

GROUP MEMBERS: RABIA IMRAN AND UNAIZA

ROLL NO: 21F-BSAI-11 and 21F-BSAI-40

SUBMITTED TO: SIR MUSTAFA

SUBJECT: ARTIFICIAL INTELLIGENCE

DEPT: BS-ARTIFICIAL INTELLIGENCE

Pfizer Vaccine Sentiment Analysis Using Machine Learning

INTRODUCTION

The Pfizer-BioNTech COVID-19 vaccine is critical in the global pandemic response. Analyzing public sentiment is crucial for health agencies and policymakers to address concerns and promote vaccination. We conduct sentiment analysis on Twitter data from GitHub to reveal trends and insights, guiding communication strategies.

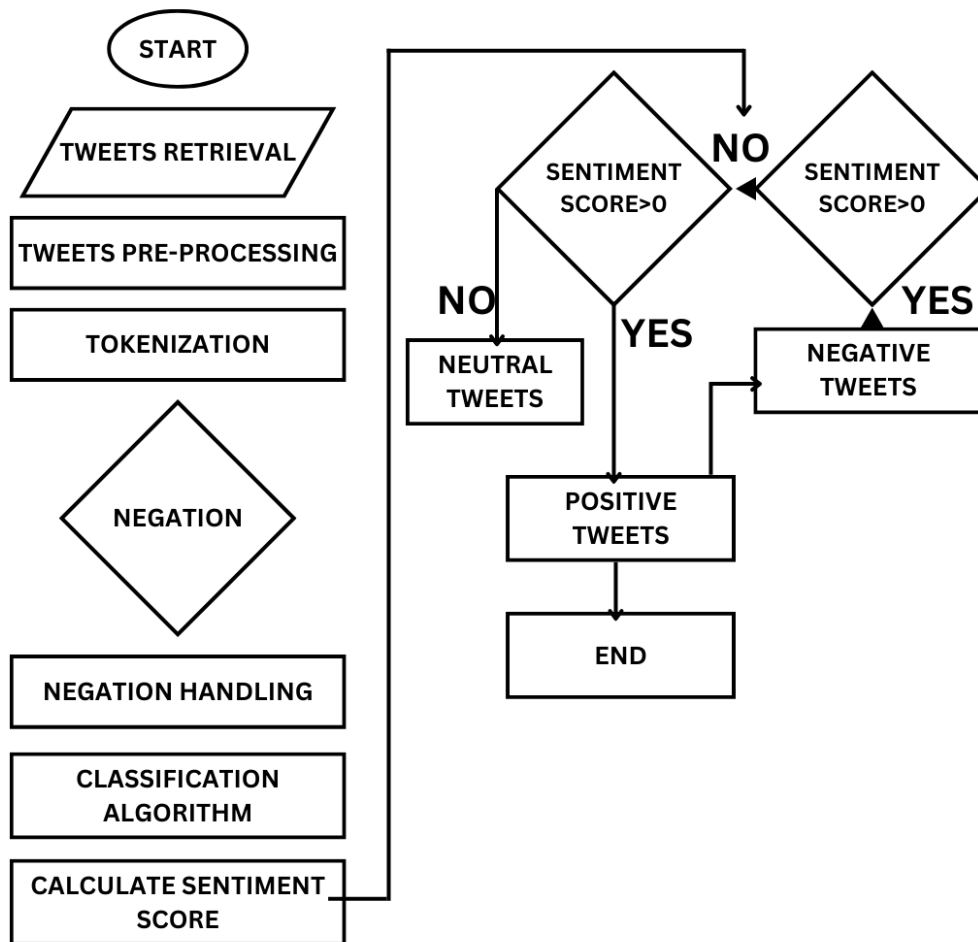
PROBLEM STATEMENT

The Pfizer-BioNTech COVID-19 vaccine has sparked diverse opinions and sentiments. To understand and address these, we analyze a dataset of tweets from GitHub. Our goal is to uncover trends, concerns, and prevailing sentiments to inform effective communication and public health strategies regarding the vaccine.

OBJECTIVE

- Identify prevailing sentiments towards the Pfizer vaccine.
- Uncover trends and patterns in public opinions and attitudes.
- Gain insights that can inform public health communication strategies.
- Provide valuable information for policymakers and health agencies to address concerns and promote vaccination.

