Multi-label Classification of Covid-19 Vaccine Tweet

Palvika Bansal[1], Sumit Das[1], Vikas Rai[1] and Shalini Kumari[1]

[1]*Thomson Reuters Lab, Bangalore, India*

Abstract

This research paper presents a novel approach to multi-label classification of tweets expressing concerns about Covid-19 vaccines. It introduces fine-tuned BERT based model, customized for this task, which achieves good performance in accurately categorizing specific concerns within tweets. Through extensive data preprocessing, the model accommodates a wide range of concerns. Our findings have significant implications for public health communication, as they enable precise monitoring of public sentiment and vaccine-related concerns. This research contributes to natural language processing and demonstrates the practical application of advanced machine learning techniques in addressing real-world challenges. It underscores the potential for innovative AI-driven solutions in public health communication.

Keywords

COVID-19 Vaccine Tweets, Sentiment Analysis, Multi label Classification, BERT, Prefix-Tuning

1. Introduction

Vaccination plays a crucial role in mitigating the risk and transmission of a wide range of diseases. Over the past few years, vaccination has emerged as a critical tool in combating the COVID-19 pandemic. Moreover, large-scale vaccination efforts are essential to reduce the prevalence of various diseases. Nonetheless, skepticism towards vaccines persists among many individuals, primarily due to a variety of reasons, including political factors and concerns about potential vaccine side effects.

It is imperative to acknowledge and address these diverse concerns surrounding vaccines. Social media platforms have proven to be invaluable sources of data for gauging public sentiment and opinions regarding vaccination. Leveraging platforms like these allows us to rapidly gather insights from conversations and discussions about vaccines [1]. To facilitate this understanding, our work has utilized training data sourced from a prior project called ”CAVES: A dataset designed to facilitate the transparent classification and summarization of concerns related to COVID vaccines.” [2].

Our rigorous methodology entailed a systematic experimentation with a wide spectrum of techniques in the realms of deep learning and machine learning. We experimented with these approaches to facilitate the precise categorization of tweets that revolved around vaccine-related concerns. Within our experimental framework, we started with foundational models including TF-IDF and LSTM and advanced towards more contextual models which involved BERT [3] based models. One noteworthy experimentation involved the implementation of prefix

Forum for Information Retrieval Evaluation, December 15–18, 2023, India

palvika.bansal@thomsonreuters.com (P. Bansal); sumit.das@thomsonreuters.com (S. Das); vikas.rai@thomsonreuters.com (V. Rai); shalini.kumari@thomsonreuters.com (S. Kumari)

tuning [4], a refinement technique integrated with state-of-the-art transformer models. This intricate synergy enabled us to extract nuanced insights from the tweets under examination, enhancing the accuracy and depth of our classification efforts. To further extract the contextual meaning of tweet, we experimented with various data processing approaches such as identifying named entities in the tweets, expansion of tweets, analyzing sentiment of tweet and analysis of keywords in the tweets. We also experimented with state of the art GPT-4 [5] model to identify concerns related to the tweet by providing it with few-shot examples.

Furthermore, our investigative pursuits were not confined solely to the broad spectrum of techniques. We ventured into the specialized domain of model fine-tuning to accommodate the idiosyncrasies inherent in tweet data. This approach allowed us to harness the unique characteristics of Twitter's concise and informal language style, ensuring our models were finely attuned to capture the subtle intricacies of vaccine-related discourse.

2. Task

Our primary aim is to develop a highly efficient multi-label classification model that can accurately assign labels to a social media post, specifically tweets. These labels will correspond to the specific concerns and sentiments expressed by the post's author regarding vaccines. This task involves not only identifying the presence of various concerns but also understanding the nuances and context in which they are discussed, enabling a comprehensive analysis of public sentiment and discourse surrounding vaccines on social media platforms.

In the context of this study, the classification task is centered around a set of predefined concerns pertaining to vaccines. These concerns serve as the labels for categorizing social media (tweet) posts, providing a structured framework for analyzing and understanding public discourse on vaccine-related topics. To gain deeper insights, kindly refer to the following topics:

- **Unnecessary:** The tweet indicates vaccines are unnecessary, or that alternate cures are better.
- **Mandatory:** Against mandatory vaccination — The tweet suggests that vaccines should not be made mandatory.
- **Pharma:** Against Big Pharma — The tweet indicates that the Big Pharmaceutical companies are just trying to earn money, or the tweet is against such companies in general because of their history.
- **Conspiracy:** Deeper Conspiracy — The tweet suggests some deeper conspiracy, and not just that the Big Pharma want to make money (e.g., vaccines are being used to track people, COVID is a hoax)
- **Political:** Political side of vaccines — The tweet expresses concerns that the governments/politicians are pushing their own agenda though the vaccines.
- **Country:** Country of origin — The tweet is against some vaccine because of the country where it was developed/manufactured
- **Rushed:** Untested/Rushed Process — The tweet expresses concerns that the vaccines have not been tested properly or that the published data is not accurate.

