SOCIAL MEDIA SENTIMENT ANALYSIS

Prepared by : Hagar Abdelsalam

AGENDA

Objective

Dataset Overview

Steps and Methodology

Key Insights

Tools & Libraries Used

Key Visuals & Recommendation

Conclusion

OBJECTIVE

The objective of this project was to analyze social media (Twitter) data to understand public sentiment towards general topics. The task involved applying **Natural Language Processing** (**NLP**) techniques to preprocess tweets, extract sentiment scores, build a classification model, and visualize sentiment trends over time.

DATASET OVERVIEW

Source: Sentiment140 Twitter Dataset

Size: 1.6 million labeled tweets

Sentiment Labels:

•0 ⟶ Negative

•4 ⟶ Positive

Fields Used:

- **sentiment:** Original sentiment score (0 or 4)
- **text:** Raw tweet content

STEPS AND METHODOLOGY

1. Data Loading & Preprocessing

- Loaded the dataset using pandas with latin-1 encoding.
- Renamed columns and retained only sentiment and text fields.
- Mapped numerical sentiment scores to labels: 0 → Negative, 4 → Positive.

STEPS AND METHODOLOGY

2. Text Cleaning (NLP Preprocessing)

- Removed URLs, mentions, hashtags, and special characters.
- Converted all text to lowercase.
- Removed stopwords using NLTK's English stopword list.
- Applied stemming using the Porter Stemmer.
- Created a new column clean_text with the cleaned tweets.

STEPS AND METHODOLOGY

3. Visualization: Sentiment Distribution

- Plotted sentiment distribution using Seaborn.

- **Insight :** Dataset is balanced with nearly equal Positive and Negative tweets.

4. Word Cloud

- Generated separate word clouds for Positive and Negative tweets using WordCloud.

- **Insight:**
  - Positive tweets commonly used words like *love, good, great*.
  - Negative tweets frequently included words like *hate, bad, sad*.

STEPS AND METHODOLOGY

5. Sentiment Trend Over Time

- Simulated tweet timestamps using pandas.date_range.
- Aggregated tweet counts per day and plotted trends.
- Applied 7-day rolling average to smooth the lines.

Insight:

1- Positive and Negative tweet frequencies show general consistency over time.

2- Sudden spikes or dips might indicate events or reactions, depending on the actual dataset.

STEPS AND METHODOLOGY

6. Sentiment Classification Model

- Used a basic Multinomial Naive Bayes classifier.
- Converted text to numerical vectors using CountVectorizer.
- Split the data (80% training, 20% testing).
- Evaluated model performance using classification_report.
- **Model Results:**
  - **Accuracy:** ~77%
  - **Precision & Recall:** Fairly balanced between Positive and Negative classes
- **Insight:** Even simple models like Naive Bayes can achieve strong results with well-preprocessed data.

KEY INSIGHTS

1- Sentiment Balance: Dataset has roughly equal Positive and Negative tweets useful for unbiased training.

2- Keyword Patterns:

- Positive sentiment is associated with emotionally strong, uplifting words.
- Negative sentiment leans towards frustration and complaints.

3- Trends Over Time: Patterns suggest relatively stable sentiment flow, though further topic-based filtering could reveal trends during real-world events (e.g., holidays, crises).

