TASK- 2

Social Media Sentiment Analysis using NLP

Intern Name : Amit Gond

Organization : Tecnia Institute of of Advance Studies

Company Name : Brainwave matrix solution

Domain : Data Science

Submission Date: 8 September 2025

Index (Table of Contents)

Abstract

This project analyzes social media data to understand public sentiment on various topics. Using NLP preprocessing and sentiment analysis techniques (TextBlob/VADER), we classified tweets into positive, negative, or neutral categories. Visualizations such as sentiment distribution, trends over time, and word clouds highlight public opinions. The project demonstrates how Data Science can extract valuable insights from unstructured text data.

Introduction

Social media has become a powerful platform where people express their opinions about products, services, events, and daily experiences. This massive volume of unstructured text provides valuable insights for businesses, organizations, and researchers. Sentiment analysis, a key application of Natural Language Processing (NLP), helps in understanding whether the expressed opinion is positive, negative, or neutral. In this project, sentiment analysis is applied on social media data using NLP techniques to preprocess text, classify sentiments, and visualize public opinion trends over time.

Objective

The main objective of this project is to analyze social media data to understand public opinion on various topics. Using NLP techniques, the aim is to preprocess text data, classify sentiments into positive, negative, or neutral categories, and identify overall trends. The project further seeks to visualize sentiment distribution and highlight topic-wise insights that can support better decision-making.

Methodology

DatasetDescription:

The dataset used in this project consists of 10,000 synthetic social media posts collected across a variety of topics such as iPhone, Cricket, Movies, Traffic, Festivals, Laptops, and Restaurants. Each record contains attributes like tweet_id, username, date, and tweet_text. This dataset was chosen to simulate real-world social media conversations and provide a diverse mix of both positive and negative opinions.

5.2 NLP Preprocessing Steps"

Since social media text is often noisy and unstructured, Natural Language Processing (NLP) preprocessing techniques were applied. The preprocessing pipeline included:

- Converting text to lowercase.
- Removing URLs, mentions (@usernames), and special characters.
- Eliminating stopwords (common words like "is", "the", "at" that do not add meaning).

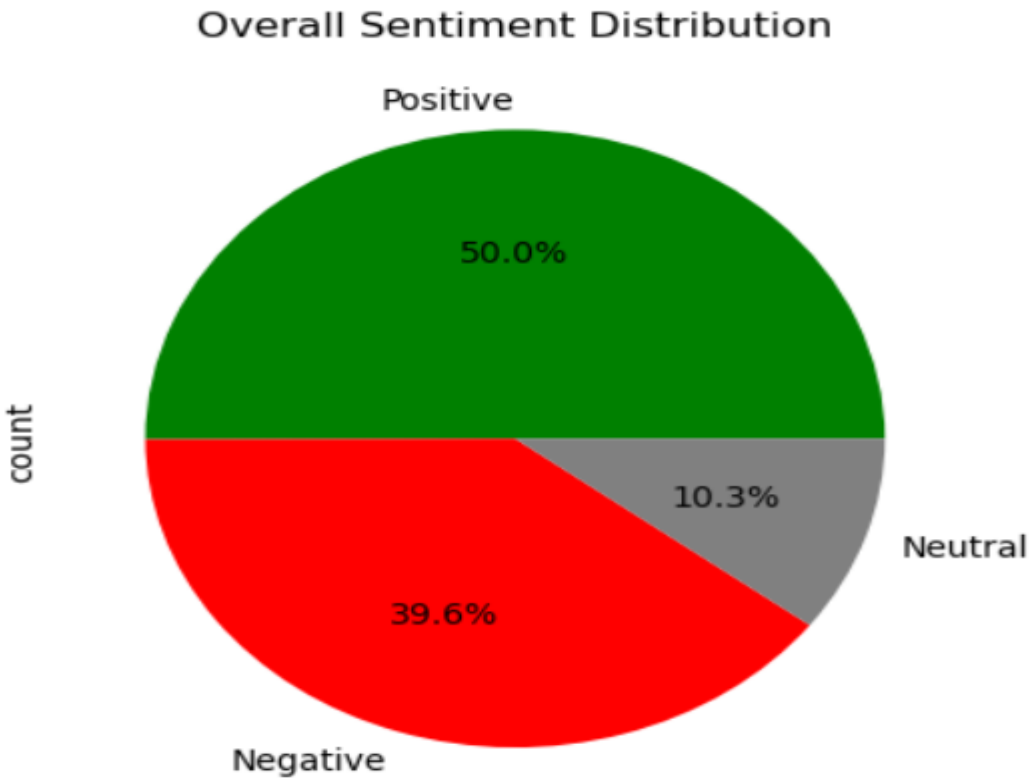
- Tokenization of text into words.
- Lemmatization to reduce words to their root forms (e.g., *running* → *run*).
  This cleaning ensured that the text was standardized and ready for sentiment classification.

