Analyzing COVID-19 Vaccine Sentiments: An Aspect-Based Sentiment Analysis (ABSA) Approach

Shubham Saha
Chalmers University
of Technology
Gothenburg, Sweden
shubhams@chalmers.se

Sifat Nawrin Nova
Chalmers University
of Technology
Gothenburg, Sweden
sifatn@chalmers.se

Sudeshna Lama
Chalmers University
of Technology
Gothenburg, Sweden
sudeshna@chalmers.se

Abstract

This project aims to analyse public sentiment on COVID-19 vaccinations using text classification. Through the use of Linear Regression, achieving an AUC score of 0.921 along with Aspect-Based Sentiment Analysis (ABSA) and TF-IDF vectorization, we identified key sentiments, uncovering both support and discouragement for vaccination. Our findings, derived from well-processed data, serve as a crucial tool for informing health communication strategies. The insights helped to observe the potential of analytical models in navigating public health challenges, guiding future efforts towards enhancing model precision and broadening the scope of analysis.

1 Introduction

After the outbreak of Covid-19 pandemic in 2020, though the innovation of covid-19 vaccine brought relief to the public, it also created a lot of controversies from the anti-vaccine and pro-vaccine groups (Miftahul Qorib 1, 2023). Text classification using NLP method,is a useful application that helps to determine the perspective or stance expressed in a piece of text towards a particular topic.

This project aims to carry forward a Aspect Based Sentimental Analysis (ABSA) to extract the postive and negative aspects from public comments regarding Covid-19 vaccine. The dataset is collected and formulated to implement a model based on a text classification algorithm. Our classification method is based on the machine learning algorithms such as Linear Regression, Bernoulli Naïve Bayes, Multinomial Naïve Bayes, and Linear SVC. Common evaluation metrics like ROC AUC, accuracy, precision, and recall are used to analyze model performance.

2 Method

2.1 Data Collection and Annotation

The data was collected by the students of the course DAT341/DIT867 through crowdsourcing information from comments in English, relating to Covid-19 vaccine from different social media and English language newspaper articles that allow public commenting. The final dataset was provided by the course teacher which is a mix of annotations and comments after the first and second round given by the students. To assess the consensus between annotators in our dataset, we've utilized Krippendorff's Alpha, which is a measure of inter-rater reliability. With a value of approximately 0.735 according to our calculation, it suggests a moderate to substantial level of agreement among annotators, making them suitable for further analysis and modeling.

2.2 Data Preprocessing

2.2.1 Comment Preprocessing

A key part of our method is to change the text data into a format that machine learning programs could use. The first step of data preprocessing is data cleaning. This process includes:

- Converting all text to lowercase.
- Expanding contractions.
- Removing non-alphanumeric characters, including punctuation and special symbols.
- Stripping emojis and URLs.
- Applying tokenization and lemmatization to break down text into its base components and reduce words to their root forms.

2.2.2 Annotation Preprocessing

After that we have simplified the annotations. As we received a number of annotations on every comment, so for the sake of simplicity to use the valid annotations we measured the frequency of each annotation. Then we filtered out annotations that appeared less than 100 times. Maintaining a refined dataset we proceeded to simplify the annotation by converting the multi annotation value to a single annotation value. As we are aiming for binary classification, we dropped all the rows that contains the annotation -1, thus having a perfectly cleaned dataset which is ready for model training and testing.

2.3 Data Representation

After cleaning, we converted the text data into numerical vectors using Term Frequency-Inverse Document Frequency (TF-IDF) vectorization. This method weighs words based on their document frequency to focus on the most informative features, thus reducing data complexity. We used scikit-learn's TfidfVectorizer to transform the comments into a sparse matrix of TF-IDF features for machine learning analysis.

3 Model Training and Analysis

3.1 Selection of the Model

The selection of learning algorithms for this project was a strategic approach aimed at identifying the most effective algorithm for our text classification task. We selected Linear Regression, Multinomial Naive Bayes, Bernoulli Naive Bayes, and Linear SVC. An essential aspect of our evaluation involved comparing the performance of our text classification system against a trivial baseline. For that we used DummyClassifier as our baseline model.

Linear Regression: Linear Regression is primarily known for its predictive modeling and continuous outcome prediction. This approach involves representing each document as a vector in a high-dimensional space. The dimensions here corresponds to the features extracted from the text, often through TF-IDF vectorization, which quantifies the importance of words within the documents relative to a corpus.

Bernoulli Naive Bayes: The Bernoulli Naive Bayes classifier is designed for binary features. In the context of text classification with TF-IDF, it considers the presence or absence of features (words).

Multinomial Naive Bayes: The Multinomial Naive Bayes classifier is more directly suited for counting words and thus it works naturally with TF-IDF vectors, which are essentially weighted counts.

Linear Support Vector Classification (Linear SVC): Linear SVC is a non-probabilistic approach, variant of Support Vector Machine (SVM) that is specifically designed for binary classification tasks. It has capacity for achieving high accuracy in distinguishing between categories.