- **Ingredients:** Vaccine Ingredients/technology — The tweet expresses concerns about the ingredients present in the vaccines (eg. fetal cells, chemicals) or the technology used (e.g., mRNA vaccines can change your DNA)
- **Side-effect:** Side Effects/Deaths — The tweet expresses concerns about the side effects of the vaccines, including deaths caused.
- **Ineffective:** Vaccine is ineffective — The tweet expresses concerns that the vaccines are not effective enough and are useless.
- **Religious:** Religious Reasons — The tweet is against vaccines because of religious reasons
- **None:** No specific reason stated in the tweet, or some reason other than the given ones.

3. Related Work

Users frequently turn to micro-blogging platforms such as Twitter, motivated by a diverse range of objectives. These include expressing their viewpoints on the Coronavirus pandemic, disseminating personal health updates to their online connections, flagging symptoms, and sharing alerts regarding their well-being or that of acquaintances. Robust discussions take place concerning COVID-19 vaccines and vaccination campaigns, often preceding individuals' receipt of their vaccine doses. The extraction of valuable insights from these textual tweets represents a common application within the field of social computing.

In the realm of text classification, traditional machine learning techniques such as the Naive-Bayes classifier, Linear classifier, Support Vector Machine (SVM), and cutting-edge deep learning methods including Long Short Term Memory (LSTM) networks and Bidirectional Recurrent Neural Networks (RNNs) have demonstrated their effectiveness.

Recent advancements in natural language processing have given rise to notable language models, with BERT (Bidirectional Encoder Representations from Transformers) [3] and its domain-specific counterpart CT-BERT (COVID-Twitter-BERT) [6] at the forefront. Additionally, VaccineBERT [7], a BERT-based model specialized in classifying COVID-19 vaccine-related tweets, has garnered attention.

4. Dataset

The dataset in its entirety consists of 9,921 tweets records, and it is worth noting that there are no missing values within this dataset, ensuring a comprehensive and complete collection of Twitter data for analysis.

4.1. Data Exploration

Within the scope of this classification task, it is imperative to acknowledge that individual tweets may be linked with multiple labels. Consequently, it is of utmost importance to undertake a comprehensive examination of the distribution of these labels within the dataset. This understanding is vital for effectively categorizing and interpreting the complex and diverse nature of the tweets in our dataset.For an in-depth analysis and a complete overview of the results from this analysis, Refer Table 1.

Table 1
Label distribution within dataset

Label	**Count**
side-effect	3805
ineffective	1672
rushed	1477
pharma	1273
mandatory	783
unnecessary	722
none	629
political	626
conspiracy	487
ingredients	436
country	201
religious	64

In addition to this, it is crucial to examine the distribution of the number of labels assigned to each individual tweet. Upon analyzing the entire dataset, we observed that approximately 7,936 tweet texts were assigned only one label, indicating a prevalent singularity of classification. Furthermore, around 1,716 tweets exhibited a dual-label configuration, suggesting a moderate level of complexity in label assignment. Intriguingly, a subset of 269 tweets challenged this convention by being concurrently linked to three distinct labels, underscoring the presence of intricately categorized content within the dataset. This meticulous examination of label distribution not only enhances our understanding of the dataset's characteristics but also provides valuable insights into the diverse nature of the classification challenge at hand.
Furthermore, we have undertaken an examination of the distribution of tweet lengths. For a more comprehensive view of the length distribution, Refer to the Appendix Figure 1.1.

4.2. Trends in the dataset

Label-Entity Mapping in Tweet Text An analysis aimed at mapping training data labels to the most prevalent entity types found within the tweet text. This analysis was carried out for both individual training data labels and when multiple labels were present. For a comprehensive breakdown of this analysis and its results you can refer to Appendix Tables 1.1 and 1.2.
Extraction and Parsing of URL-Embedded HTML Content in Tweet Text We performed a two-fold analysis involving the extraction of URLs from tweet text and the subsequent parsing of HTML content from these URLs. The purpose was to examine the HTML content, particularly the headlines, associated with each URL and compare it with the tweet text. It was observed that the majority of these URLs referred to either other tweet threads or news media reports. Among the complete list of URLs, approximately 20% of the web pages were found to be non-existent.