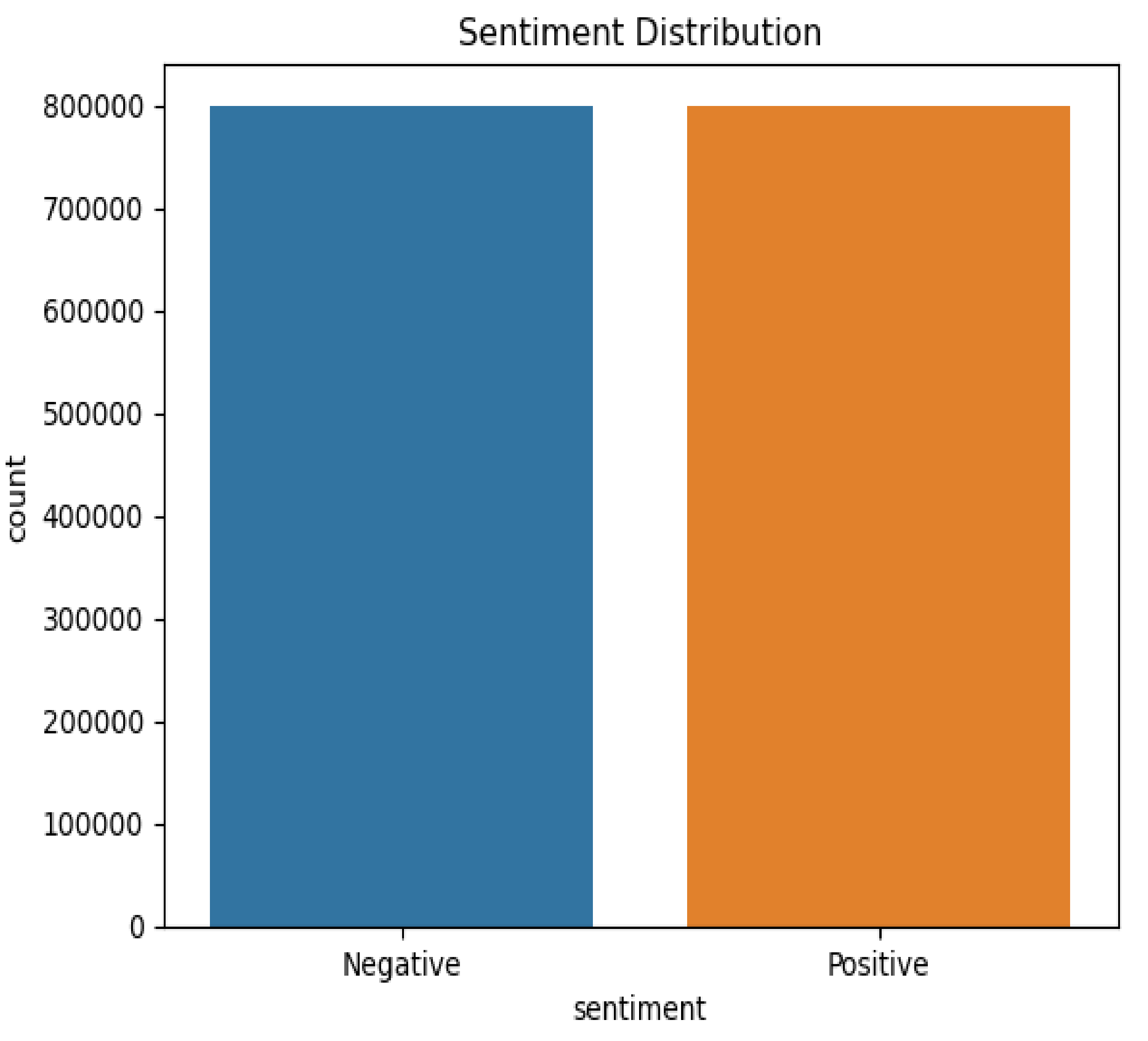
4- Model Performance: A basic NLP pipeline and simple classifier can achieve **~77% accuracy**, demonstrating the power of classical ML when applied effectively.

TOOLS & LIBRARIES USED

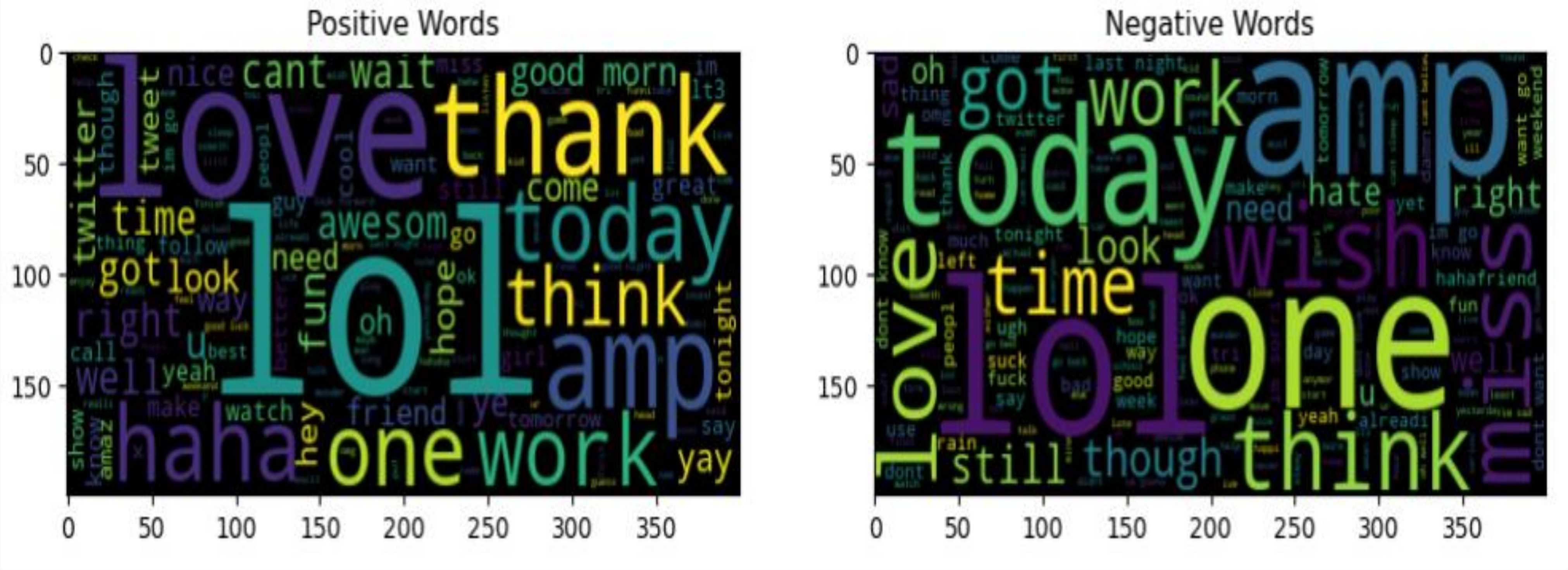
- **Pandas** – Data manipulation
- **Matplotlib / Seaborn** – Visualization
- **NLTK** – Text preprocessing (stopwords, stemming)
- **Scikit-learn** – Vectorization and modeling
- **WordCloud** – Word cloud generation