5.3 Sentiment Analysis Approach

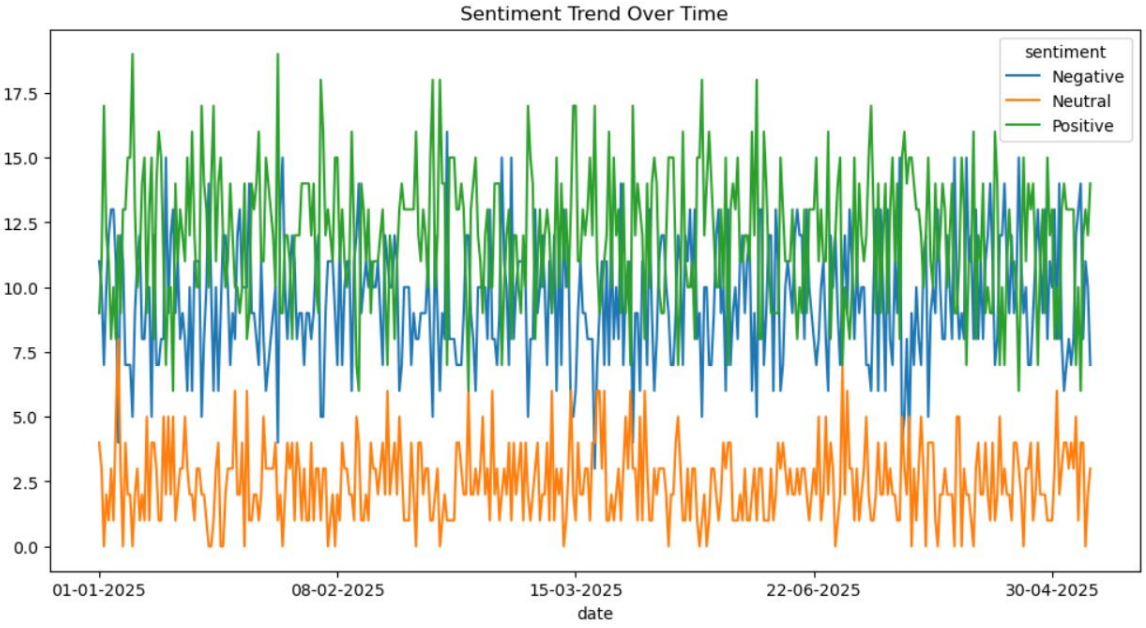
Two NLP-based sentiment analysis approaches were used. First, **TextBlob**, which calculates polarity scores ranging from –1 (negative) to +1 (positive). Second, **VADER (Valence Aware Dictionary for Sentiment Reasoning)**, which is specifically designed for analyzing sentiments in social media text. Based on the scores, each post was categorized as Positive, Negative, or Neutral. Using both models helped in validating the results and ensuring higher reliability of sentiment classification.

Results & Visuals

- **Sentiment Distribution**
  The analysis showed that most posts were positive in nature, while a smaller portion expressed negative opinions, and some remained neutral. A bar chart and pie chart were used to present the percentage share of each sentiment. The distribution highlighted that users were more inclined towards positive experiences compared to negative ones.



