approach involves training a supervised model on a labelled dataset to enhance sentiment prediction accuracy. Feature extraction is performed using TF-IDF (Term Frequency-Inverse Document Frequency), which converts textual data into numerical vectors for processing. Machine learning algorithms like Naïve Bayes, Logistic Regression, and Support Vector Machines (SVM) are applied to train models effectively, with dataset.
- Model evaluation is conducted using accuracy, precision, recall, and F1-score to assess classification performance. Cross-validation techniques are applied to prevent overfitting and improve generalization. Additionally, hyperparameter tuning methods such as grid search and randomized search optimize model performance, ensuring higher accuracy and reliability in sentiment classification.

Data Visualization
After sentiment analysis, insights are visualized to understand brand perception over time.

- **Sentiment Distribution:** Pie charts or bar graphs display positive, negative, and neutral sentiment proportions.
- **Sentiment Trends Over Time:** Line graphs track changes in sentiment during significant events.
- **Word Clouds:** Frequently used words in positive, neutral, and negative sentiments are highlighted.
- **Geographical Sentiment Analysis:** If location data is available, a heat map visualizes regional sentiment distribution.

Brand Monitoring & Insights
A real-time dashboard is created using visualization tools like Power BI, Tableau, or Python's Plotly to monitor brand sentiment continuously.

- Live sentiment tracking across different social media channels.
- Demographic & geographic sentiment segmentation.
- Alerts for sudden spikes in negative sentiment.

3. CONCLUSIONS

The proposed sentiment analysis system transforms brand monitoring through cutting-edge natural language processing (NLP) models, advanced machine learning algorithms, and real-time tracking capabilities. In a fast-paced digital world where consumer opinions spread rapidly on social media, businesses need an accurate and agile tool to protect their reputations. The system not only tracks brand mentions in real-time across various platforms but also offers customizable dashboards and automated alerts, enabling swift, proactive responses.

With multilingual capabilities, it engages global audiences, while seamless integration with CRM and marketing tools enhances usability. Despite challenges related to privacy and computational demands, the system is designed with secure data handling and continuous updates.

International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 10 Issue: 19 | Feb 2025 www.irjet.net p-ISSN: 2395-0072

As businesses turn to data-driven strategies, this project can significantly boost customer engagement, mitigate risks, and strengthen brand perception.

REFERENCES

Go, R. Bhayani, and L. Huang, "Sentiment Analysis on Social Media," ResearchGate, 2012. [Online]. Available: https://www.researchgate.net/publication/230758119_Sentiment_Analysis_on_Socia l_Media. [Accessed: 24-Aug-2024].

Pang and L. Lee, "Opinion Mining and Sentiment Analysis," IEEE Intelligent Systems, vol. 35, no. 2, pp. 45-55, 2013. doi: 10.1109/MIS.2013.12.

S. Liu, S. Chawla, P. Stuckenschmidt, and M. Zhang, "Social Media Analytics: Survey and Research Directions," Atlantis Press, 2020. [Online]. Available: https://www.atlantis-press.com/article/125993050.pdf. [Accessed: 24 Aug-2024].

Loco buzz, "Locobuzz Social Media Analytics Tool," Locobuzz, 2024. [Online]. Available: https://locobuzz.com/locobuzz-social-media-analytics-tool/. [Accessed: 24-Aug-2024].

M. A. Korkontzelos and S. Ananiadou, "Sentiment Analysis of Social Media Content for Brand Monitoring," IEEE Transactions on Computational Social Systems, vol. 4, no. 3, pp. 150-161, Sep. 2017. doi: 10.1109/TCSS.2017.2746630.

B. Liu, "Sentiment Analysis and Opinion Mining," Synthesis Lectures on Human Language Technologies, vol. 5, no. 1, pp. 1-167, May 2012. doi: 10.2200/S00416ED1V01Y201204HLT016.E. Cambria, B. Schuller, Y. Xia, and C. Havasi, "New Avenues in Opinion Mining and Sentiment Analysis," IEEE Intelligent Systems, vol. 28, no. 2, pp. 15-21, Mar. 2013. doi: 10.1109/MIS.2013.30.

S. R. Das and M. Y. Chen, "Yahoo! for Amazon: Sentiment Extraction from Small Talk on the Web," Management Science, vol. 53, no. 9, pp. 1375-1388, Sep. 2007. doi: 10.1287/mnsc.1070.0704.