

Deciphering Vaccine Sentiments: Transformer Models in Social Media Analysis

S.Sushmitha^{1,*}, M.S.Shriram^{2,*} and S. Karthika^{3,*}

¹Department of Information Technology, Sri Sivasubramaniyanadar College of Engineering, Kalavakkam, Chennai

²Department of Information Technology, Sri Sivasubramaniyanadar College of Engineering, Kalavakkam, Chennai

³Department of Information Technology, Sri Sivasubramaniyanadar College of Engineering, Kalavakkam, Chennai

Abstract

The Forum for Information Retrieval Evaluation, a platform emphasizing on data evaluation and analysis, has conducted a novel challenge in the NLP domain titled Artificial Intelligence on Social Media. The main idea behind this task is to formulate an effective multi-label classifier that labels social media posts according to specific concerns towards vaccines expressed by authors of the posts. The critical point here is that each post has multiple unlikely characters and multiple labels, this would ideally require models with better attention mechanisms for accurate classifications. Upon appropriate pre-processing, our team performed three different transformer based approaches that provided optimal results compared to other models with lesser attention span. The models used were BERT, RoBERTa and XLNet. The BERT and RoBERTa models gave the same macro-f1 score of 0.57 but had a mild difference with the Jaccard score with RoBERTa giving 0.57 and BERT giving 0.56. The XLNet model gave out a 0.48 macro-f1 score and 0.49 Jaccard score. Our research primarily aims to throw light on why these state-of-the-art transformer approaches prove to be deal breakers in modern day NLP.

Keywords

NLP, Transformers, BERT, RoBERTa, XLNET


1. Introduction


In more recent times there is no greater illustration of how crucial vaccinations are as a vital weapon in fighting off diseases than how this has played out and the impact vaccines have had on controlling the COVID-19 pandemic. While there are numerous advantages to vaccination, there is still an extensive group of individuals who don't trust in them. Such skepticism has been fed by a number of issues including doubts over the safety and effectiveness of vaccines, conspiracy theories about pharmaceutical companies and governments having hidden agendas. These issues need to be understood and addressed in order to encourage mass vaccination and secure the public health. In this paper, we propose an exploratory method to analyze social media sentiment regarding vaccines in general. They aim at building an efficient multi-label classification system for the task of identifying and labeling different types of issues raised in tweets about vaccines. Our typology includes many issues such as concerns over vaccine

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

✉ sushmitha2010422@ssn.edu.in (S.Sushmitha); shriram2010160@ssn.edu.in (M.S.Shriram); skarthika@ssn.edu.in (S. Karthika)

ORCID 0009-0002-2396-266X (M.S.Shriram); 0000-0001-8919-5841 (S. Karthika)

 © 2023 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

 CEUR Workshop Proceedings (CEUR-WS.org)

safety and efficacy, concerns about government influence or manipulation, various forms of conspiracies, and more. To accommodate the possibility that one tweet could address more than one issue, we have built our classification framework in such a way that multiple labels can be assigned to each tweet. We utilize cutting edge NLP models namely BERT, RoBERTa, XLNET for the task, which can learn subtle linguistic features and context information with great effectiveness. These tools allow us to get a pretty accurate reading of what people are saying in tweet-form; it's quite useful for understanding feelings and topics people are discussing on social media. The goal of these studies is to advance knowledge about public attitudes and receptivity towards immunizations to better develop public health information campaign and communication strategies. With the power of social media data analytics and sophisticated advanced NLP models at our disposal, we are attempting to illuminate what may contribute to anti-vaccine attitudes so as to increase awareness and immunization rates across the globe

