Twitter Sentiment Analysis of Public Reaction to COVID-19 News

Prepared and Presented by Alexandra Yakovleva

Project Overview

Motivation:

By leveraging Natural Language Processing (NLP) and sentiment analysis models, we aim to gain insights into online conversations surrounding the pandemic.

Project Overview

Potential Stakeholders:

Public Health Officials, Media Outlets, Government Agencies, Researchers.

Business Problem and Objectives

***Problem Statement*:**

Potential Stakeholders need **better understanding** of how their announcements influence public sentiment on Social Media.

Business Problem and Objectives

Key Questions:

- What is the sentiment?
- Can we identify features related to specific sentiment?
- Can we identify patterns or correlations between events and sentiments?

Business Problem and Objectives

Project Objectives:

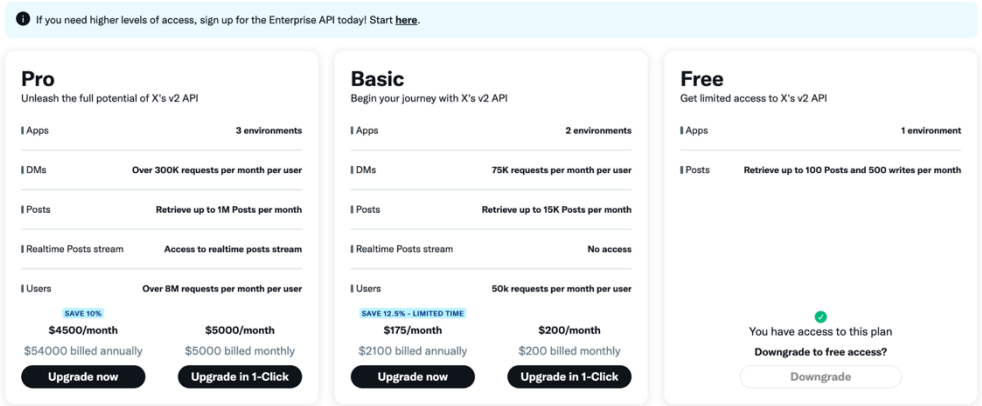
- Develop NLP pipeline for Social Media data analysis
- Deploy sentiment analysis classification models
- Visualize and interpret results

Data Acquisition and Preparation

Covid-19 Twitter Dataset: A large collection of COVID-19-related tweets from Kaggle.

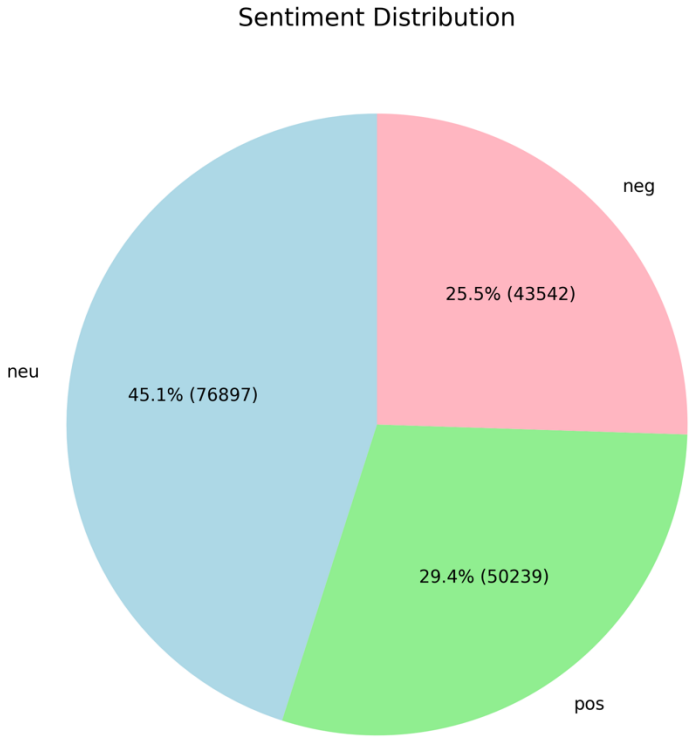
GloVe Embeddings: Pre-trained word embeddings from Stanford NLP Group.