In the course of our analysis, we discovered that in most cases, the tweet text was concise and often a partial excerpt from the parsed URL contents. Additionally, there were instances where the context of the tweet text contradicted the information present in the HTML content of the URLs. Consequently, we arrived at the conclusion that incorporating this HTML content into the tweet text would not provide added value and could potentially introduce confusion to the model.

Analysis of @Mentioned Users in Tweet Text Furthermore, we conducted an analysis of the mentions of user profiles (@user) within the tweet text. The intention was to explore whether the profiles of mentioned users could offer supplementary information related to the type of tweet. However, it is important to note that our efforts were hindered by the unavailability of data due to restrictions imposed by the Twitter API, which prevented access to user profiles.

Exploring Entity Types Within Tweet Text In the course of our research, we leveraged the Hugging Face's bertweet-tb2_wnut17-ner API as a cornerstone for detecting entities within the tweet texts. This API, tailored for the intricacies of social media data, harnessed the power of advanced Named Entity Recognition (NER) techniques, specifically fine-tuned for Twitter contexts, to accurately categorize entities amid the informality, hashtags, and mentions characteristic of tweets. However, it's noteworthy that given the constraints of time, our exploration did not yield significant outcomes, warranting further investigation in the future.

```
[
    {'entity_group': 'PER', 'score': 0.9401, 'word': 'JoshBloom@@'},
    {'entity_group': 'ORG', 'score': 0.8423, 'word': 'Pfizer'}
],
[
    {'entity_group':'ORG', 'score': 0.9042, 'word': 'Lifesitenews'}
]
```

Label association with tweet length Investigated the Correlation Between Label Assignments and Tweet Length to Explore Potential Label Preferences in Terms of Tweet Length. However, No Direct Relationship Between Label and Tweet Text Length Was Identified. Analysis and its results you can refer to Appendix Table 1.3.

Sentiment Analysis Used SentimentIntensityAnalyzer package for a quick sentiment analysis. However, given that all labels were primarily linked to negative emotions, the average sentiment scores across the board leaned toward negativity. Analysis and its results you can refer to Appendix Table 1.4.

Determining Key Terms in Tweets for the Primary Entity Group of Each Label Examined the prevalent words associated with each class label, omitting stopwords and applying tweet cleaning. For additional details on this analysis and its outcomes, consult the Appendix Figure 1.2.

5. Pre-processing

To enhance the quality of word embeddings that we leveraged in modeling process, we pre-processed the tweets. Tweets generally encompass distinctive lexical elements such as hashtags,

@username mentions, URLs, RT and special characters. These elements, if left unprocessed, tend to hinder the model's performance. Consequently, we implemented a specific data cleansing procedure as an integral component of our tweet pre-processing strategy within the dataset:

- **Removing stop words:** In this phase, stop words, which are commonly used words such as ”the,” ”and,” and ”in,” are systematically removed from the text. We also removed some words specific to tweets data such as rt which depicts retweets. This step helped in reducing noise and improving the efficiency of the tasks by focusing on the most meaningful words and phrases in the text.
- **Removing URLs:** Initially we explored using external URL content to enhance tweet meaning but it didn't add much value to the core meaning of tweet and was distorting results. So, We removed these extraneous web links using regular expression.
- **Removing Username mentions:** Removing username mentions in tweets analysis data is crucial to preserve privacy and reduce bias, as mentions often refer to specific individuals or accounts. This step ensured that the analysis remains impartial.
- **Convert words to lowercase:** Converting words to lowercase in tweets analysis data standardizes text and enhances consistency, ensuring that words with different capitalization patterns are treated as identical. This step prevents discrepancies in analysis and simplifies text processing.
- **Remove non-alphanumeric characters:** We removed special symbols, and punctuation marks that often don't contribute significantly to the analysis. This step helped in focusing on the core linguistic content.
- **Tweet text expansion:** For labels with less data, We utilized GPT-3.5 to augment tweet content for labels such as country, political, conspiracy, religious, and none, in order to provide richer context and enhance the relevance of the tweet in accordance with its label. This initiative aims to assess whether text expansion can contribute to the enhancement of the model's performance, particularly for these challenging labels. For additional details on this analysis and its outcomes, consult the Appendix Table 1.5.