3.2 Hyperparameter tuning

Hyperparameter tuning was essential to enhance our models' accuracy and effectiveness. We employed GridSearchCV on our algorithms for an exhaustive parameter search, combined with 5-fold cross-validation and n_jobs = -1 for full processor utilization for speed. Our primary metric was ROC AUC for its class distinction capability, selecting the highest score as the optimal model configuration.

3.3 Evaluation

To thoroughly assess the quality of our classification system, we used several key metrics like AUC (area under the ROC curve), Accuracy, Precision, and Recall. However, we have used AUC as our primary evaluation metric.

3.4 ABSA Performance

To evaluate the effectiveness of our model we executed ABSA. This process was fundamental in understanding how well our model could generalize its sentiment predictions across unseen data, thereby providing a robust measure of its predictive accuracy and reliability. Our core strategy was to extract relevant aspects from the text and predict the sentiment associated with it. Utilizing spaCy for NLP to extract aspects for sentiment analysis. The model then assessed the sentiment, categorizing them as either 'Positive' or 'Negative'.

Topic Modeling with NMF: After ABSA, we used Non-Negative Matrix Factorization (NMF) to show sentiment distribution across topics, revealing the nature and sentiment of discussions with 20 distinct topics for a detailed yet concise analysis.

Sentiment Analysis Enhancement: We refined our sentiment analysis by filtering out irrelevant terms, reducing noise, and enhancing the precision of our sentiment assessment for a focused analysis.

Sentiment Counter Initialization: Our analysis included counters to measure sentiment in comments, helping us identify key positive and negative aspects. This provided clear insights to improve positive perceptions and tackle negative issues.

4 Results

After completing GridSearchCV, we found that Linear Regression holds the maximum ROC AUC score of 0.8934 across all models. The best parameters for this was, C: 1, max_iter: 100, solver: saga. A comparison chart of the GridSearchCV results for all the models is shown in Figure 1.

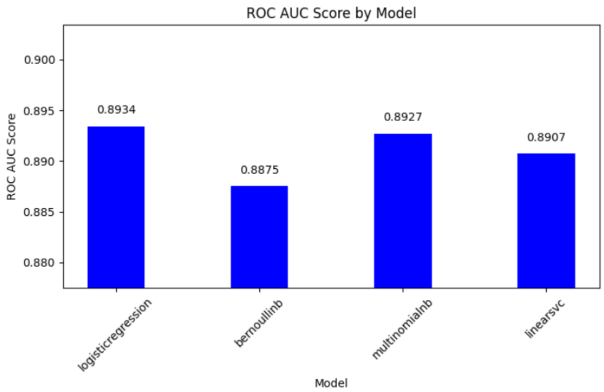

Figure 1: ROC AUC comparison

Also based on the borderline which we have selected Dummy classifier, in Table 1 the evaluated the performance of the trained model using their best parameters on the test set as shown.

From Table 11, we conclude that Linear Regression and Linear SVC had the best (Area under ROC curve) AUC score, 0.923 and 0.921. However Linear Regression performed better in all metrics of evaluation and we created a confusion matrix, plotted it using a heatmap in Figure 2 to investigate its errors further.

The model correctly predicted 857 instances of the positive class. The model correctly predicted 854 instances of the negative class. The model incorrectly predicted 166 instances as positive that were actually negative. The model incorrectly predicted 162 instances as negative that were actually positive.

4.1 Error Analysis

False Positives: It might be that the model picked up on certain keywords or phrases that it associated with positive sentiment, but in context, these did not truly indicate a positive outcome. It could also be that the model is overfitting to certain features in the training data that aren't generalizable to the test data.

False Negatives: This could be because the positive instances are more nuanced or less represented in the training data, leading the model to be less sensitive to the features that indicate a positive outcome. It could also be that negative instances share common features with positive ones, making it harder for the model to distinguish between them accurately.

4.2 ROC Curve

Our model's ROC curve, with an AUC of 0.921, shows a strong distinction between positive and negative classes, confirming the effectiveness of our feature selection and training. It is represented in Figure 3.

Our analysis revealed a distinct sentiment within the public comments about COVID-19 vaccines. By employing ABSA, we were able to classify the sentiments associated with the most prevalent aspects(shown in Figure 4 and Figure 5).

Table 1: Performance results on test set using AUC, accuracy, precision, and recall.

Model	**AUC**	**Accuracy**	**Precision**	**Recall**
Dummy	0.500	0.500	1.00	0.0
Linear Regression	0.923	0.839	0.84	0.84
Bernoulli NB	0.913	0.832	0.82	0.85
Multinomial NB	0.916	0.833	0.86	0.80
Linear SVC	0.921	0.836	0.84	0.83