Data Limitations: Paywalled Access to Twitter Data

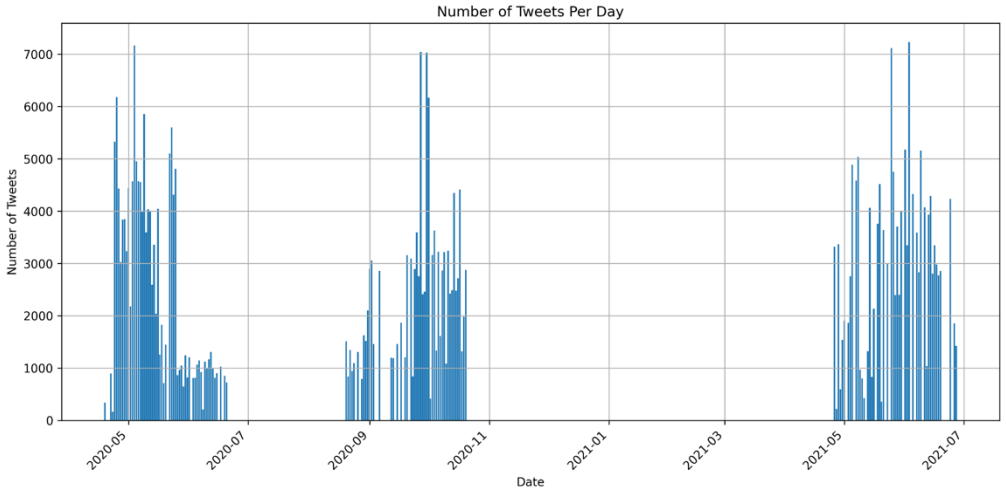


Data Limitations: Sentiment Labels

- Imbalanced Distribution.
- Unverified Labels.
- Unknown Labeling Algorithm.



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'*

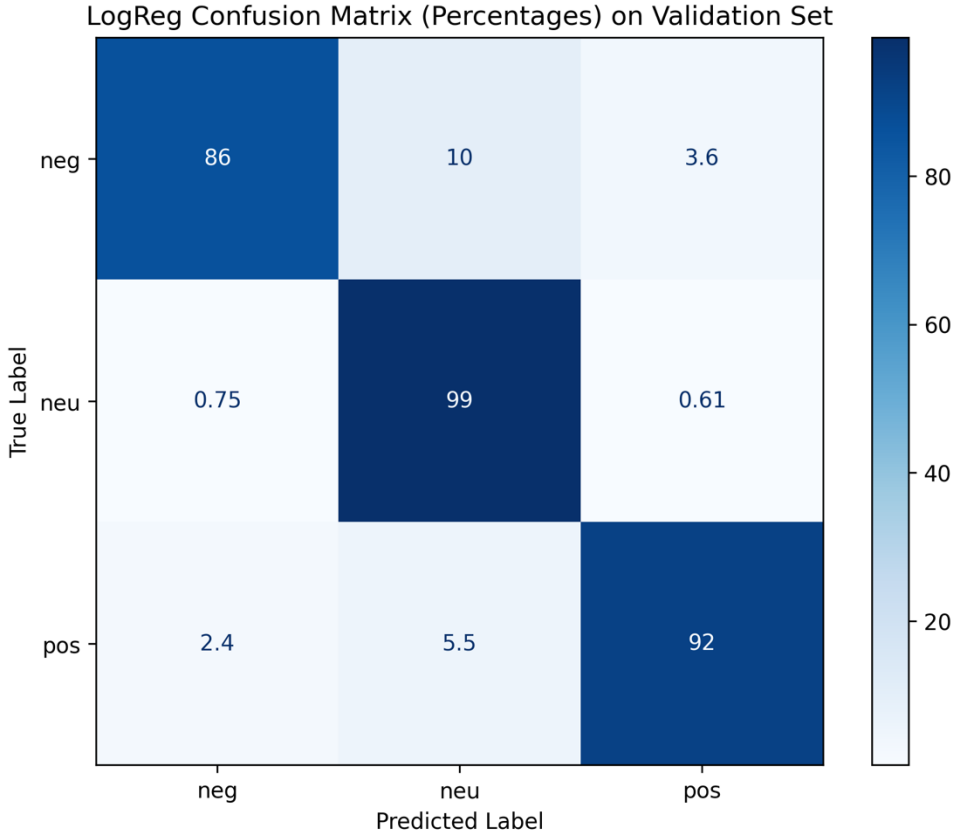
Sentiment Analysis using Supervised Machine Learning

Methods:

- **Features:** X (170679 x 10200)
- **Target:** y (3 Sentiment categories)
- **Data Split**: Train (70%) – Test (15%) –Validate(15%)
- **Sentiment Classification:** Logistic Regression
- **Performance Metric:** accuracy.

Results and Visualization: Performance

Weighted Accuracy 94%



Results and Visualization: Predictive Features

Positive	*Bliss, Adapt, Acceptable*
Neutral	*Address, Absolutely, Actually*
Negative	*Administration, Activist, Ability*

Conclusion

- **Summary:** *Developed a system to classify sentiment and extract informative features.*
- **Strengths:** *Model performs well on new data and can provide actionable insights for the stakeholders.*
- **Limitations:** *Model is limited by a specific sentiment classification. Model covers extensive time-period and is not specific to a particular stakeholder.*
- **Future Work:** *Implement advanced sentiment classification and feature engineering. Re-train model on more specific subset of data.*