

Twitter Sentiment Analysis of Public Reaction to COVID-19 News

Prepared and Presented by Alexandra Yakovleva

Project Overview

- **Project Overview**

- This project analyzes COVID-19-related tweets to understand public sentiment evolution and spread in response to news and events. By leveraging Natural Language Processing (NLP) and sentiment analysis models, we aim to gain insights into online conversations surrounding the pandemic.

- **Importance and Motivation**

- Understanding public sentiment during a global crisis is crucial for:
- Public Health Officials: Gauging response to policies and interventions.
- Media Outlets: Assessing the impact of news coverage on public perception.
- Government Agencies: Monitoring public opinion and tailoring communication.
- Researchers: Studying information and misinformation spread.
- This project contributes by providing a comprehensive analysis of Twitter data, revealing trends and patterns in public sentiment.

Business Problem and Objectives

- **Problem Statement**

- Media outlets and public health organizations need a better understanding of how their COVID-19-related news and announcements influence public sentiment on Twitter.

- **Key Questions**

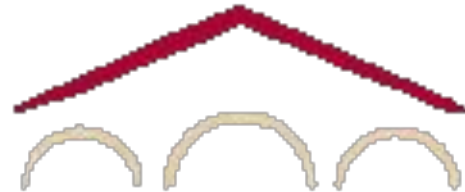
- How does sentiment spread following a COVID-19 news announcement?
- What are the key topics and themes associated with sentiment trends?
- Can we identify patterns or correlations between news events and sentiment changes?

- **Project Objectives**

- Develop a robust NLP pipeline for Twitter data analysis.
- Apply sentiment analysis models to classify tweets and track trends.
- Visualize and interpret results for actionable insights.
- Potentially identify key influencers and networks driving sentiment.

Data Acquisition and Preparation

- **Covid-19 Twitter Dataset:** A large collection of COVID-19-related tweets from Kaggle (450.000).
- **GloVe Embeddings:** Pre-trained word embeddings from Stanford NLP.



Data Limitations: Paywalled Access to Twitter Data

 If you need higher levels of access, sign up for the Enterprise API today! Start [here](#).

Pro

Unleash the full potential of X's v2 API

Apps 3 environments

DMs Over 300K requests per month per user

Posts Retrieve up to 1M Posts per month

Realtime Posts stream Access to realtime posts stream

Users Over 8M requests per month per user

SAVE 10%

\$4500/month

\$54000 billed annually

Upgrade now

\$5000/month

\$5000 billed monthly

Upgrade in 1-Click

Basic

Begin your journey with X's v2 API

Apps 2 environments

DMs 75K requests per month per user

Posts Retrieve up to 15K Posts per month

Realtime Posts stream No access

Users 50k requests per month per user

SAVE 12.5% - LIMITED TIME

\$175/month

\$2100 billed annually

Upgrade now

\$200/month

\$200 billed monthly

Upgrade in 1-Click

Free

Get limited access to X's v2 API

Apps 1 environment

Posts Retrieve up to 100 Posts and 500 writes per month



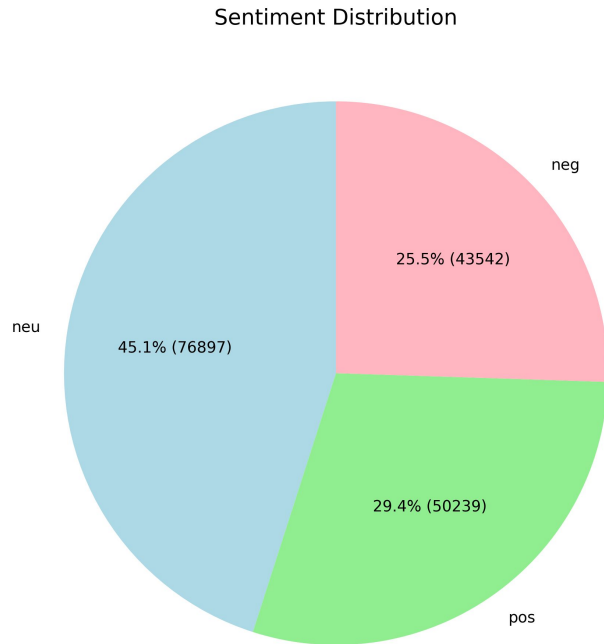
You have access to this plan

Downgrade to free access?

