

A word cloud background featuring various terms related to COVID-19, such as 'coronavirus', 'health', 'vaccine', 'new', 'positive', 'lockdown', 'trump', 'spread', 'mask', 'vaccination', 'recovery', 'social', 'team', 'want', 'said', 'part', 'post', 'sign', 'dose', 'join', 'family', 'right', 'may', 'every', 'world', 'time', 'thing', 'coronavirus', 'trump', 'health', 'vaccine', 'new', 'positive', 'lockdown', 'spread', 'mask', 'vaccination', 'recovery', 'social', 'team', 'want', 'said', 'part', 'post', 'sign', 'dose', 'join', 'family', 'right', 'may', 'every', 'world', 'time', 'thing', 'coronavirus', 'trump', 'health', 'vaccine', 'new', 'positive', 'lockdown', 'spread', 'mask', 'vaccination', 'recovery', 'social', 'team', 'want', 'said', 'part', 'post', 'sign', 'dose', 'join', 'family', 'right', 'may', 'every', 'world', 'time', 'thing'. The words are in various colors (green, yellow, blue, purple) and sizes, creating a dense, textured background.

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

❗ 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

1 Apps	3 environments
1 DMs	Over 300K requests per month per user
1 Posts	Retrieve up to 1M Posts per month
1 Realtime Posts stream	Access to realtime posts stream
1 Users	Over 8M requests per month per user

SAVE 10%

\$4500/month
\$54000 billed annually

\$5000/month
\$5000 billed monthly

Upgrade nowUpgrade in 1-Click

Basic

Begin your journey with X's v2 API

1 Apps	2 environments
1 DMs	75K requests per month per user
1 Posts	Retrieve up to 15K Posts per month
1 Realtime Posts stream	No access
1 Users	50k requests per month per user

SAVE 12.5% - LIMITED TIME

\$175/month
\$2100 billed annually

\$200/month
\$200 billed monthly

Upgrade nowUpgrade in 1-Click

Free

Get limited access to X's v2 API

1 Apps	1 environment
1 Posts	Retrieve up to 100 Posts and 500 writes per month

✔

You have access to this plan

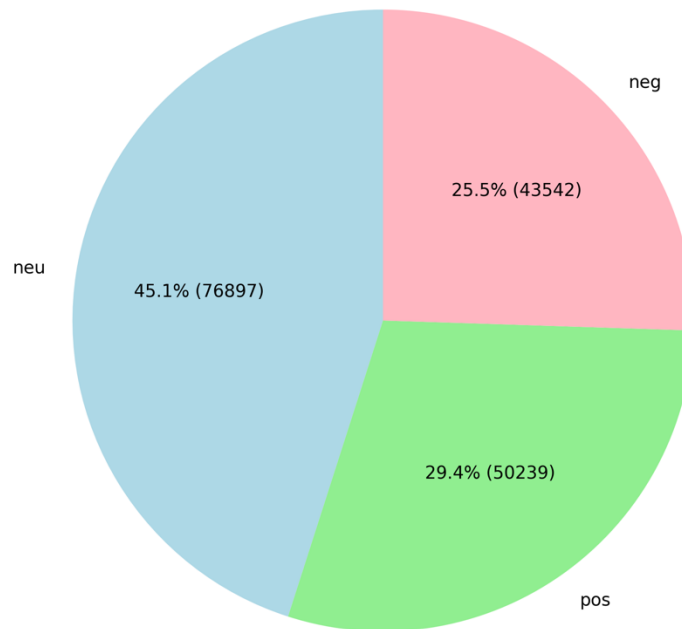
Downgrade to free access?