2. Related Work

This section provides an overview of research in developing architectures for comment classification in the context of the COVID-19 pandemic's impact on social media information flow. The urgency of reliable classification methods for the surge in COVID-19-related tweets has led to various approaches. One study [1] achieved remarkable F1 scores of 0.93 and 0.92 with two machine learning approaches for classifying tweets into three categories. Another work [2] focused on sentiment analysis of COVID-19-related Twitter posts from March to mid-April 2020, utilizing seven deep learning models with F1 scores consistently exceeding 0.90. In [3], advanced machine learning techniques resulted in 95.35% accuracy for sentiment analysis and 91.49% for topic classification. COVID-19 vaccine sentiment on Twitter was analyzed in [4], employing diverse data collection and preprocessing methods, along with classification using BERT Transformers. [5] introduced four deep learning models combining BERT with BiLSTM and BiGRU algorithms, excelling over classical machine learning models. [6] harnessed Bidirectional Long Short-Term Memory (Bi-LSTM) to analyze COVID-19 sentiments from Twitter and Reddit, surpassing conventional LSTM. [7] introduced a deep learning sentiment analysis model that achieved an accuracy of 78.062% on COVID-19-related tweets by focusing on sentence embeddings. [8] analyzed COVID-19 sentiment on Twitter, with BERT emerging as the most accurate model. [9] conducted sentiment analysis on Twitter discussions about COVID-19 vaccines, while [10] employed various sentiment analysis models, with BERT outperforming all. [11] conducted a comprehensive analysis of COVID-19 vaccine-related tweets, highlighting a shift from initial negativity to a more positive outlook. [12] investigated public perceptions of COVID-19 vaccine adverse effects through social media data, with LSTM achieving the highest accuracy. [13] analyzed #sideeffects social media conversations during early COVID-19 vaccine rollout, with BERT outperforming other methods. [14] examined over 200,000 COVID-19 vaccination tweets, achieving up to 90% accuracy in classifying misleading content. [15] focused on classifying COVID-19 vaccine-related tweets into negative, neutral, and positive sentiments, with Bidirectional LSTM (BiLSTM) standing out with a 94.12% accuracy rate. A variety of machine learning and deep learning models have been successful in capturing the subtleties of public opinion during the pandemic, as shown by the most current work in this field, which

shows the extensive efforts made to assess sentiments in tweets related to COVID-19.

3. Dataset

The training dataset[16] provided by AiSoMe FIRE 2023[17] contains 9921 records about anti-vaccine tweets about COVID vaccines that were posted during 2020-21. The dataset contains tweet id tweet and labels. The important pattern to look into is that each tweet can be categorized under multiple labels. This dataset encompasses 12 unique categories, each explicitly defined by distinct concerns and sentiments regarding vaccines. These categories include "Unnecessary," where tweets suggest vaccines are unnecessary or that alternate cures are superior; "Mandatory," which pertains to objections against mandatory vaccination; "Pharma," focusing on skepticism towards pharmaceutical companies and their profit motives; "Conspiracy," delving into deeper vaccine-related conspiracies; "Political," addressing concerns about government or political influences on vaccination; "Country," involving objections based on the country of origin of vaccines; "Rushed," which expresses concerns about rushed or inadequately tested vaccine processes; "Ingredients," centering around worries regarding vaccine components and technology; "Side-effect," focusing on concerns about vaccine side effects and associated deaths; "Ineffective," which questions vaccine effectiveness; "Religious," indicating objections to vaccines for religious reasons; and finally, "None," representing tweets with no specified reason or concerns outside of the defined categories. A test dataset[16] was also provided which contained 486 records which were used for prediction by the models.

4. Methodology

Upon receiving the training dataset for the AISoMe track as part of the FIRE 2023 conference we first analyzed a few inconsistencies that the textual data had. To overcome such inconsistencies and to use a clean form of data for the multi-label classifier model, we first used certain methods to preprocess the data. In the provided dataset, each "post" started off with the usernames being tagged. In social media tagging a user usually involves using the "@" symbol. So the first task involved removing the "@" symbol. Post this, we removed html tags, brackets and emojis and converted the characters to lower cases. This was followed by stop word removal and lemmatization. This preprocessing technique provided us with clean data that can be used for further processing. Since there are multiple labels for a single post, this task can be considered multi label classification. The processing of these labels involved splitting them up and encoding the labels, using multi-label binarizer technique to the appropriate numerical format. Upon understanding the attention mechanism of transformer based models, we proceeded to incorporate three such models namely BERT, RoBERTa and XLNET for the classification task. These models were chosen because of the self-attention mechanism that they possess which enables it to weigh the significance of different elements within a sequence simultaneously, this gives an edge over traditional models that relied on sequential data processing.