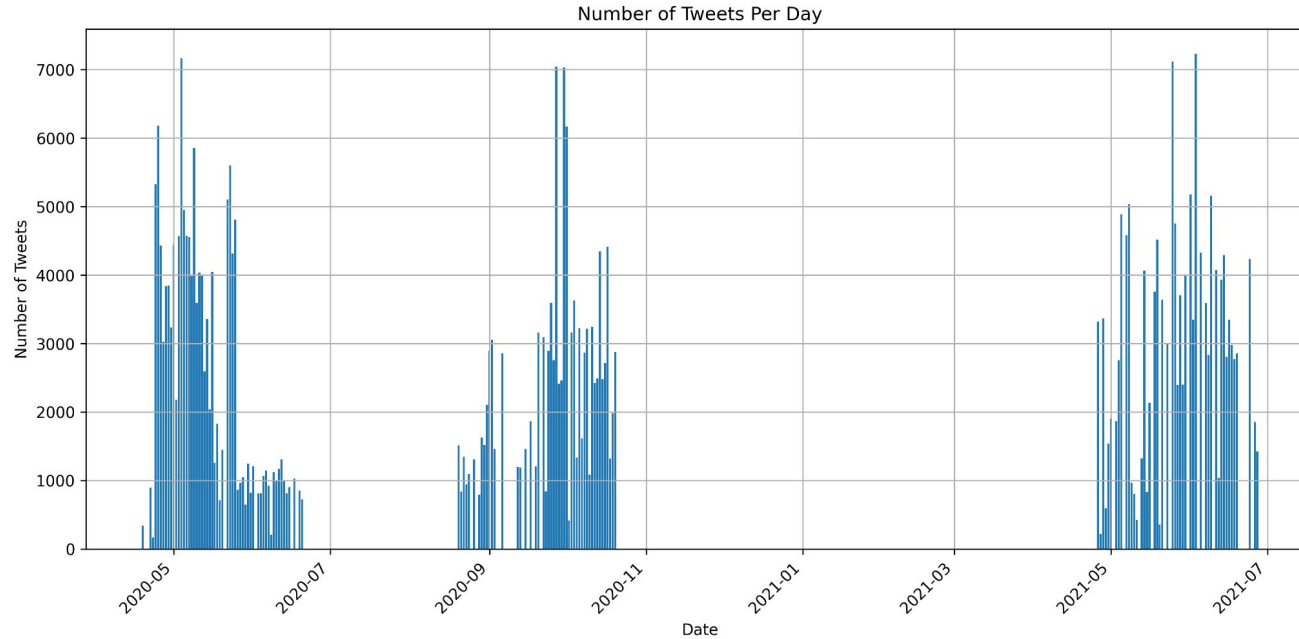
Downgrade

Data Limitations: Sentiment labels

- Imbalanced distribution.
- Unverified Labels.



Data Limitations: Date Ranges



Data Cleaning and EDA

- **Date:** Temporal trends of tweet data are extracted.
- **Language:** Relevant language is subset to increase relevant tweet ratio.
- **Location:** Location data is standardized and processed for geocoding.
- **Source:** Platforms (e.g., Twitter for Android) are identified.
- **Sentiment:** Sentiment labels are explored and distributions analyzed.
- **Social Connections:** Mentions and retweets are analyzed for network insights.

Text Preprocessing and Feature Engineering:



Cleaning: *another day in paradise grinning face with big eyes*

Preprocessing: *another day paradise grinning face big eye*

Feature Extraction: *'another day', 'another day paradise', 'grinning face'*

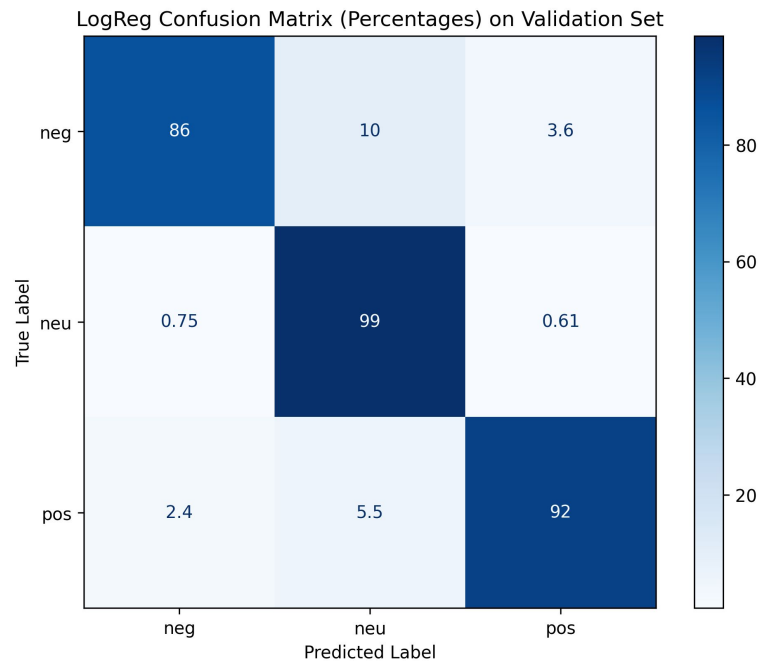
Sentiment Analysis using Supervised Machine Learning

- **Methodology:**

**Initially Features $X(170679 \times 10200)$,
target Sentiment(3 categories)**

- Train-Test-Validate Data Split (70%-15%-15%)
- Sentiment Classification using Machine Learning models (Logistic Regression, others).

Results and Visualization: Performance



Results and Visualization: Predictive Features

Positive	<i>Bliss, Adapt, Acceptable</i>
Neutral	<i>Address, Absolutely, Actually</i>
Negative	<i>Administration, Activist, Ability</i>

Conclusion

- **Summary:** *Developed a system to analyze real-time feedback, classify sentiment, track sentiment trends.*
- **Strengths and Limitations:** *Model performs well on new data and can provide actionable insights for the stakeholders.*
- **Future Work:**