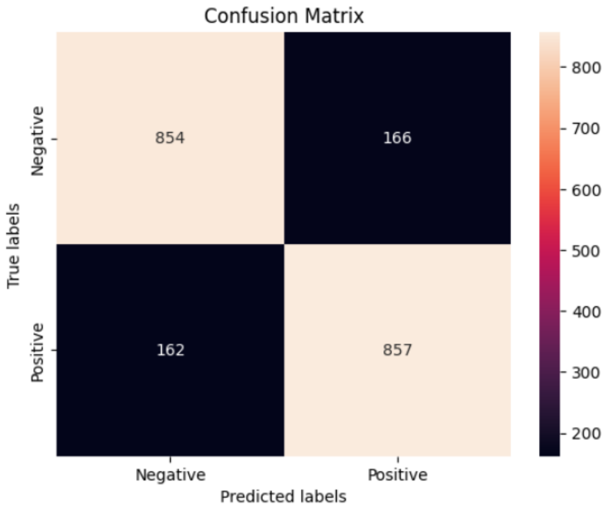

Figure 2: Confusion matrix for Linear Regression model

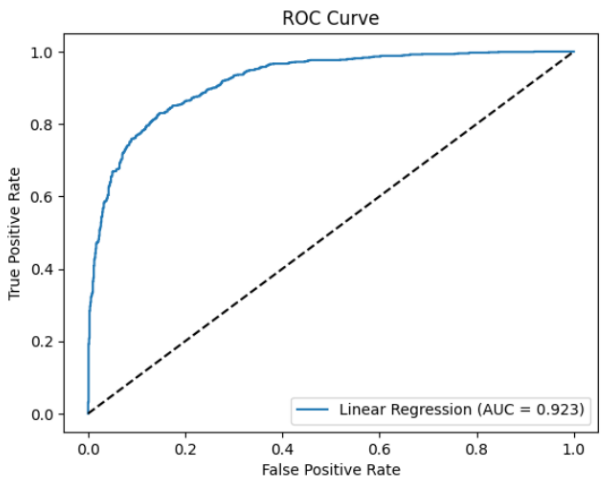

Figure 3: ROC Curve of Linear Regression

4.3 Positive Sentiment Aspects

The discourse positively highlighted 'booster' shots, reflecting trust in their efficacy. 'Science', 'Moderna', and 'Pfizer' were also seen positively, indicating confidence in these vaccines. Aspects like 'hope' and 'second dose' suggest optimism about vaccine effectiveness, while 'fact' and 'vaccinated people' show a community valuing factual information and progress through vaccination.

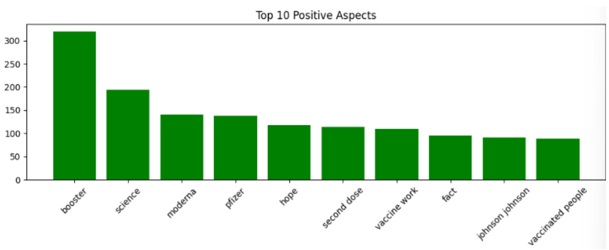

Figure 4: Positive Sentiment Aspects

4.4 Negative Sentiment Aspects

Negative sentiments focused on 'god' and 'virus', suggesting a sense of helplessness. Concerns about 'mRNA' vaccines and 'side effects' indicate fear of new technologies and adverse reactions. Distrust was also aimed at the 'FDA', and terms like 'jab', 'poison', 'death', 'immune system', and 'mask' reflect skepticism towards vaccination and pandemic handling.

5 Discussion

In our project, we achieved a best AUC score of 0.921 using Linear Regression model that underscoring the effectiveness in distinguishing sen-

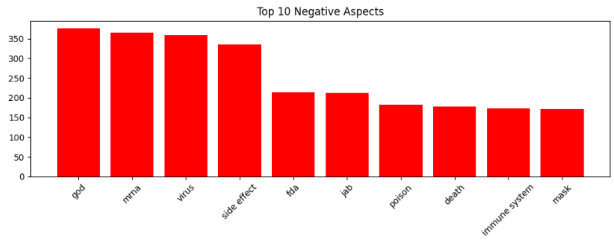

Figure 5: Negative Sentiment Aspects

timent with high precision, with ABSA playing a crucial role in deciphering the nuanced public sentiment towards COVID-19 vaccinations. The preprocessing phase, including text cleaning, annotation and the use of TF-IDF vectorization, laid the groundwork for accurate data representation and analysis. Our ABSA results notably highlighted 'booster shots' and 'science' as top positive aspects, reflecting public trust and optimism, whereas 'mRNA vaccines' and 'side effects' emerged as primary concerns, indicating prevalent apprehensions. Despite limitations such as potential dataset biases and the representativeness of online comments, our findings offer insightful perspectives on public sentiment and model efficacy in textual analysis. Future work could expand dataset diversity and integrate more sophisticated language processing methods, aiming to refine the accuracy and applicability of our sentiment analysis.

6 Conclusion

Our project's analysis of COVID-19 vaccination sentiments concluded with a high AUC score of 0.921 and an impressive accuracy using linear regression, showcasing our model's ability to effectively differentiate sentiments. Despite challenges like dataset bias and capturing language nuances, our use of TF-IDF vectorization and ABSA provided deep insights into public perceptions. Future efforts will focus on expanding data diversity and exploring advanced techniques to further improve our model's accuracy and applicability in public health discourse.

References

Max Denis 3 Esther Ososanya 4 Paul Cotae Miftahul Qorib 1, Timothy Oladunni 2. 2023. *Expert systems with applications*, 212(118715).