4.1. BERT

The first model devised for this task is the Bidirectional Encoder Representation from Transformers, shortly known as BERT. It is a state of the art model used in a variety of NLP tasks but in this particular case, we are formulating one for multi-label classification. Any text classification model requires in depth context extraction/learning. BERT has this intricate feature called MLM that helps it learn this concept more efficiently. MLM refers to a masked language model which means, certain tokens in a text are intentionally masked or hidden, and the model's task is to predict or fill in the missing tokens. BERT is bidirectional which means it takes into account both left and right contexts of the masked tokens in general. While using BERT for The input text is tokenized and fed into the BERT model, and the final hidden state corresponding to the [CLS] token (a special token added to the input) is used as the representation for the entire input sequence.classification, we append a classification layer upon the pre-trained BERT model.The output of the BERT model's [CLS] token representation is passed through this classification layer. The number of output units in the classification layer corresponds to the number of classes in the classification task, for this task it is 12 output units representing 12 unique classes.

4.2. RoBERTa

RoBERTa, an extension of the BERT model, excels in various NLP tasks, including multi-label classification. It utilizes a Masked Language Model (MLM) during pre-training, similar to BERT, where certain tokens in the input text are concealed, and the model's objective is to predict these hidden tokens, taking both left and right contexts into account. The key difference between BERT and RoBERTa here is in BERT architecture, the masking is performed once during data preprocessing, resulting in a single static mask. To avoid using the single static mask, training data is duplicated and masked 10 times, each time with a different mask strategy over 40 epochs thus having 4 epochs with the same mask. This strategy is compared with dynamic masking in which different masking is generated every time we pass data into the model.. In multi-label classification, RoBERTa adapts by adding a classification layer atop its pre-trained layers. Input text is tokenized, and the final hidden state associated with the [CLS] token is used as the sequence representation. The number of output units in the classification layer matches the number of classes in the classification task. During inference a sigmoid activation was used and a threshold is applied to determine the assigned labels.

4.3. XLNET

Much similar to BERT and RoBERTa, XLNET also has a bidirectional approach for concept learning but what sets XLNET apart is the autoregressive method that it employs. This helps XLNET to capture contextual dependencies .Much like RoBERTa, XLNet employs a Masked Language Model (MLM) during pre-training, akin to BERT, where specific tokens within the input text are concealed, and the model's primary objective is to predict these concealed tokens, factoring in both left and right contexts. The pivotal distinction between BERT and XLNet lies in the methodology of masking. While BERT employs a static mask, applied once during data preprocessing, XLNet opts for a more dynamic approach. During training, the data is duplicated and masked tenfold, with each instance employing a distinct masking strategy. This dynamic

masking strategy evolves with each pass through the model, enriching context comprehension. In the realm of multi-label classification, XLNet adapts seamlessly by introducing a classification layer atop its pre-trained layers. Text inputs are tokenized, and the ultimate hidden state corresponding to the [CLS] token functions as the sequence representation. The number of output units in our case is 12 mirroring the number of classes in the classification task. During inference, sigmoid activation function was used, with a threshold applied to establish the assigned labels similar to the other 2 models.

5. Evaluation and Results

The training phase of our models was a meticulous process that involved fine-tuning several critical hyperparameters, each with its unique role in shaping model performance. Weight decay, represented by λ and set at 0.001, acted as a vital regularizer. It controlled the influence of model weights on the loss function, effectively curbing overfitting by penalizing large weight values. The learning rate, denoted as η and thoughtfully configured at 0.0001 or 0.00001, played a pivotal role in determining the step size for weight updates during training. This parameter significantly impacted convergence, with an appropriate learning rate ensuring that the model effectively learned from the data without overshooting optimal parameter values. Dropout rates, which ranged between 0.4 and 0.5, were judiciously chosen to regulate neuron activations. Dropout is a regularization technique that prevents over-reliance on specific neurons by randomly deactivating a fraction during each training iteration, promoting generalization. Furthermore, the "epochs" parameter governed the number of complete iterations through the training data, allowing models to progressively learn and adapt over time. Finally, the "maximum sequence length" parameter was set at 300, ensuring that input sequences were appropriately processed during training, thereby enhancing the model's overall ability to achieve peak accuracy. These hyperparameters collectively contributed to the fine-tuning process, facilitating the models' optimization and ensuring their robust performance. Table 1 specifies the various hyperparameters and their corresponding values that gave maximum training accuracy