KEY VISUALS

Recommendation:
1-investigating the reasons behind the negative sentiments to address potential issues (e.g., product flaws, service complaints).
2-leverage the positive sentiments to highlight strengths in marketing or customer engagement strategies.

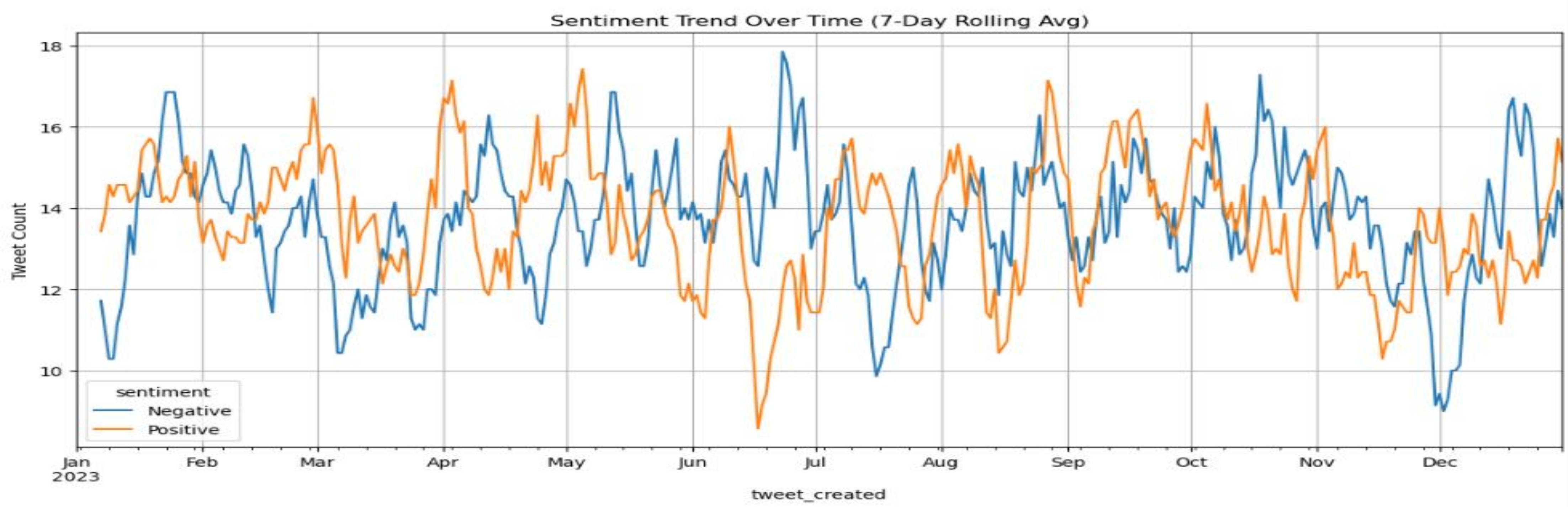


KEY VISUALS



Recommendation:
1- To boost positive sentiment, amplify what users love humor, community, and timely engagement (e.g., events or campaigns around "today" or "tonight").
2- For negative sentiment, address pain points like work-related stress or unmet expectations by offering solutions, support, or empathetic communication.

KEY VISUALS



Recommendation:

1- Investigate the causes of sentiment spikes (e.g., March, May, November) by correlating with events, product releases, or external factors during those periods.

2- Address negative spikes with targeted improvements, and amplify positive spikes through marketing.

3-Since sentiment is balanced and event-driven, focus on real-time monitoring and rapid response to emerging trends or issues to maintain or shift the sentiment in your favor.

CONCLUSION

This project demonstrated a complete NLP pipeline for analyzing Twitter sentiment:
- From data cleaning and visualization to modeling and insights extraction.
- The workflow can easily be adapted to monitor sentiment about specific events, products, or brands — making it valuable for businesses, marketing teams, or social researchers.

THANK YOU

Hagar Abdelsalam
My Github : https://github.com/hagarabdelsalam
My Email : hagarabdelsalam64@gmail.com