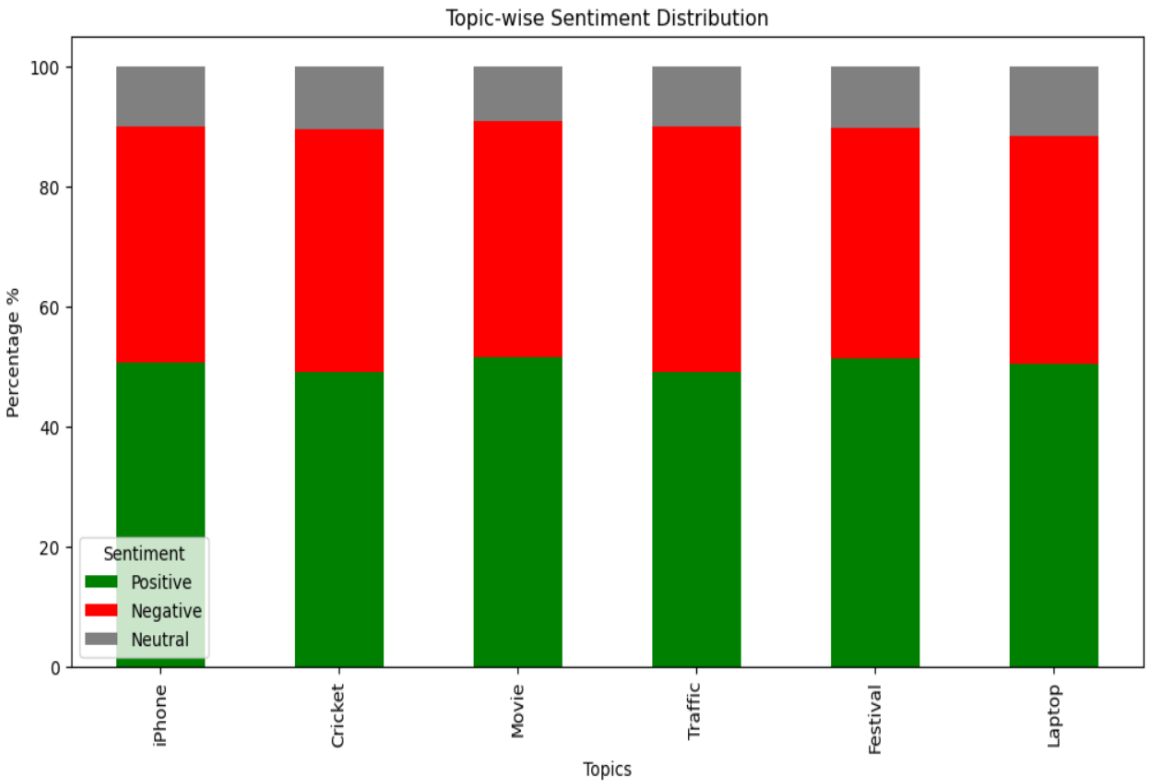
- **Sentiment Trend Over Time**
  A time-series analysis was conducted to observe daily sentiment changes. The line graph revealed stable trends in positive and neutral sentiments but also showed noticeable spikes in negative sentiments on certain days. These fluctuations suggested that negative experiences were often tied to specific events or situations.



- **Word Clouds (Positive & Negative)**
  Word clouds were generated separately for positive and negative tweets. In the positive cloud, frequent words included *love, amazing, great, fantastic*, which reflected appreciation and satisfaction. On the other hand, the

negative cloud highlighted terms such as *worst, frustrated, terrible, slow*, clearly showing areas of dissatisfaction. This comparison helped in understanding the tone and context of user opinions.

- **Topic-wise Sentiment Analysis**
  A stacked bar chart was created to compare sentiment across different topics. Topics like *Festivals* and *Movies* showed predominantly positive reactions, while *Traffic* and *Laptop performance* attracted higher negative sentiments. This breakdown emphasized that public opinion varies greatly depending on the subject, and sentiment analysis can capture these differences effectively.


Topic-wise Sentiment Distribution
Percentage %
100
80
60
40
20
0
Sentiment
Positive
Negative
Neutral
iPhone
Cricket
Movie
Traffic
Festival
Laptop
Topics

Findings / Insights

- Most of the social media posts reflected a **positive sentiment**, showing that people generally shared favorable opinions.
- Topics such as **Festivals** and **Movies** received higher positivity, indicating excitement and enjoyment among users.
- **Traffic** and **Laptop performance issues** were major sources of negative sentiment, highlighting common public frustrations.
- Word clouds revealed that positive posts frequently included words like *love, amazing, great*, while negative posts commonly contained words such as *worst, slow, frustrated*.
- The trend analysis showed fluctuations in sentiment over time, with certain days witnessing a sharp increase in negative posts, often linked to service-related complaints.

Conclusion

This project successfully demonstrated how NLP techniques can be applied to analyze social media data and classify public opinion into positive, negative, or neutral categories. The results showed clear sentiment trends and provided meaningful insights into user opinions on different topics.

For future work, real-time data can be collected directly from platforms like Twitter using APIs. Advanced deep learning models such as BERT or RoBERTa can be used to improve accuracy. Additionally, interactive dashboards in Power BI or Streamlit may be developed for real-time visualization and monitoring of sentiment trends.

References

- Python Libraries: *Pandas, NumPy, NLTK, TextBlob, Matplotlib, Seaborn, WordCloud*
- NLTK Documentation: https://www.nltk.org
- TextBlobDocumentation: https://textblob.readthedocs.io
- Dataset: https://drive.google.com/file/d/1B5J06qtYT5HOvIM4zBOVBmLHFq9NQDSj/view?usp=drive_link
- Jupyternotebook: https://drive.google.com/file/d/1sKBtygOWGV-0xbdHz-SNSO7RmZin4yD7/view?usp=drive_link
- Tutorials & Guides: Official documentation of libraries and standard NLP practices