SOCIAL MEDIA SENTIMENTS ANALYSIS DATASET

REPORT

Presented By
Abouri Abdelhamid
Laamari Youusef

2024

Introduction

Social media sentiment analysis is a powerful tool for extracting insights and patterns from user-generated content. The dataset under consideration encompasses various dimensions, including text, sentiment, timestamp, user details, platform, engagement metrics, geographical information, and more.

The analysis aims to uncover valuable information that can be used for strategic decision-making, content optimization, and understanding user behavior.

General Problematique

The overarching problematique revolves around the need to comprehend and leverage the vast amount of information available on social media platforms. Key challenges include deciphering sentiment trends, understanding user engagement dynamics, exploring temporal and geographical variations, and extracting meaningful insights from user-generated content. The complexity lies in the interplay of various factors and the dynamic nature of social media data.

Solutions Provided

Sentiment Analysis :

Distribution Analysis: Utilize descriptive statistics to understand the distribution of sentiments, helping to identify prevalent sentiments and their proportions in the dataset

Platform-based Sentiment Trends: Explore sentiment patterns specific to different social media platforms to discern platform-specific user sentiments.

User Engagement

- Correlation Analysis: Investigate the relationship between likes and retweets to identify content with high engagement, aiding in the creation of impactful social media strategies.
- Hashtag Analysis: Identify popular hashtags and analyze their correlation with high engagement metrics, allowing for the optimization of content with trending hashtags.

Temporal Analysis

- Time-of-Day Sentiment Analysis: Investigate how sentiments change over different times of the day to understand temporal patterns in user sentiments.
- User Activity Peaks: Identify peak times for user activity, likes, and retweets, facilitating optimal content scheduling.

Geographical Analysis

- Country-wise Sentiment Variations: Investigate sentiment variations across different countries to understand regional sentiment trends.
- Regional Trending Topics: Correlate user engagement metrics with geographical locations, uncovering regional trending topics and sentiments.

Content Analysis

- Hashtag Insights: Extract insights from frequently used hashtags, shedding light on prevalent topics and themes.
- Topic Modeling: Apply natural language processing techniques to identify prevalent themes and topics in user-generated content.

User Behavior

Highly Engaging Users: Analyze the behavior of users with high engagement metrics, exploring posting frequency, sentiments expressed, and patterns in engagement.

Platform Comparison

Sentiment and Engagement Metrics: Compare sentiment distribution, user engagement, and popular hashtags across different social media platforms, highlighting platform-specific trends.

Time-Series Analysis

- Long-term Trend Identification: Utilize time-series analysis to identify long-term trends in sentiment and engagement, aiding in strategic planning.
- Forecasting Models: Apply forecasting models to predict future trends based on historical data.

Cross-Column Analysis

- Correlation Studies: Investigate correlations between sentiment and other columns (e.g., country, hashtags), unveiling patterns and associations.
- Influence Factors: Explore how factors like the time of day or specific hashtags influence sentiment.

Text Analysis

- NLP for Key Phrases: Implement natural language processing to extract key phrases, topics, and keywords from the text column.
- Topic Modeling: Conduct topic modeling to identify prevalent themes in user-generated content.

In this machine learning project, we build a binary text classifier to classify the sentiment behind the text. We use the various NLP preprocessing techniques to clean the data and utilize the LSTM layers to build the text classifier.

Tools and Libraries Used :

- Jupyter Notebook: Utilized for interactive development, data analysis, and documentation of the project.
- Google Colab: Employed as a cloud-based platform for running and collaborating on Jupyter Notebooks, providing access to GPU resources for deep learning tasks.
- Python Libraries:
  - pandas: Used for data manipulation and analysis.
  - matplotlib: Employed for data visualization.
  - re: Utilized for regular expression operations.
  - seaborn: Utilized for statistical data visualization.
  - TensorFlow: Utilized for developing and training machine learning models.
    - Tokenizer and pad_sequences: Preprocessing text data for input to neural networks.
    - Sequential, LSTM, Dense, Dropout, SpatialDropout1D, Embedding: Components of the deep learning model architecture.