6. Methodology

6.1. Models

Fine Tuning DeBERTa Large: In one of our experiments we finetuned DeBERTa (Decoding-enhanced BERT with Disentangled Attention) [8] ”large” variant. It builds on RoBERTa [9] with disentangled attention and enhanced mask decoder training with half of the data used in RoBERTa. It is a Transformer-based neural language model that aims to improve the BERT [3]and RoBERTa models with two techniques: a disentangled attention mechanism and an enhanced mask decoder. The disentangled attention mechanism is where each word is represented unchanged using two vectors that encode its content and position, respectively, and the attention weights among words are computed using disentangle matrices on their contents and relative positions. The enhanced mask decoder is used to replace the output softmax layer to predict the masked tokens for model pre-training. In addition, a new virtual adversarial training method is used for fine-tuning to improve model's generalization on downstream tasks. We used max length as 128 with padding to right. We used learning rate ad 2e-5 and batch size of 10 to fine tune model for 15 epochs. To prevent overfitting, We used early stop monitoring the validation loss with patience value 5.

Prefix Tuning of RoBERTa Large: In our experiment, we employed the RoBERTa (A Robustly Optimized BERT Pretraining Approach) [9] ”large” variant, which is among the state-of-the-art transformers in the domain of natural language processing. RoBERTa builds on BERT [3] model architecture using a more effective training procedure and was trained on a much larger dataset. This variant is pre-trained on 160GB of text from the BookCorpus, OpenWebText, English Wikipedia etc., making it adept at grasping linguistic nuances and contextual representations of text. We chose prefix tuning [4] for RoBERTa large because it allows us to adapt the pre-trained model for our specific multi-label classification task without overhauling the underlying patterns the model had previously learned. By adding a task-specific prefix to the input sequence, prefix tuning effectively guides the model to tailor its representations for the given task while leveraging the extensive pre-existing knowledge encoded in the model. We kept 128 virtual tokens at the prefix of the prompt and 100 tokens to encode the tweet looking at the distribution of tweet lengths. We used learning rate of 1e-2 and batch size of 8 to fine-tune the model for 15 epochs. We used BCEWithLogitsLoss loss function to suit the multi label classification problem. To prevent overfitting, We used early stop monitoring the validation loss with patience value 5. For this experiment, we selected probability threshold of 0.5 to assign classes above this threshold to any tweet.

6.2. Experimental Setup

Our experimental framework was designed to ensure robust model development and evaluation. We started by randomly shuffling the dataset and then splitting it into an 80% training set and a 20% validation set. We pre-processed the training and validation set using the pre-processing steps mentioned in Section 5. Given the nature of tweets with multiple labels, we applied a Multilabel Binarizer to appropriately encode and handle these labels. Additionally, to prevent overfitting, we employed early stopping techniques with configurable parameters. For each experiment, we systematically varied model hyperparameters. Detailed information on these parameters and experiment configurations can be found in the Section 6.1.

6.3. Predictions

For the predictions over the final test data provided, we fine-tuned different language based model architectures with the objective of multi label text classification, details of which are mentioned in Section 6.1. We predicted the probability scores of each test tweet against all classes. We also experimented with different probability thresholds to assign classes for different models and selected thresholds based on Macro-F1 performance metric. Classes with probability score greater than the selected threshold were assigned as the predicted classes for that tweet. We also did some post-processing for scenarios where the model was predicting other class labels along with “none” class label, so we removed “none” class label in those scenarios and kept the other predicted class labels as is. Based on our thresholds, there might be a few scenarios, where the model didn’t make any prediction to ensure precise results. We submitted 3 prediction files from different models containing Tweet ID and predicted classes.

6.4. Additional Modeling Experiments

In addition to the submitted models, we conducted a series of experiments utilizing diverse feature sets and model architectures. However, these experiments did not yield superior results and were consequently not included in the final submission. This section provides insights into our exploration of alternative approaches, offering valuable context for the chosen model's selection.