Table 1

Comparison of models based on different hyperparameters and their corresponding values

Models	weight decay	learning rate	dropout	epochs	max length	accuracy
BERT	0.001	0.0001	0.5	10	300	99.2
RoBERTa	0.001	0.00001	0.4	10	300	98.12
XLNET	0.001	0.0001	0.5	10	300	96.02

In the evaluation phase, a dedicated test dataset was provided to rigorously assess the performance of the trained models. Each model underwent a single run, generating predictions that were then submitted for evaluation. The evaluation process hinged on two pivotal metrics: the Macro F1 Score and Jaccard Score. Together, these two reliable assessment criteria provided reliable benchmarks for carefully assessing and contrasting the performance of the models

The Macro F1-Score is a metric used to evaluate the performance of a classification model. It is calculated by taking the harmonic mean of precision and recall. The Macro F1-Score is useful in multiclass classification problems where the classes are imbalanced. It gives equal weight to each class, making it a good metric to use when the data is imbalanced.

$$F1Score = 2 * (Precision * Recall) / (Precision + Recall) \quad (1)$$

$$Macro\ F1\ Score = \frac{\sum F1\ scores}{Number\ of\ classes} \quad (2)$$

The Jaccard Index, also known as the Jaccard similarity coefficient, is a statistic used in understanding the similarities between sample sets. The Jaccard Score is particularly valuable for evaluating models dealing with binary or multiclass classification, as it doesn't take into account true negatives, focusing solely on the overlap between predicted and actual positive instances.

$$Jaccard(U, V) = \frac{|U \cap V|}{|U \cup V|} \quad (3)$$

Table 2
Run Results

Models	Macro F1	Jaccard
RoBERTa	0.57	0.57
BERT	0.57	0.56
XLNET	0.48	0.49

Table 2 specifies the final Macro-F1 and Jaccard scores of the runs for different models submitted.

6. Conclusion

This study harnessed advanced transformer-based models, BERT, RoBERTa, and XLNET, to address the complex task of multi-label classification in the context of tweets expressing concerns about vaccines, primarily comprising anti-vaccine sentiments. Notably, RoBERTa demonstrated superior performance, followed closely by BERT and XLNET. In this predominantly anti-vaccine dataset, these models showcased their prowess in discerning and categorizing various concerns. To further enhance our model's performance, future investigations could explore data augmentation techniques, given the data-hungry nature of transformer-based models. Additionally, fortifying model robustness through methods like adversarial training presents a promising avenue. These efforts align with our mission of accurately categorizing concerns within the domain of vaccine-related discussions.

