Multi-Class Classification using BERT models for Detecting Depression Signs from Social Media Text

# Introduction

The sentiment expressed by the public on social media platforms is gaining significance in controlling the spread of pandemics. In this regard, it is crucial to comprehend the emotions and perceptions of the public regarding COVID-19 vaccinations. The research article under consideration explores the sentiment of Japanese Twitter users towards tweets associated with COVID-19 vaccination. The study utilized the WRIME dataset, which provides emotion ratings for Japanese tweets, and fine-tuned a BERT model to predict the emotional intensity levels. The model achieved an impressive training accuracy of MSE = 0.356.

# Related Work

In the past, the analysis of emotions in Tweets relied on a combination of feature engineering, lexicon-based methods, and traditional off-the-shelf classifiers before deep learning was adopted. (CBalabantaray et al., 2012) and (Wang et al., 2012) both created sets of features from Tweets, with features including n-grams, POS, adjectives, and lexicon-based sentiment polarity scores. (CBalabantaray et al., 2012) used an SVM to process the features, while (Wang et al., 2012) employed linear and Naive Bayes classifiers. (Roberts et al., n.d.) followed a similar approach, using an SVM and a set of features that included synonym rings, hypernyms, and LDA topic scores instead of sentiment polarity scores. However, these classical methods have been superseded by more advanced and specialized sequence modelling techniques such as RNNs and LSTMs. (Vateekul & Koomsubha, 2016) have demonstrated the superiority of LSTMs in emotion analysis over SVMs and Naive Bayes on Thai Twitter text, while (Colneric & Demsar, 2020) have demonstrated the effectiveness of character-based RNNs.

# Experiments

## Dataset

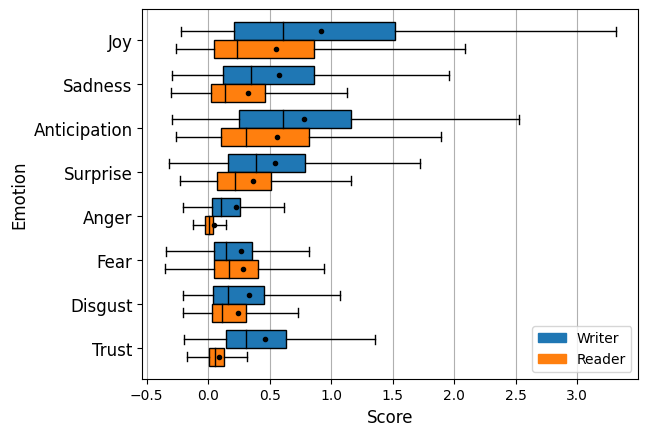
The research article being discussed aimed to analyze the emotions of Japanese Twitter users towards COVID-19 vaccines and related topics. To accomplish this, the researchers gathered a dataset of 20,254 vaccine-related Tweets from December 2021. The dataset included Tweets containing any of the following keywords: “ワクチン” (“vaccine”), “モデルナ” (“Moderna”), “ファイザー” (“Pfizer”), or “オミクロン” (“Omicron”). The keywords "Moderna" and "Pfizer" were specifically chosen as they represent the brands of COVID-19 vaccines that are commonly administered in Japan. The dataset was constructed by sampling 15 random minutes from each day of December 2021 for each keyword and scraping all Tweets containing the assigned keyword for each sampled minute.

## Implementation

The implementation of the paper # was performed using the GitHub repo provided by the author of the paper at this link: <https://github.com/PatrickJohnRamos/BERT-Japan-vaccination>

## Results

The results of the fine-tuned BERT model were evaluated on the samples extracted from the train set. The test set of the selected dataset is not publicly available now. So, we extracted the 2000 samples from the train set to evaluate the performance of the model. Now the scores of the labels on the extracted samples are presented in the below figure.



The comparison of the reproduced results in also shown in the below Table. Although the manuscript calculated the MSE for reader and writer emotions separately and collectively also, but using the code repository we are able to extract the MSE collectively.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Emotion | Reader | Writer | Reader + Writer | Reader + writer (Reproduced) |
| Joy | 0.658 | 0.192 | 0.425 | 0.125 |
| Sadness | 0.688 | 0.178 | 0.433 | 0.129 |
| Anticipation | 0.746 | 0.211 | 0.479 | 0.174 |
| Surprise | 0.542 | 0.139 | 0.341 | 0.189 |
| Anger | 0.486 | 0.032 | 0.259 | 0.109 |
| Fear | 0.462 | 0.147 | 0.304 | 0.062 |
| Disgust | 0.664 | 0.123 | 0.394 | 0.101 |
| Trust | 0.40 | 0.029 | 0.214 | 0.756 |
| Overall | 0.581 | 0.131 | 0.356 | 0.205 |

