TweetClass: COVID-19 Vaccine Tweet Classification with scikit-learn

Baivab Chakraborty[1], Subhajit Srimani[2] and Souvit Biswas[3]

[1]*Amity University, Kolkata, West Bengal, Kolkata 700135*

[2]*Netaji Subhash Engineering College, Kolkata, West Bengal, Kolkata 700152*

[3]*Amity University, Kolkata, West Bengal, Kolkata 700135*

Abstract

Our team has proposed TweetClass as a solution in the FIRE 2023 AISoMe Track. We propose to utilize the sci-kit-learn library, which consists of several classifiers that can be used to categorize tweets as Unnecessary, Mandatory, Pharma, Conspiracy, Political, Country, Rushed, Ingredients, Side-effect, Ineffective and Religious. As the vaccination process continued globally to combat COVID-19, analyzing people's tweets seemed to provide valuable insights into their opinions on the entire vaccination episode. This enormous dataset, on correct utilization, can help the government create effective vaccination strategies in case of future pandemics. Our submitted model achieves a Macro-Fl score of 0.39 and a Metric Jaccard score of 0.46, earning us one of the top spots amongst other submissions.

1. Introduction

The world faced its toughest challenge in the form of the COVID-19 pandemic. Over time, vaccines have proven to be a safe and effective way to combat and eradicate infectious diseases. As a result, there emerged a race to discover effective vaccines that could prevent the havoc of COVID-19 and eventually, this led to the worldwide availability of these vaccines. However, the discussions concerning vaccination progress, accessibility, efficacy, and side effects had been ongoing, and people had both positive and negative opinions about it. Some took to social media sites, like Twitter, to share their concerns regarding the vaccine and it proves beneficial for the governments and health organizations, like WHO, to understand people's thoughts regarding the new COVID-19 vaccines. They look to use such insights to plan their future strategies and encourage everyone to get fully vaccinated. It has been crucial to stop spreading misinformation about these vaccines and appreciate the efforts of governments to have worked to restrict the pandemic from spreading further. Twitter had also attempted to block tweets that contained incorrect or misleading information about the virus, its preventative measures, and treatments. The manual classification of tweets has proved time-consuming and prone to

error. Therefore, there had been an urgent need to develop machine-learning models that could assist in classifying tweets about the COVID-19 vaccines.

2. Task

The Artificial Intelligence on Social Media (AISoMe) 2023 Track, organized as a part of the 15th meeting of Forum for Information Retrieval Evaluation (FIRE) 2023, has tasked us to build an effective multi-label classifier to label a social media post (particularly, a tweet) according to the specific worries about vaccines that the post's author stated, and we use this research to show how we think the problem might be solved. The tweets consist of distinct concerns towards vaccines owing to reasons such as the politics involved, potential side-effects of vaccines, etc. To be precise, the tweets are categorized into the following 12 classes, described with instances:

- `Unnecessary`: Suggests that vaccines are suspected to be unnecessary. Example: *Besides the fact that those who already had covid are immune, so the vaccine will not be needed. The 'vaccine' ain't the only solution to immunity, in fact it doesn't even provide immunity, just the symptoms are decreased.*
- `Mandatory`: Suggests that it is preferred if vaccination is not mandatory. Example: *A vaccine passport will not be issued for those who don't want vaccination, how are you gonna deal with these people, your objective seems like to push us to one corner and penalise, I suppose this is what you call justice.*
- `Pharma`: Suggests that COVID-19 is hoax for the big pharmaceutical companies to earn more money. Example: *Fauci and Bill Gates are the real owners. Fauci owns a stake in Moderna through NIH, and Gates invested 100 million dollars in 2016 for the development of mRNA vaccines.*
- `Conspiracy`: Suggests that there is some deeper conspiracy than just pharmaceutical companies earning money. Example: *First they came for your minds, then your freedoms with Lockdown, your rights to earn a living, then your body with vaccination. Now they are coming for your kids. Finally they will own everything you posses including you.*
- `Political`: Suggests that there is an attempt of propagating some political agenda. Example: *We don't need any twisted, insane politicians directing us whether or not to get vaccinated. They do not have the right to instruct us on how to live our lives. The best course of action is to push back while ignoring the absurd story.*
- `Country`: Suggests that the vaccine is denied because of the country of origin or manufacturing. Example: *I regret to inform you that I will never take a vaccine produced in Russia.*
- `Rushed`: Suggests that the vaccines lack proper testing or that inaccurate data is published. Example: *Since these tests typically take years, the long-term safety of any vaccination cannot possibly be tested. What will make it work? Why do you think they can heal diseases when they can't even treat the common cold despite years of research?*
- `Ingredients`: Suggests that the ingredients of the vaccines are worth concerned. Example: *Is created using the kidney cells of a young girl who was aborted in the 70s.*