BERTweet Large: As the cornerstone of our approach, we selected the BERTweet Large model due to its specialization in processing Twitter data. This model is pre-trained on a massive Twitter corpus, making it adept at capturing the linguistic nuances and contextual intricacies of tweets. BERTweet [10] is the first public large-scale language model pre-trained for English Tweets. BERTweet is trained based on the RoBERTa pre-training procedure. The corpus used to pre-train BERTweet consists of 850M English Tweets (16B word tokens, 80GB), containing 845M Tweets streamed from 01/2012 to 08/2019 and 5M Tweets related to the COVID-19 pandemic. The BERTweet Large model was fine-tuned on our training dataset. During fine-tuning, we optimized model weights to align with the specific multi-label classification task. This step included adjusting model parameters, learning rates, and batch sizes. We used learning rate of 2e-5 and batch size of 10 to fine-tune the model for 10 epochs. We used early stopping threshold of 0.001 for preventing model overfitting. For this experiment, we selected probability threshold of 0.2 to assign classes above this threshold to any tweet.

tf-idf vectorizer with Deep Neural Network: After pre-processing the text, we used tf-idf vectorizer to create numerical representations of text features. Then, we used Deep Neural Network model on these features by adding dense layers and also drop out layers to handle overfitting.

LSTM with GloVe Twitter Embedding: We did another experimentation by building an LSTM model. We used GloVe Twitter(2B tweets, 27B tokens, 1.2M vocab, uncased, 100d) embedding [1] to create features. Then we used a dropout layer for handling overfitting, an LSTM layer and a Dense layer for building the multi-label classifier. We used sigmoid as the activation function at the output layer, binary cross-entropy as loss. With this experiment we got Macro Average F1 score of 0.296 on the validation set using threshold of 0.2.

Experiment with GPT-4: We experimented with GPT-4 [5] to generate labels for tweets in validation set by giving it few shots examples of all the class labels along with system and user prompt, details of which are mentioned in Appendix B. We used temperature of 0 to be more deterministic and top_p of 1.0. We analyzed the results to find that most of the times GPT-4 was predicting at least 2 labels for a tweet, even though our data distribution has majority of the times 1 label for each tweet. Hence, it was significantly lowering the precision of the results.

Expanded Tweet Experiment: As mentioned in the pre-processing section, we expanded the tweets for certain classes to improve the performance of those classes. We did prefix tuning of Roberta Large model using expanded tweets for certain classes and normal tweet for other classes in the train set. In the validation set, we didn't expand tweets to evaluate performance.

[1] https://nlp.stanford.edu/projects/glove/

We didn't see any performance improvement over the prefix tuning of Roberta Large model on normal tweets.

Label Enhancement and Similarity Matching: In this experiment we tried to enhance the label by using GPT-3.5 model. After having enhanced labels we calculated its embedding using BERTweet model. In runtime we calculated the cosine similarity of embeddings of enhanced labels and tweets. We noticed the with threshold as 0.8 it was not performing well.

7. Evaluation

This task was evaluated using Macro-F1 score on the 12 different classes as metric. The result of our submitted automated runs on test set for this Task is shown in Table 2.

Sr No.	Team_name	Model Details	macro-F1 score	Jacc score	Rank
1	Cognitive Coders	DeBERTa Large Fine-tuning	0.67	0.70	5
2	Cognitive Coders	RoBERTa Large Prefix tuning	0.64	0.65	9

Table 2
Results on submitted test set

8. Conclusion and Future Work

In the final evaluation of this study, we conducted fine-tuning experiments using different language models: DeBERTa Large and Prefix Tuning of RoBERTa Large. Our objective was to explore the performance of these models in the context of a complex dataset where none of the labels exhibited a direct correlation with entity, sentiment, length, or word characteristics.

Our findings revealed that transformer-based models outperformed traditional classifiers in handling the intricacies of our dataset. This observation underscores the potential of transformer-based architectures in addressing multifaceted classification tasks.

Furthermore, we explored different data augmentation strategies, such as utilizing language models (LLM) to expand tweet text and provide additional context with the objective that this approach can potentially enhance model performance, particularly for labels with limited data points, such as religious, country, and ingredients. Increasing the dataset size for these labels may lead to improved classification accuracy, as transformer-based models are known to benefit from larger datasets due to their data-hungry nature. Also, In our research, we employed Hugging Face's bertweet-tb2_wnut17-ner API to detect entities in tweet texts. This API, specialized for social media data, enhanced our Named Entity Recognition (NER) capabilities. It allowed us to categorize entities effectively in the context of Twitter's informal language, hashtags, and mentions. This integration could enable comprehensive analyses of label assignments, sentiment, and tweet length, shedding light on the intricate entity-label relationships within our dataset. However, due to time constraints, our exploration yielded limited outcomes, suggesting

the need for further investigation in the future.
In summary, our study highlights the promising performance of transformer-based models in tackling complex multi-label classification tasks. Additionally, we recommend future research efforts that focus on data augmentation and dataset expansion to further enhance model effectiveness, particularly in scenarios with limited labeled data.

