



Mahavir Education Trust's
SHAH & ANCHOR KUTCHHI ENGINEERING COLLEGE
Chembur, Mumbai - 400 088
UG Program in Information Technology

- **Student Name: Purva Mahamunkar**
- **Roll Number: 12/28**
- **PRN : 22UF17661IT090**
- **Department: IT**
- **Batch/Division: BTECH-1**
- **Experiment Title: Medical Reviews Analysis from Social Media Data**
- **Date: 11/10/2025**



Mahavir Education Trust's
SHAH & ANCHOR KUTCHHI ENGINEERING COLLEGE
Chembur, Mumbai - 400 088
UG Program in Information Technology

Experiment No. 7				
Date of Performance:	11/10/2025			
Date of Submission:	11/10/2025			
Program Execution/ formation / correction/ ethical practices	Timely Submission	Viva	Experiment Total	Sign with Date

Experiment No 7

7.1 Aim: Medical Reviews Analysis from Social Media Data

7.2 Course Outcome: Assess various NLP techniques used for healthcare applications using suitable methods on real-world datasets.

7.3 Learning Outcome: Analyze patient reviews and sentiments from social media platforms for public health insights.

7.4 Requirement: Python (with TensorFlow/PyTorch), Jupyter/Colab, OpenCV, NumPy, Matplotlib, and a labeled Diseases Risk Patient dataset.

7.5 Related Theory:

1. Introduction

- In the digital era, millions of people share their health experiences, medicine feedback, and treatment opinions on social media platforms such as Twitter, Reddit, Facebook, and health forums.
- Analyzing this data helps researchers, doctors, and healthcare organizations understand public perception, treatment effectiveness, and patient satisfaction in real time.

2. Concept of Social Media Data Analysis

- **Social media data** is unstructured, meaning it consists of free-text posts, comments, emojis, and hashtags.
- Using **Natural Language Processing (NLP)** and **Machine Learning (ML)** techniques, we can extract valuable insights such as sentiment (positive, negative, neutral), trending medical issues, and common side effects mentioned by users.



Mahavir Education Trust's
SHAH & ANCHOR KUTCHHI ENGINEERING COLLEGE
Chembur, Mumbai - 400 088
UG Program in Information Technology

3. Data Collection

Data can be collected using **APIs** (e.g., Twitter API) or from publicly available datasets on platforms like **Kaggle**.

- Collected data typically includes user posts, timestamps, hashtags, and mentions related to medicines, diseases, or healthcare topics.

4. Text Preprocessing

- Since social media text contains noise (emojis, URLs, spelling errors), preprocessing is crucial.
- Common preprocessing steps include:
 - Removing stop words, punctuation, and special symbols
 - Tokenization and lemmatization
 - Lowercasing text
 - Handling hashtags and mentions

5. Feature Extraction

- To make text understandable to machines, we convert it into numerical form using techniques like:
 - **Bag of Words (BoW)**
 - **TF-IDF (Term Frequency–Inverse Document Frequency)**
 - **Word2Vec or BERT embeddings** for contextual understanding

6. Sentiment Analysis

- The core of this experiment is **sentiment classification**—identifying whether a user's review is positive, negative, or neutral
- Pre-trained models like **VADER**, **TextBlob**, or **transformer-based models (BERT)** are often used for social media text due to their ability to understand context and emotion.

7. Entity Extraction (Optional)

- Apart from sentiment, entities like **medicine names**, **diseases**, and **side effects** can be extracted using **Named Entity Recognition (NER)** to gain deeper medical insights.

8. Visualization & Interpretation

- The results are often visualized using graphs such as bar charts or word clouds to show the distribution of sentiments, commonly discussed drugs, or trending health concerns.

9. Applications

- **Pharmacovigilance:** Detecting adverse drug reactions (ADR) early.
- **Public Health Monitoring:** Tracking disease outbreaks through user discussions.
- **Healthcare Marketing:** Understanding public response to healthcare products or campaigns.
- **Patient Awareness:** Identifying misinformation or misconceptions about treatments.

