

SOCIAL MEDIA SENTIMENTS ANALYSIS(NLP PROJECT)

1. Introduction

Social media is one of the most powerful platforms where people freely express their opinions, experiences, and feedback about products, services, brands, and global events. Extracting valuable insights from this massive data can help businesses improve their services and understand customer perception better.

This project focuses on **Sentiment Analysis using Natural Language Processing (NLP)** techniques to classify public sentiments from tweets or reviews into three categories: **Positive**, **Neutral**, and **Negative**. We aim to build a complete pipeline starting from data preprocessing to visualization and recommendation generation based on sentiment trends.

2. Objective

The objective of this project is to:

- Clean and preprocess social media text data.
 - Apply sentiment analysis using TextBlob and VADER.
 - Classify sentiments into three classes.
 - Visualize common keywords and sentiment trends.
 - Detect sentiment spikes and generate brand improvement recommendations.
-

3. Dataset Description

- **Source:** Kaggle
 - **Link:** [Social Media Sentiment Analysis Dataset by Kashish Parmar](#)
 - **Format:** CSV
 - **Main Column:** `text` (contains tweet/review content)
 - **Size:** ~30,000 entries
-

4. Tools and Libraries

- **Python:** Programming language
 - **Google Colab:** Cloud-based IDE
 - **Libraries:**
 - `pandas`, `numpy`: Data handling
 - `nltk`, `re`: NLP preprocessing
 - `TextBlob`, `vaderSentiment`: Sentiment analysis
 - `seaborn`, `matplotlib`: Data visualization
 - `wordcloud`: Word cloud generation
-

5. Methodology

5.1 Data Collection

- Uploaded the dataset manually in Google Colab.
- Read the dataset using `pandas.read_csv()`.
- Verified data integrity: checked for nulls, duplicates, and column names.

5.2 Data Preprocessing

Steps performed:

1. Converted text to lowercase.
2. Removed URLs, punctuation, numbers, mentions, and hashtags using regex.
3. Tokenized the text.
4. Removed stopwords using NLTK.
5. Lemmatized words using WordNetLemmatizer.

Result: Cleaned text was saved in a new column `clean_text`.

5.3 Sentiment Analysis

Two models were used:

TextBlob:

- Computes a polarity score.
- Labels:
 - Positive (polarity > 0)
 - Neutral (polarity = 0)
 - Negative (polarity < 0)

VADER (Valence Aware Dictionary):

- Especially designed for social media text.
- Uses compound score:
 - Positive (≥ 0.05)
 - Neutral ($-0.05 < x < 0.05$)
 - Negative (≤ -0.05)

Columns Added:

- `sentiment_textblob`
- `sentiment_vader`

VADER was used as the final classifier due to its better accuracy on social media text.

5.4 Visualizations

A. Word Cloud:

Generated from cleaned text using `WordCloud` library to display the most frequent words.

B. Sentiment Distribution:

Bar plot showing counts of positive, neutral, and negative tweets.

C. Sentiment Trend:

- Used the index of tweets as a pseudo-time series.
- Converted sentiment to numerical values: Positive = 1, Neutral = 0, Negative = -1.
- Applied a rolling average to observe trend movement.

6. Sentiment Spike Analysis

- Rolling average calculated over a window of 10 tweets.
- Plotted line graph to detect spikes:
 - Positive spikes: Sudden jump in positive sentiment.
 - Negative spikes: Sudden dip in sentiment.
- Highlighted spikes on the graph for easy observation.






7. Key Insights

- **Most tweets were positive**, indicating a good brand perception overall.

- **Common negative keywords:** delay, issue, refund.
 - **Common positive keywords:** quick, love, great, satisfied.
 - **Spikes:** Negative spikes related to delivery or support issues.
-

8. Recommendations

Based on the analysis, the following strategies are suggested:

-  Respond to negative tweets quickly and transparently.
 -  Retweet and promote positive user content.
 -  Train customer care teams for online reputation management.
 -  Monitor keyword trends regularly to catch issues early.
 -  Use influencer partnerships to strengthen positive engagement.
-

9. Recommend Strategies to Improve Brand Perception

To enhance how the public views a brand based on sentiment data, the following actions are recommended:

1. **Active Social Listening:** Regularly monitor social media platforms for brand mentions and emotional keywords to act promptly.
 2. **Prompt Customer Support:** Ensure that support teams address negative feedback quickly and politely to regain trust.
 3. **Encourage User-Generated Content (UGC):** Promote positive stories and testimonials from happy users to amplify credibility.
 4. **Address Pain Points:** If repeated negative terms like "late", "delay", or "support" appear, focus on solving those operational issues.
 5. **Engagement Campaigns:** Run interactive campaigns like polls and contests to create positive buzz.
 6. **Transparency in Communication:** Publicly acknowledge and correct mistakes to build authenticity and trust.
-

10. Challenges Faced

- Some tweets were very short or sarcastic, making them hard to classify.
- Dataset had no timestamp column, so index was used for trend analysis.
- Preprocessing needed careful handling to avoid removing useful context.

11. Future Scope

- Use Twitter API for real-time tweet collection.
- Apply topic modeling (e.g., LDA) to explore themes behind sentiments.
- Implement live dashboards using Streamlit.
- Extend to multilingual sentiment analysis.

12. Conclusion

This project provided hands-on experience with text data preprocessing, sentiment classification, and data visualization. I learned how NLP can be applied to extract meaningful insights from raw social media content and support real-world brand strategies.

The overall project helped me build confidence in working with unstructured data and using Python libraries for end-to-end analysis.

13. Personal Reflection

Working on this project gave me a clear understanding of how language data can be translated into useful business insights. It was exciting to see how simple tools like sentiment analysis can detect trends and customer feelings from just words. I enjoyed every step, from cleaning the data to creating visual graphs.