- `Side-effect`: Suggests that there are side-effects of being vaccinated including death. Example: *It starts. Please look for secure substitutes for this vaccine. Following patient illnesses, the UK publishes an allergy warning for the Pfizer COVID-19 vaccination.*
- `Ineffective`: Suggests that the vaccines may be ineffective altogether. Example: *What exactly is the point? Are you giving your old and frail patients a second shot because Pfizer claims the vaccine is worthless if administered outside of the recommended time frame?*
- `Religious`: Suggests that there are religious restrictions against the vaccines. Example: *A vaccine would go against my religion. The strongest defense available is the 91st Psalm.*
- `None`: The tweet doesn't coincide with any of the above reasons or with those not stated above. Example: *Total garbage. I could have taken the vaccine but will pass this one*

3. Related Work

Information extraction from social media posts containing textual data has become an integral part of social computing. Social media upholds the diverse opinions and expressions of people regarding almost every aspect of this world and, the concerns about the COVID-19 vaccines is no exception. Hence, the use of textual tweets as data to classify people's stances on vaccines employing machine learning methods like Multinomial Naive Bayes Classifier [1] and Multi-Output Classifier [2] is very efficient and useful.

3.1. Multinomial Naive Bayes

Multinomial Naive Bayes Classifier is a probabilistic machine learning algorithm based on Bayes' theorem. and assumes that features are independent and follow a multinomial distribution. It is vital for roles like text categorization and sentiment analysis, where the frequency of words in documents is essential for classification.

3.2. Multi-Output Classifier

A Multi-Output Classifier is a machine learning model designed to handle multiple target variables simultaneously, making it suitable for multi-label or multi-task learning problems. Instead of predicting a single output, it produces multiple outputs, each corresponding to a different target variable. Multi-output classifiers are used in various domains, including natural language processing, where multiple aspects or labels need to be predicted for a single input. They extend traditional classification and regression techniques to tackle complex, multi-dimensional prediction tasks.

4. Dataset

A training dataset, containing 9921 tweets exhibiting the varied apprehensions of people about the vaccines, is provided. Since it is used for training the model, the tweets in this file possess both their IDs and their respective classes. The test dataset only contains 486 tweets along with their respective IDs. Our approach makes the best use of the CAVES dataset [?] through

various pre-processing techniques utilized and application of appropriate classifiers to develop an efficient model.

4.1. Trends in the Dataset

On examining the training dataset, it can be stated that *side-effects* emerges as the most frequent class of concerns with a whopping 38.4% portion, followed by *ineffective* (16.9%), *rushed* (14.9%), *pharmaa* (12.8%), *mandatory* (7.9%), *unnecessary* (7.3%), *none* (6.3%), *political* (6.3%), *conspiracy* (4.9%), *ingredients* (4.4%), *country* (2.0%) and *religious* (0.7%).

5. Pre-processing

A textual tweet tends to comprise of plain text, special characters and emojis which must be dealt with otherwise, this suppresses the performance of the model. Therefore, we pre-processed the tweets to enhance the quality of the data for further processing and reduce any chances of hampering performance. The following are the steps of pre-processing adopted in our model:

- `Feature and Lable Extraction`: Feature extraction selects and transforms the raw data from the tweets into a set of features that can be used as input for the applied machine learning algorithms. Similarly, the labels are also extracted which the model will predict.
- `Vectorization`: The TF-IDF Vectorizer further transforms the text data into a numerical format that the machine learning algorithms, the model uses, can understand. It also removes the English stop words from the text that does not add any extra meaning to the sentence on their own.
- `Binarization`: The Multi-Label Binarizer assists in the multi-label classification task as required in case of our dataset where each tweet can belong to more than one class. It converts these categories into binary labels (0 or 1) for each class.