References

[1] 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.

[2] 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, 2022. arXiv:2204.13746.

[3] J. Devlin, M.-W. Chang, K. Lee, K. Toutanova, Bert: Pre-training of deep bidirectional transformers for language understanding, 2019. arXiv:1810.04805.

[4] X. L. Li, P. Liang, Prefix-tuning: Optimizing continuous prompts for generation, 2021. arXiv:2101.00190.

[5] R. OpenAI, Gpt-4 technical report, arXiv (2023) 2303–08774.

[6] M. Müller, M. Salathé, P. E. Kummervold, Covid-twitter-bert: A natural language processing model to analyse covid-19 content on twitter, 2020. arXiv:2005.07503.

[7] S. Bithel, S. Verma, Vaccinebert: Bert for covid-19 vaccine tweet classification, in: Working Notes of FIRE-13th Forum for Information Retrieval Evaluation, FIRE-WN 2021, 2021, pp. 1199–1203.

[8] P. He, X. Liu, J. Gao, W. Chen, Deberta: Decoding-enhanced bert with disentangled attention, 2021. arXiv:2006.03654.

[9] Y. Liu, M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, O. Levy, M. Lewis, L. Zettlemoyer, V. Stoyanov, Roberta: A robustly optimized bert pretraining approach, 2019. arXiv:1907.11692.

[10] D. Q. Nguyen, T. Vu, A. T. Nguyen, BERTweet: A pre-trained language model for English Tweets, in: Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations, 2020, pp. 9–14.

A. Data Exploration and Observations

In this section, we delve into an extensive exploration of our data, unveiling key insights across various dimensions. Specifically, we scrutinize tweet length, dissect word frequency patterns within each label category, extract entities relevant to each label, investigate the correlations between label assignments and tweet length, and conduct a thorough analysis of original versus expanded tweets. The subsequent subsections provide a comprehensive account of these analyses and observations.

