

Real-Time Public Sentiment Dashboard (Twitter/X)

Report for Samsung

Harshabardhana Parida

May 16, 2025

Introduction

This report presents the findings of a real-time public sentiment analysis conducted on tweets related to Samsung. The analysis was performed using the Twitter API, specifically utilizing the Tweepy library for data retrieval and the NLTK library for sentiment analysis. The primary objective was to gauge public sentiment towards Samsung by analyzing the latest tweets, categorizing them as Positive, Negative, or Neutral based on their sentiment scores.

Methodology

Data Collection

- **Library Installation:** The necessary libraries were installed, including `tweepy`, `pandas`, and `nltk`.

```
!pip install tweepy pandas nltk
```

- **Sentiment Analysis Setup:** The VADER sentiment analyzer from NLTK was initialized to evaluate the sentiment of the tweets.

```
import nltk
nltk.download('vader_lexicon')
from nltk.sentiment.vader import SentimentIntensityAnalyzer
sia = SentimentIntensityAnalyzer()
```

- **Twitter API Access:** A Bearer Token was used to authenticate and access the Twitter API.

```
BEARER_TOKEN = "your_token_here"
client = tweepy.Client(bearer_token=BEARER_TOKEN)
```

- **Tweet Retrieval:** The script fetched the latest 100 tweets containing the keyword "Samsung", filtering out retweets and non-English tweets.

```
query = "Samsung -is:retweet lang:en"
tweets = client.search_recent_tweets(query=query, tweet_fields=['created_at', 'text'])
```

Sentiment Analysis

Each tweet was analyzed for sentiment using the VADER sentiment analyzer. The sentiment was categorized based on the compound score as follows:

- **Positive:** $\text{Score} \geq 0.05$
- **Negative:** $\text{Score} \leq -0.05$
- **Neutral:** $-0.05 < \text{Score} < 0.05$

The results were stored in a DataFrame and saved to a CSV file for further analysis.

Data Aggregation

To facilitate comprehensive analysis, multiple CSV files containing sentiment data were combined into a single dataset. The combined dataset was saved in both CSV and Excel formats.

Word Frequency Analysis

A word frequency analysis was conducted to identify the most common words associated with each sentiment category. The key steps included:

- **Text Preprocessing:** Removing URLs, mentions, hashtags, non-alphabetic characters, and stopwords.
- **Tokenization:** Breaking down cleaned text into words.
- **Frequency Calculation:** Computing word frequency for each sentiment category and extracting the top 50 words.

The results were saved in a CSV file for visualization purposes.

Results

- **Sentiment Distribution:** Tweets were categorized into Positive, Negative, and Neutral groups. Visualizations such as pie charts or bar graphs can illustrate this distribution effectively.
- **Word Frequency:** Identification of common words per sentiment category aids in understanding topics discussed within each sentiment.
- **Sentiment Scores:** Each tweet was assigned a sentiment score, enabling nuanced analysis beyond categorical labels.

Conclusion

The sentiment analysis of tweets related to Samsung provides valuable insights into public perception. Categorizing tweets into Positive, Negative, and Neutral sentiments helps stakeholders understand consumer opinions and brand sentiment. Word frequency analysis enriches this understanding by highlighting key discussed topics corresponding to each sentiment category.