Slide Title: Twitter Sentiment Analysis of Public Reaction to COVID-19 News

roject 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.

Slide Title: 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.

Slide Title: Data Acquisition and Preparation

- Data Sources
- Covid-19 Twitter Dataset: A large collection of COVID-19-related tweets from Kaggle.
- GloVe Embeddings: Pre-trained word embeddings from Stanford NLP.
- Data Relevance
- The dataset provides a valuable source of public opinion and sentiment during the pandemic.
- It includes sentiment labels for training and evaluating models.
- Data Limitations
- Restricted access to recent Twitter data due to API changes.
- Pre-computed sentiment labels might have unknown accuracy.

Slide Title: 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.

Slide Title: Text Preprocessing and Feature Engineering

• Cleaning: Remove noise, links, mentions, and hashtags. Preprocessing: Tokenize, lemmatize, and remove stop words. Feature Extraction: Generate n-grams, TF-IDF features, or BERT embeddings.

•

Slide Title: Sentiment Analysis

- Methodology:
- Train-Test Data Split
- Sentiment Classification using Machine Learning models (Logistic Regression, others).
- Model Evaluation: Accuracy, Precision, Recall.
- Feature Reduction and Hyperparameter Tuning for optimization.
- Feature Interpretation.

Slide Title: Results and Visualization

Slide Title: Conclusion

- Summary: (Recap your project goals, methods, and key findings.)
- Strengths and Limitations: (Discuss the strengths of your analysis, as well as any limitations or potential biases.)
- Future Work: (Suggest areas for improvement or further research based on your findings.)