FLOW CHART



LITERATURE REVIEW

Author	Summary
Aman Kharwal	This is how you can analyze the sentiments of Twitter users about the Pfizer vaccine. In the conclusion of this sentiment analysis, I can just say that the discussion of the Twitter users was about the awareness of the Pfizer

	vaccine rather than sharing its benefits or drawbacks. I hope you liked this article on Pfizer vaccine sentiment analysis using Python. Feel free to ask your valuable questions in the comments section below.
THOMAS KONSTANTIN	The sentiment strength of a tweet offers us a domain from which we can learn how the population is reacting to the vaccine; such insight can allow the different government to channel their advertisements towards more negative groups that usually refuse to believe in the integrity of the vaccine and observing the change over time similarly as we saw between the 13th and 27th of December.

METHODOLOGY

1. **Data Collection:** Describe how you collected the dataset from GitHub. Include details such as the keywords used for searching, the date range of tweets collected, and any filters applied to ensure the relevance of the dataset.

2. **Data Preprocessing:** Outline the steps taken to clean and prepare the dataset for analysis. This may include removing duplicates, handling missing data, and tokenizing the text.
3. **Exploratory Data Analysis (EDA):** Explain how you explored the dataset to understand its characteristics. This could involve analyzing the distribution of sentiment labels, identifying common themes or topics, and visualizing key insights.
4. **Feature Engineering:** Detail the features extracted from the text data that are used as input to the machine learning model. This could include bag-of-words features, TF-IDF features, or word embeddings.
5. **Model Selection:** Describe the machine learning models considered for sentiment analysis and the criteria used for selecting the final model. Mention any baseline models used for comparison.
6. **Model Training:** Explain how the selected model was trained using the preprocessed data. Include details on the training/validation split, hyperparameter tuning, and any cross-validation strategies employed.
7. **Model Evaluation:** Outline the evaluation metrics used to assess the performance of the model. This could include accuracy, precision, recall, F1 score, and confusion matrix analysis.

8. **Results Interpretation:** Discuss the findings of the sentiment analysis, including any insights gained into public sentiment towards the Pfizer vaccine. Highlight any patterns or trends observed in the data.
9. **Future Enhancements:** Provide suggestions for future work to improve the sentiment analysis methodology. This could include using more advanced ML models, incorporating additional features, or exploring alternative data sources.
10. **Conclusion:** Summarize the key findings of the analysis and their implications. Discuss the limitations of the methodology and any recommendations for further research.

RESULTS

RESULTS

WordCloud Visualization of Hashtags



Response

```
Positive: 743.187
Negative: 356.012
Neutral: 6982.802
```

Project Accuracy.

Accuracy: 0.8

User Verified

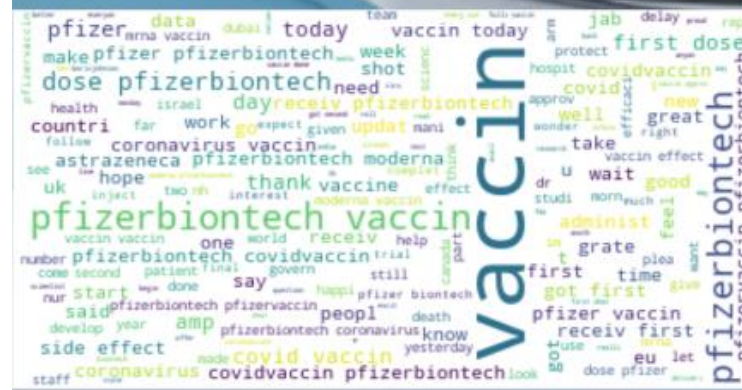
```
user_verified
False      7323
True       759
Name: count, dtype: int64
```

Sentiment Analysis

Summary.

Neutral 😊

WordCloud Visualization of Text Data



FUTURE ENHANCMENT

Emotion-aware sentiment analysis: Instead of just analyzing whether a tweet is positive, negative, or neutral towards the Pfizer vaccine, you could incorporate emotion analysis. This would involve detecting not just the sentiment but also the specific emotions expressed in the tweet (e.g., joy, anger, fear, sadness).

How to implement it: we could use a pre-trained emotion detection model or train your own using a dataset where each tweet is labeled with both sentiment (positive, negative, neutral) and emotion (joy, anger, fear, sadness, etc.). By combining sentiment and emotion analysis, you can gain a deeper understanding of how people feel about the Pfizer vaccine and potentially uncover more nuanced insights.

Example scenario: A tweet might express a negative sentiment towards the vaccine but convey a sense of hope or determination (e.g., "Despite the side effects, I'm glad I got the Pfizer vaccine"). Emotion-aware sentiment analysis would capture both the negative sentiment and the positive emotion, providing a more nuanced interpretation of the tweet's overall sentiment.