

Sentiment Analysis and Social Media Insights

Overview

This project involves analyzing social media data to extract insights such as sentiment distribution, platform performance, trending hashtags, and user engagement. The analysis includes data cleaning, sentiment analysis, topic modeling, and advanced visualizations.

Steps in the Analysis

1. Import Libraries

The following libraries were used for data manipulation, visualization, and analysis:

- `pandas`, `numpy` for data handling
- `matplotlib`, `seaborn` for visualizations
- `nlTK`, `TextBlob` for text processing and sentiment analysis
- `sklearn` for topic modeling

2. Data Loading

- Data was loaded from a CSV file (`raw_data.csv`).
- Necessary NLTK resources (`punkt`, `stopwords`) were downloaded.

3. Data Cleaning and Preprocessing

- Removed URLs, mentions, and special characters.
- Converted text to lowercase and applied tokenization, stopword removal, and stemming.

4. Sentiment Analysis

- Sentiment polarity was calculated using `TextBlob`.
- Sentiments were categorized as Positive, Neutral, or Negative.
- Visualized sentiment distribution using bar plots.

5. Platform Analysis

- Analyzed average sentiment scores by platform.
- Visualized platform sentiment comparison using bar charts.

6. Hashtag Analysis #

- Extracted hashtags from posts and identified the top 20 trending hashtags.
- Visualized hashtag frequencies using bar plots.

7. Time Series Analysis

- Analyzed monthly average sentiment trends.
- Visualized sentiment trends over time using line plots.

8. Topic Modeling

- Applied Latent Dirichlet Allocation (LDA) to identify topics in the text.
- Displayed the top words for each topic.

9. User Engagement

- Identified top users based on retweets and likes.
- Visualized user engagement using grouped bar charts.

10. Time-based Patterns

- Analyzed average sentiment by hour of the day.
- Visualized hourly sentiment trends using line plots.

11. Platform Comparison

- Compared platforms based on total posts, average likes, retweets, and sentiment scores.
- Visualized platform performance using bar plots.

12. Emotion Detection

- Detected emotions (e.g., happy, sad, angry) using a predefined lexicon.
- Visualized emotion distribution using a donut chart.

13. Additional Visualizations

- Generated a word cloud for cleaned text.
- Visualized sentiment by country and engagement vs. sentiment.

14. Summary of Findings

- **Overall Sentiment Distribution:** Neutral (395), Positive (240), Negative (97).
- **Top Platforms by Sentiment:** Twitter (0.123), Instagram (0.107).
- **Trending Hashtags:** #Serenity, #Gratitude, #Excitement, etc.
- **Top Users by Engagement:** WinterWarmth, CosmosExplorer, etc.
- **Emotion Distribution:** Neutral (642), Happy (51), Sad (19).

Outputs

- Processed data saved to `processed_data.csv`.
- Visualizations include sentiment trends, hashtag analysis, and platform comparisons.

Tools and Libraries

- Python: `pandas`, `numpy`, `matplotlib`, `seaborn`, `nltk`, `TextBlob`, `sklearn`, `wordcloud`.

Conclusion

This analysis provides actionable insights into social media trends, user engagement, and sentiment patterns, enabling better decision-making for content strategies.