References

- [1] Anupam Mondal , Sainik Kumar Mahata , Monalisa Dey , Dipankar Das, "Classification of COVID19 tweets using Machine Learning Approaches",SMM4H 1 June 2021.
- [2] Vernikou, S., Lyras, A. & Kanavos, A. Multiclass sentiment analysis on COVID-19-related tweets using deep learning models. *Neural Comput & Applic* 34, 19615–19627 (2022). <https://doi.org/10.1007/s00521-022-07650-2>
- [3] Gurkan, Caglar , Kozalioglu, Sude , Akdag, Ismail , Göçen, Cem , Palandoken, Merih. (2022). COVID-19 Related Tweets Classification and Sentiment Analysis Based on Machine Learning Approaches and Deep Learning Architecture Designs: A Comprehensive Analysis. 10.13140/RG.2.2.18200.47365/1.
- [4] K. K. Agustini, E. Utami and H. Al Fatta, "Sentiment Analysis of COVID-19 Vaccine on Twitter Social Media: Systematic Literature Review," 2021 IEEE 5th International Conference on Information Technology, Information Systems and Electrical Engineering (ICITISEE), Purwokerto, Indonesia, 2021, pp. 121-126, doi: 10.1109/ICITISEE53823.2021.9655960.
- [5] Talaat, A.S. Sentiment analysis classification system using hybrid BERT models. *J Big Data* 10, 110 (2023). <https://doi.org/10.1186/s40537-023-00781-w>
- [6] Arbane M, Benlamri R, Brik Y, Alahmar AD. Social media-based COVID-19 sentiment classification model using Bi-LSTM. *Expert Syst Appl*. 2023 Feb;212:118710. doi: 10.1016/j.eswa.2022.118710. Epub 2022 Aug 30. PMID: 36060151; PMCID: PMC9425711.
- [7] Fattoh IE, Kamal Alsheref F, Ead WM, Youssef AM. Semantic Sentiment Classification for COVID-19 Tweets Using Universal Sentence Encoder. *Comput Intell Neurosci*. 2022 Oct 5;2022:6354543. doi: 10.1155/2022/6354543. PMID: 36248924; PMCID: PMC9556213.
- [8] U. N. Wisesty, R. Rismala, W. Mungguna and A. Purwarianti, "Comparative Study of Covid-19 Tweets Sentiment Classification Methods," 2021 9th International Conference on Information and Communication Technology (ICoICT), Yogyakarta, Indonesia, 2021, pp. 588-593, doi: 10.1109/ICoICT52021.2021.9527533.
- [9] J. Philip, V. N. Thatha, M. Harshini, I. V. S. L. Haritha, S. Patil and B. Veerasekhar Reddy, "Classification of Covid-19 Vaccines tweets using Naïve Bayes Classification," 2022 6th International Conference on Electronics, Communication and Aerospace Technology, Coimbatore, India, 2022, pp. 1384-1387, doi: 10.1109/ICECA55336.2022.10009511.
- [10] V. Battula, S. G. Goli and J. Nasigari, "Identification of Optimal Model for Multi-Class Classification of COVID Tweets," 2022 9th International Conference on Computing for Sustainable Global Development (INDIACom), New Delhi, India, 2022, pp. 495-499, doi: 10.23919/INDIACom54597.2022.9763291.
- [11] S. Gottipati and D. Guha, "Analysing Tweets on COVID-19 Vaccine: A Text Mining Approach," 2022 IEEE 12th Annual Computing and Communication Workshop and Conference (CCWC), Las Vegas, NV, USA, 2022, pp. 0467-0474, doi: 10.1109/CCWC54503.2022.9720793.
- [12] K. R. S. N. Kariyapperuma, K. Banujan, P. M. A. K. Wijeratna and B. T. G. S. Kumara, "Classification of Covid19 Vaccine-Related Tweets Using Deep Learning," 2022 International Conference on Data Analytics for Business and Industry (ICDABI), Sakhir, Bahrain, 2022, pp. 1-5, doi: 10.1109/ICDABI56818.2022.10041615.
- [13] A. R. Manjrekar, D. J. McConnell and S. S. Gokhale, "Analyzing Twitter Conversations

- on Side Effects of Covid-19 Vaccine,” 2022 2nd International Conference on Intelligent Technologies (CONIT), Hubli, India, 2022, pp. 1-8, doi: 10.1109/CONIT55038.2022.9848134.
- [14] S. Sharma, R. Sharma and A. Datta, ”(Mis)leading the COVID-19 Vaccination Discourse on Twitter: An Exploratory Study of Infodemic Around the Pandemic,” in IEEE Transactions on Computational Social Systems, doi: 10.1109/TCSS.2022.3225216.
- [15] N. Mansouri, M. Soui, I. Alhassan and M. Abed, ”TextBlob and BiLSTM for Sentiment analysis toward COVID-19 vaccines,” 2022 7th International Conference on Data Science and Machine Learning Applications (CDMA), Riyadh, Saudi Arabia, 2022, pp. 73-78, doi: 10.1109/CDMA54072.2022.00017.
- [16] Soham Poddar, Azlaan Mustafa Samad, Rajdeep Mukherjee, Niloy Ganguly, and Saptarshi Ghosh. 2022. 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 (SIGIR '22). Association for Computing Machinery, New York, NY, USA, 3154–3164. <https://doi.org/10.1145/3477495.3531745>
- [17] Soham Poddar , Moumita Basu, Kripabandhu, and Saptarshi 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.