Downgrade

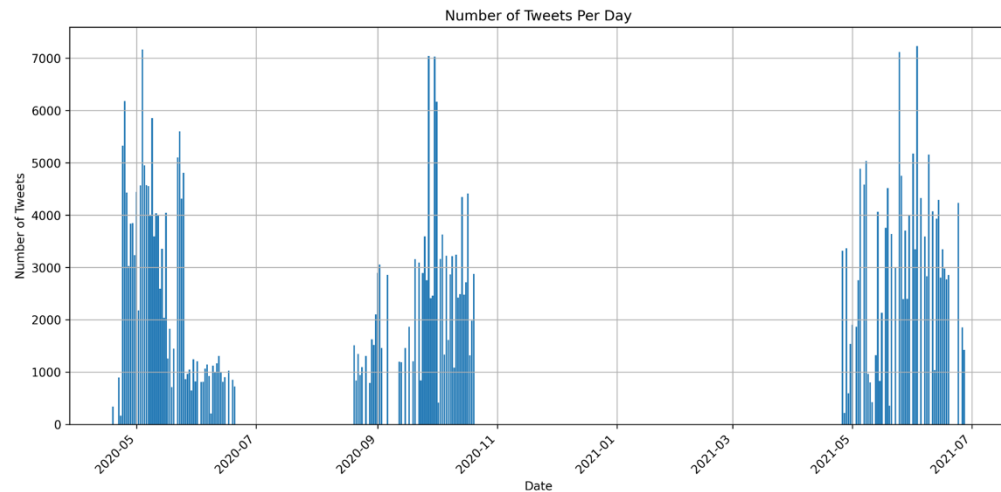
Data Limitations: Sentiment Labels

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

Sentiment Distribution



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'*



Name

@Username



RT @user456: Another day in paradise! 😊 #sunnyday

12:00 PM · Jun 1, 2021



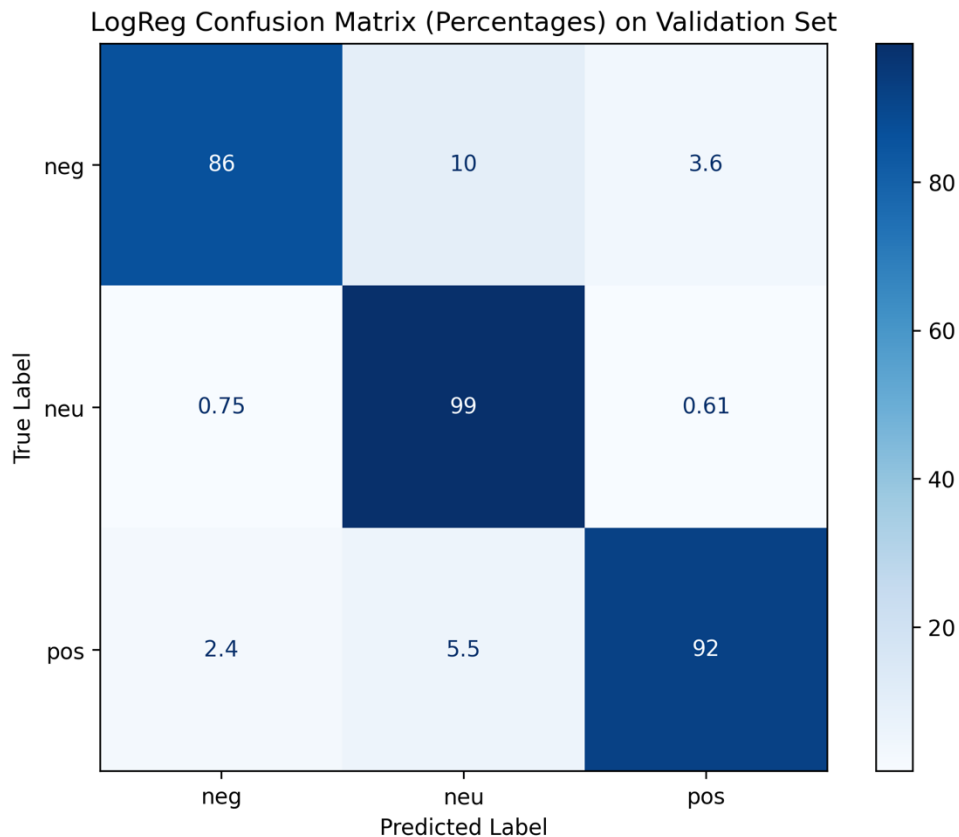
Sentiment Analysis using Supervised Machine Learning

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

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.*