Figure 1.1: Tweet lengths analysis

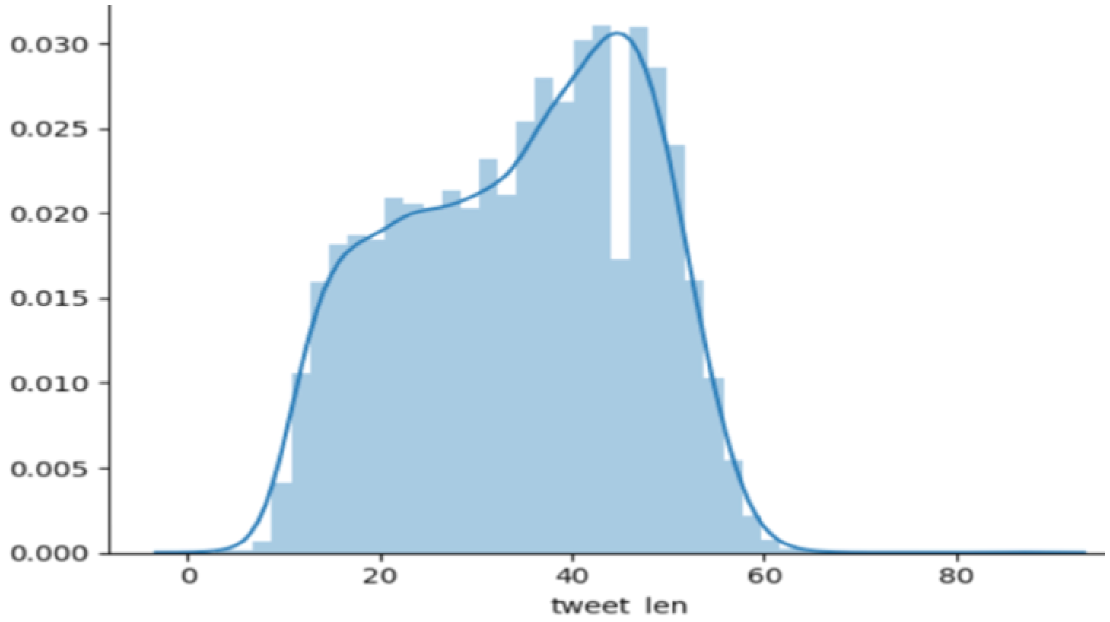
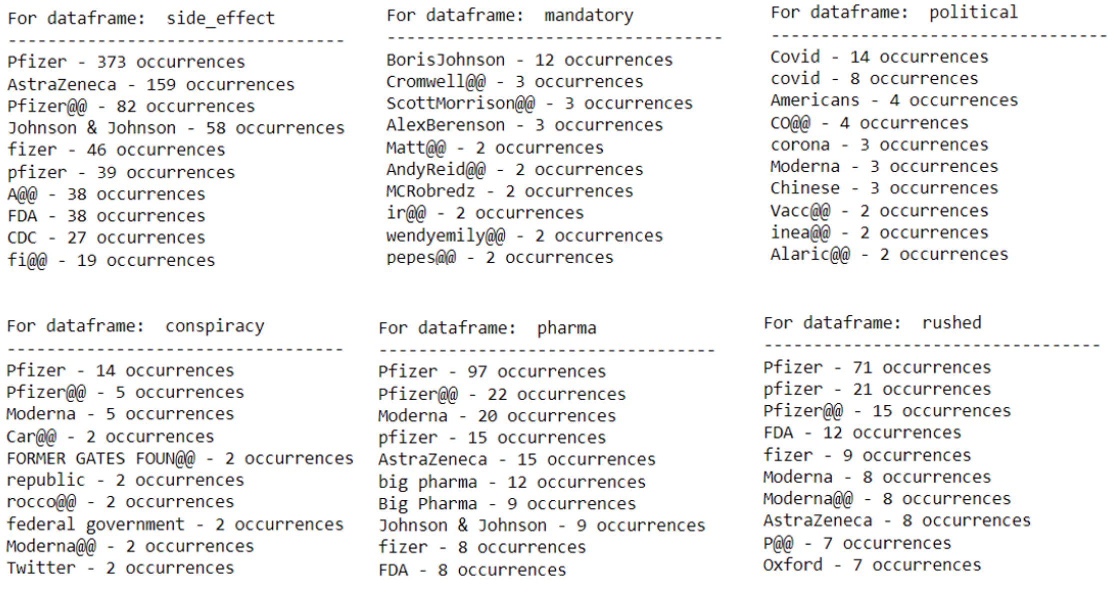


Figure 1.2: Word Frequency Analysis for label



```
For dataframe:  side_effect
----------------------------------
Pfizer - 373 occurrences
AstraZeneca - 159 occurrences
Pfizer@@ - 82 occurrences
Johnson & Johnson - 58 occurrences
fizer - 46 occurrences
pfizer - 39 occurrences
A@@ - 38 occurrences
FDA - 38 occurrences
CDC - 27 occurrences
fi@@ - 19 occurrences
```

```
For dataframe:  mandatory
----------------------------------
BorisJohnson - 12 occurrences
Cromwell@@ - 3 occurrences
ScottMorrison@@ - 3 occurrences
AlexBerenson - 3 occurrences
Matt@@ - 2 occurrences
AndyReid@@ - 2 occurrences
MCRobredz - 2 occurrences
ir@@ - 2 occurrences
wendyemily@@ - 2 occurrences
pepes@@ - 2 occurrences
```

```
For dataframe:  political
----------------------------------
Covid - 14 occurrences
covid - 8 occurrences
Americans - 4 occurrences
CO@@ - 4 occurrences
corona - 3 occurrences
Moderna - 3 occurrences
Chinese - 3 occurrences
Vacc@@ - 2 occurrences
inea@@ - 2 occurrences
Alaric@@ - 2 occurrences
```

```
For dataframe:  conspiracy
----------------------------------
Pfizer - 14 occurrences
Pfizer@@ - 5 occurrences
Moderna - 5 occurrences
Car@@ - 2 occurrences
FORMER GATES FOUN@@ - 2 occurrences
republic - 2 occurrences
rocco@@ - 2 occurrences
federal government - 2 occurrences
Moderna@@ - 2 occurrences
Twitter - 2 occurrences
```

```
For dataframe:  pharma
----------------------------------
Pfizer - 97 occurrences
Pfizer@@ - 22 occurrences
Moderna - 20 occurrences
pfizer - 15 occurrences
AstraZeneca - 15 occurrences
big pharma - 12 occurrences
Big Pharma - 9 occurrences
Johnson & Johnson - 9 occurrences
fizer - 8 occurrences
FDA - 8 occurrences
```

```
For dataframe:  rushed
----------------------------------
Pfizer - 71 occurrences
pfizer - 21 occurrences
Pfizer@@ - 15 occurrences
FDA - 12 occurrences
fizer - 9 occurrences
Moderna - 8 occurrences
Moderna@@ - 8 occurrences
AstraZeneca - 8 occurrences
P@@ - 7 occurrences
Oxford - 7 occurrences
```

Other labels ....