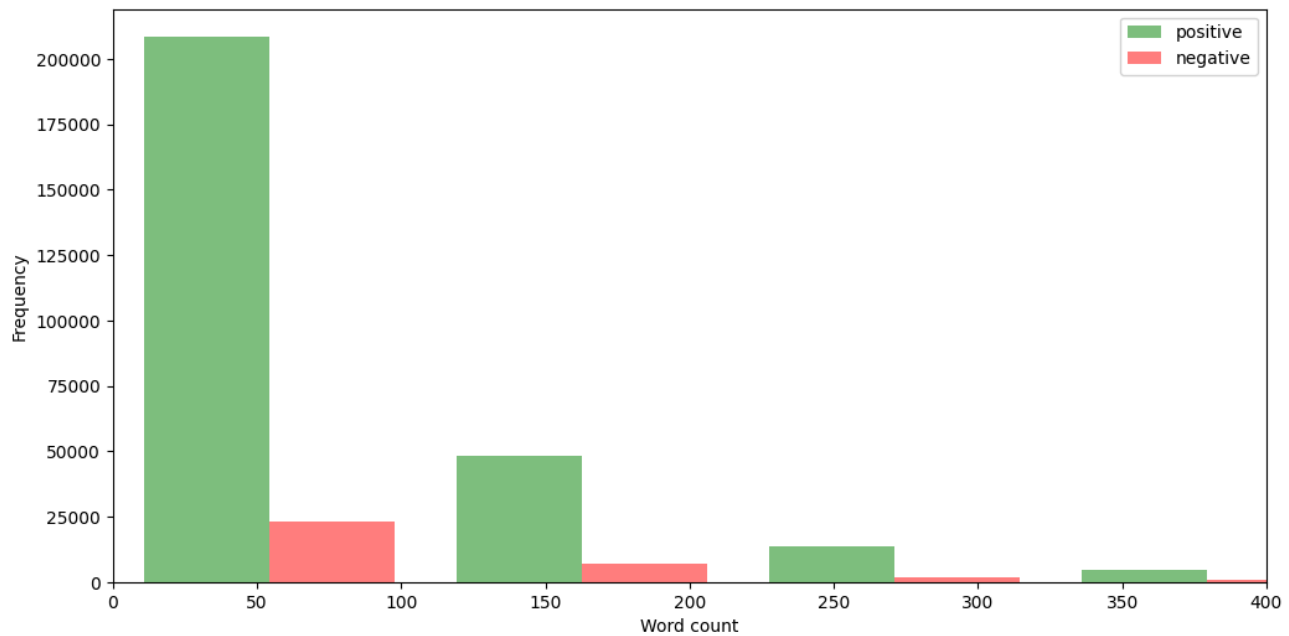


Mahavir Education Trust's
SHAH & ANCHOR KUTCHHI ENGINEERING COLLEGE
Chembur, Mumbai - 400 088
UG Program in Information Technology

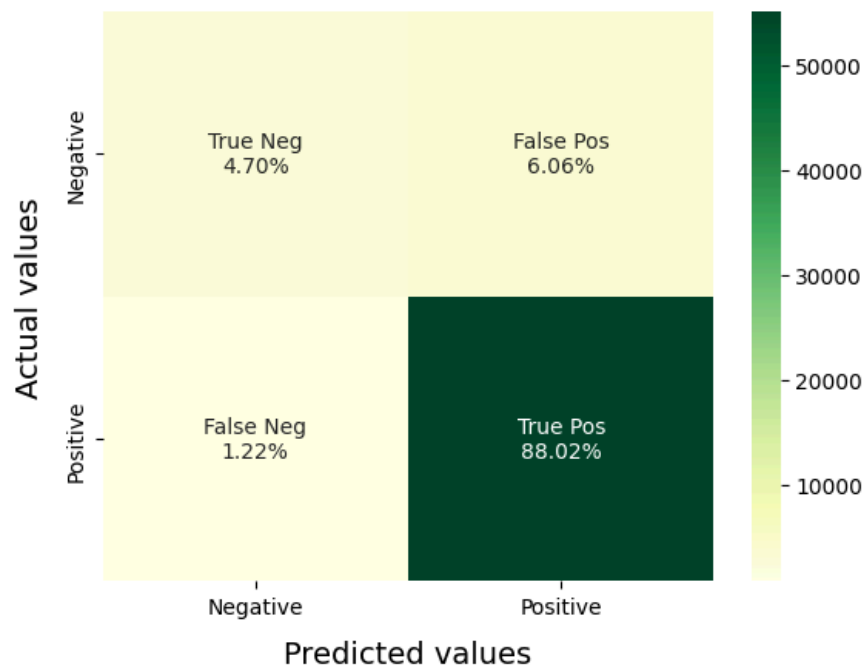
7.6 Program and Output:

Completed Google Colab [AIML7 Link](#)

7.7 Result SnapShots:



Confusion Matrix





Mahavir Education Trust's
SHAH & ANCHOR KUTCHHI ENGINEERING COLLEGE
Chembur, Mumbai - 400 088
UG Program in Information Technology

7.8 Conclusion:

In this experiment, we performed an analysis of medical reviews collected from social media platforms using Natural Language Processing (NLP) and sentiment analysis techniques. The goal was to understand public opinions, experiences, and reactions toward medicines, treatments, and healthcare services. By cleaning and preprocessing the unstructured text data, and then applying sentiment classification models, we were able to identify positive, negative, and neutral sentiments expressed by users. The experiment demonstrated how real-time social media data can provide valuable insights into patient satisfaction, treatment effectiveness, and emerging health concerns. Overall, this analysis highlights the growing importance of NLP in healthcare research and decision-making, enabling organizations to respond more effectively to public health needs.

7.9 Questions:

1. What is sentiment analysis in the context of medical reviews?

Sentiment analysis is a Natural Language Processing (NLP) technique used to determine the emotional tone behind textual data. In medical reviews, it helps identify whether patient feedback or social media discussions express positive, negative, or neutral opinions about healthcare services, medicines, or doctors.

2. Why is sentiment analysis important in healthcare applications?

Sentiment analysis helps healthcare professionals and organizations understand patient experiences, detect dissatisfaction or side effects early, and improve healthcare quality. It also supports policymakers in monitoring public health sentiments and reactions to medical treatments or policies.

3. Which tools and libraries are used for medical review sentiment analysis?

Python-based tools like Google Colab or Jupyter Notebook are used for implementation. Libraries such as NLTK, TextBlob, or VADER help in sentiment classification, while pandas, NumPy, Matplotlib, and Seaborn are used for data handling and visualization.

4. What type of dataset is used for this experiment?

A dataset containing text reviews or comments from social media platforms like Twitter, Reddit, or healthcare forums is used. Each record usually includes a user's review text and its corresponding sentiment label (positive, negative, or neutral).



Mahavir Education Trust's
SHAH & ANCHOR KUTCHHI ENGINEERING COLLEGE
Chembur, Mumbai - 400 088
UG Program in Information Technology

5. **What preprocessing steps are applied to social media data?**

Preprocessing involves cleaning the text by removing special characters, URLs, stopwords, and punctuation. Tokenization, stemming, and lemmatization are applied to standardize the text before feeding it into the sentiment analysis model.



Mahavir Education Trust's
SHAH & ANCHOR KUTCHHI ENGINEERING COLLEGE
Chembur, Mumbai - 400 088
UG Program in Information Technology