Project Report:
Social PiggyBacking

Project Title: Social PiggyBacking
Your Name: Mihir Panchal
Date: Feb 2024

Abstract

This project, "Social PiggyBacking," focuses on analyzing a dataset of COVID19 related tweets to uncover patterns and insights related to public sentiment, engagement, and user behavior. The analysis involved data loading and preparation, exploratory data analysis (EDA), and sentiment analysis using TextBlob. Key findings include insights into tweet length distributions, hashtag usage, user engagement levels, and sentiment trends, providing a comprehensive overview of public discourse surrounding COVID19 on social media.

Introduction

The COVID19 pandemic has significantly impacted global communication, particularly on social media platforms where users express their thoughts, concerns, and sentiments through tweets. Understanding the nature of these tweets can provide valuable insights into public perception and response to the pandemic.

Objectives

- To analyze the distribution of tweet lengths and user engagement metrics.
- To identify popular hashtags and tweet sources.
- To assess the sentiment expressed in tweets and its correlation with user behavior.

Importance

- This project contributes to the understanding of how social media reflects societal sentiments during crises. The insights gained can help researchers, policymakers, and public health officials gauge public reaction and adapt their strategies accordingly.

__Methodology__

- Methods and Technologies
- Data Source: A CSV file containing COVID19 related tweets.
- Technologies Used: Python, Pandas, Matplotlib, Seaborn, TextBlob.

Steps Taken

1. Data Loading and Preparation:

- Loaded the CSV file into a Pandas DataFrame.
- Renamed columns for clarity.
- Cleaned tweet text by removing *URLs, mentions, hashtags, retweets, and punctuation*, converting text to lowercase.

2. Exploratory Data Analysis (EDA):

- Analyzed tweet length distribution using a histogram.
- Identified frequent hashtags with a bar plot.
- Explored user followers distribution.
- Examined the relationship between followers and retweets using a scatter plot.
- Identified top tweet sources.
- Analyzed user verification status with a pie chart.

3. Sentiment Analysis:

- Used "TextBlob" for sentiment classification (positive, negative, neutral).
- Visualized sentiment distribution with a pie chart.
- Explored the relationship between tweet sentiment and retweets using a count plot.
- Visualized sentiment by user location for the top 10 locations.

Implementation

The project involved multiple coding phases to handle data cleaning, analysis, and visualization. Key challenges included:

- Data Cleaning: The initial dataset contained inconsistencies in formatting. This was resolved by implementing regex patterns for text normalization.

- Sentiment Analysis: The *TextBlob* library provided reliable sentiment scores, though certain tweets required additional context for accurate classification.

Results

The analysis yielded several findings:

- Tweet Length Patterns: The majority of tweets were between *"100150"* characters.

- Popular Hashtags: The most frequently used hashtags included *#COVID19* and *#StayHome.*

- User Engagement Levels: Users with a higher follower count tended to receive more retweets.

- Sentiment Distribution: The analysis indicated a predominant neutral sentiment, followed by positive and negative sentiments.

- Geographical Sentiment Trends: Locations such as *New York* and *California* exhibited a higher proportion of negative sentiments.

Visualizations accompanying these findings include histograms, bar plots, pie charts, and scatter plots, each illustrating key aspects of the dataset.

Conclusion

The "Social PiggyBacking" project successfully analyzed COVID19 related tweets, revealing significant insights into public sentiment and engagement. The findings underscore the role of social media in shaping and reflecting societal responses during crises.

Recommendations

Future work could involve:
- Analyzing the impact of major news events on tweet sentiment.
- Exploring other social media platforms for comparative analysis.
- Investigating the influence of verified accounts on tweet engagement.

References

TextBlob Documentation
Pandas Documentation
Matplotlib and Seaborn Documentation
Twitter API Guidelines and Usage Policies