6. Methodology

6.1. Model

We have made a model that executes multi-label text classification. We have basically used the scikit-learn library and its versatile features such as, the TF-IDF Vectorizer (converts the collection of raw documents to a matrix of TF-IDF features) and the Multinomial Naive Bayes Classifier (particularly useful in this case as the data set involves text data with discrete features such as word frequency counts) to perform multi-label text classification. Moreover, the implementation of Multi-Output Classifier employs one classifier per target (multi-target classification) and Multi-Label Binarizer converts the labels into a binary matrix representation, facilitating the multi-target classification process. Overall, the use of such versatile features from scikit-learn fine tunes our model enabling efficient prediction of the test dataset.

6.2. Experimental Setup

The training data is split into training and validation sets in the ratio 9:1 and the data are shuffled prior splitting so that the fraction of instances of each class are preserved in both sets. We have already explained in section 5 about the pre-processing techniques applied on the data and this training data is used to fine tune our model while the validation data is used for evaluation.

6.3. Prediction

We have designed our model to analyze and predict the available test dataset, stored in "test_data" data frame. In the dataset, the "tweet" column stores the extracted text data, and the corresponding labels are stored in a new column named "pred_labels". Then the data frame with predicted labels is saved in a new CSV file named "prediction_file.csv" and the "tweet" column is dropped from the data frame. We have also used the feature accuracy_score to calculate accuracy of each predicted label. Finally, the model opens the CSV file with predictions and reads the contents into a data frame named "result" and displays the output in the console.

7. Evaluation

AISoMe FIRE 23 Track results are evaluated using Macro-F1 score for primary evaluation and Metric Jaccard for secondary evaluation in case of tie with Macro-F1 score. The result of our two submitted automated run for the prediction of the test dataset is shown in Table 1. Our model got the 34th and 38th rank based on our two submitted run files.

Table 1
Result of AISoMe FIRE 23 Track

Sr No.	Team_Name	Macro-F1 Score	Metric Jaccard	Rank
1	APS AI&ML	0.39	0.41	34
2	APS AI&ML	0.35	0.46	38

8. Conclusion and Future Work

This paper illustrates our TweetClass model. This model, upon construction and training, performs multi- label text classification to successfully execute the prediction of an available test dataset. The model uses the versatile features of scikit-learn to perform natural language processing. We pre-processed our data through vectorization, binarization and splitting the same then, performed feature and label extraction on the resultant data. This was followed by training of our model using the training data and finally using the model for prediction of the test dataset. We also look forward to further optimizing the model by exploring the future scopes of machine learning and artificial intelligence. Our future endeavors include using more domain specific models to improve the versatility of TweetClass thus, resulting in more optimized outputs and higher accuracy scores.

References

[1] J. P. D. Delizo, M. B. Abisado, M. I. P. De Los Trinos, Philippine twitter sentiments during covid-19 pandemic using multinomial naïve-bayes, International Journal 9 (2020).

[2] J. Read, L. Martino, P. M. Olmos, D. Luengo, Scalable multi-output label prediction: From classifier chains to classifier trellises, Pattern Recognition 48 (2015) 2096–2109.

[3] S. Poddar, A. M. Samad, R. Mukherjee, N. Ganguly, S. Ghosh, Caves: A dataset to facilitate explainable classification and summarization of concerns towards covid vaccines, in: Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval, 2022, pp. 3154–3164.

[4] S. Poddar, M. Basu, K. Ghosh, S. Ghosh, Overview of the fire 2023 track:artificial intelligence on social media (aisome), in: Proceedings of the 15th Annual Meeting of the Forum for Information Retrieval Evaluation, 2023.