Table 1.1
Map Labels with Entity Types **(all labels)**

Labels	Top Entity Group
conspiracy	[ORG,PER]
conspiracy country	[PER,MISC]
conspiracy country ingredients	[ORG,MISC]
conspiracy country pharma	[MISC]
conspiracy country side-effect	[MISC,LOC]
conspiracy ineffective	[PER]
conspiracy ineffective ingredients	[ORG,LOC,MISC]
conspiracy ineffective side-effect	[PER]
conspiracy ingredients	[MISC,ORG]
conspiracy ingredients mandatory	[PER,MISC]
conspiracy ingredients pharma	[PER,ORG]
conspiracy ingredients religious	[MISC,PER]
...	...

Table 1.2
Map Labels with Entity Types **(unique labels)**

Labels	Top Entity Group
conspiracy	[ORG,PER]
country	[PER,MISC,ORG]
ineffective	[PER,MISC,ORG]
ingredients	[PER,ORG]
mandatory	[PER]
none	[]
pharma	[ORG,PER]
political	[MISC,PER]
religious	[MISC]
rushed	[ORG,LOC]
side-effect	[ORG,LOC,MISC]
unnecessary	[MISC]

Table 1.3
Label association with tweet length analysis

Label	Tweet Length
side-effect	34.258344
ineffective	37.519737
rushed	38.111036
pharma	36.656716
mandatory	35.975734
unnecessary	38.738227
none	25.817170
political	38.599042
conspiracy	37.882957
ingredients	35.243119
country	31.129353
religious	33.031250

Table 1.4
Tweet sentiment analysis

Label	Sentiment
side-effect	-0.265099
ineffective	-0.050121
rushed	-0.051898
pharma	-0.047567
mandatory	-0.098390
unnecessary	-0.113949
none	-0.057131
political	-0.125577
conspiracy	-0.152121
ingredients	-0.096524
country	-0.027262
religious	-0.180167

Table 1.5
Original and Expanded tweet text analysis

Original tweet	**Expanded tweet**	**Label**
Michael Yeadon, a former employee of Pfizer, said that the government rollout of the COVID-19 vaccine is an attempt at ”mass depopulation” with booster recipients expected to die ...	It is important to note that the claim made by Michael Yeadon, a former employee of Pfizer, that the government rollout of the COVID-19 vaccine is an attempt at ”mass depopulation” with booster recipients expected to die...	conspiracy
@MrStache9 Well i believe there won’t be an election until Trudick get enough covid vaccine into enough people to claim he did something right...	The statement seems to suggest that the Canadian government, led by Prime Minister Justin Trudeau, will likely wait until a significant portion of the population has been vaccinated against COVID-19...	political
Iâ€™d rather catch Covid than take that Russian vaccine!	It seems that the person who wrote this text is expressing their skepticism or distrust towards the Russian COVID-19 vaccine. They are saying that they would prefer to risk getting infected with COVID-19 than to take the Russian vaccine...	country

B. Experiments

GPT-4 Prompt

System Prompt:
You are a helpful assistant that will help in providing the most relevant labels to a social media post from a list of labels that express significant concern towards the vaccine.

User Prompt:
Assign most relevant labels to a social media post (particularly, a tweet) according to the specific concern(s) towards vaccines as expressed by the author of the post.
Note that a tweet can have more than one label (concern), e.g., a tweet expressing more than 1 different concerns towards vaccines will have more labels.
We consider the following concerns towards vaccines as the labels for this classification task:
{labels with description}
tweet text: {text}
Response: list of labels separated by space

Sample of Few-shot examples:

```
{
   "role": "system",
    "name": "example_user",
    "content": '''@kentlivenews Let's hope Boris Johnson isn't one of those new
                 trainees to stick people with the vaccine. Not a good picture to
                 use.'''
},
{
    "role": "system",
    "name": "example_assistant",
    "content":     'Political',
}
```

C. Online Resources

- GitHub