## Discussion

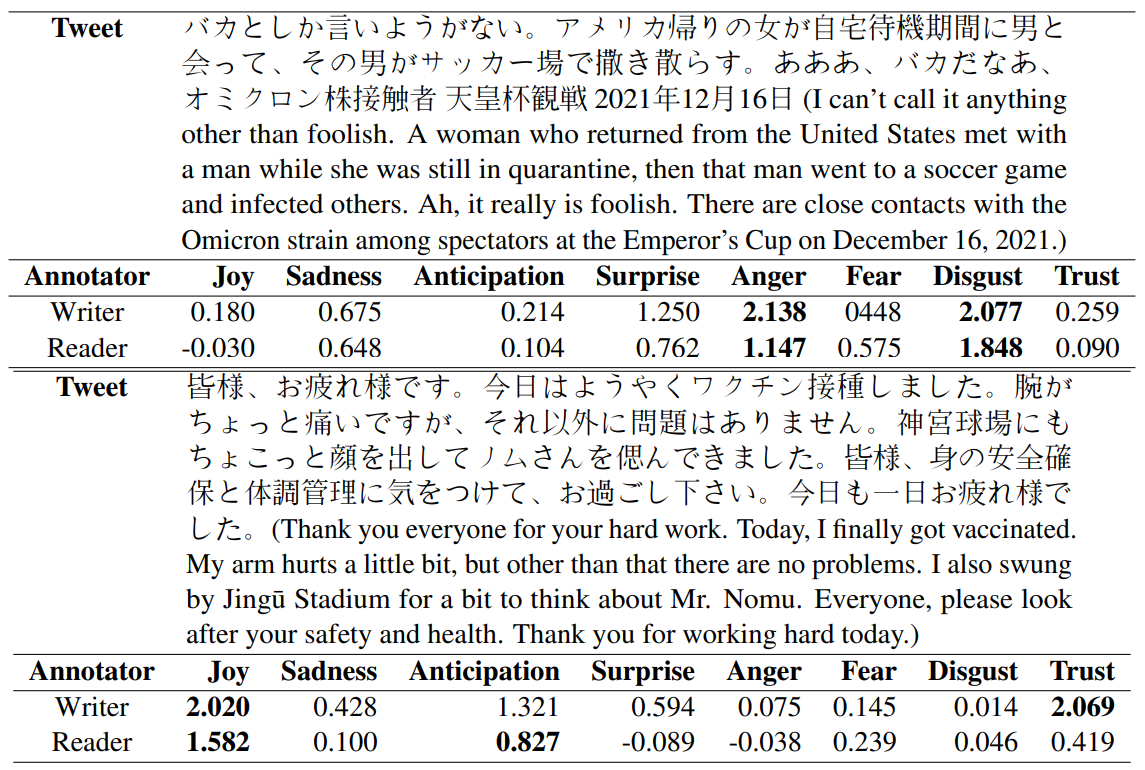
The code of the selected paper is available via the GitHub repository that was used here for the reproduction of the results. The BERT classifier showed a 0.2056 MSE after hyperparameter tuning. But these results are on the part of the training set as the test set is not publicly available. Although, they were generated not similarly due to the different datasets.

## Error Analysis

The manuscript analysis the single tweet to extract the emotions as shown in the below Figure. But the paper gives the example of a tweet that classifies correctly. For the error analysis, the paper didn’t provide any experiments and analyses. We are also unable to analyze errors with the provided code.

# Conclusion

Based on the analysis of the paper's methodology, it can be concluded that the results are reproducible through the model parameter tuning and training process. The code used in the study is available on GitHub, and the results can be replicated by implementing the code from scratch or using the provided repository. However, one limitation is the unavailability of the test set, which makes reproducing the same results on the test set difficult. Despite this limitation, the overall methodology of the paper is feasible, and the results are replicable.



# References

CBalabantaray, R., Mohammad, M., & Sharma, N. (2012). Multi-Class Twitter Emotion Classification: A New Approach. *International Journal of Applied Information Systems*, *4*(1), 48–53. https://doi.org/10.5120/IJAIS12-450651

Colneric, N., & Demsar, J. (2020). Emotion Recognition on Twitter: Comparative Study and Training a Unison Model. *IEEE Transactions on Affective Computing*, *11*(3), 433–446. https://doi.org/10.1109/TAFFC.2018.2807817

Roberts, K., Roach, M. A., Johnson, J., Guthrie, J., & Harabagiu, S. M. (n.d.). *EmpaTweet: Annotating and Detecting Emotions on Twitter*.

Vateekul, P., & Koomsubha, T. (2016). A study of sentiment analysis using deep learning techniques on Thai Twitter data. *2016 13th International Joint Conference on Computer Science and Software Engineering, JCSSE 2016*. https://doi.org/10.1109/JCSSE.2016.7748849

Wang, W., Chen, L., Thirunarayan, K., & Sheth, A. P. (2012). Harnessing twitter “big data” for automatic emotion identification. *Proceedings - 2012 ASE/IEEE International Conference on Privacy, Security, Risk and Trust and 2012 ASE/IEEE International Conference on Social Computing, SocialCom/PASSAT 2012*, 587–592. https://doi.org/10.1109/SOCIALCOM-PASSAT.2012.119