# HCMUS MediaEval 2021: Multi-Model Decision Method Applied on Data Augmentation for COVID-19 Conspiracy Theories Classification

Tuan-An To*[1,3], Nham-Tan Nguyen*[1,3], Dinh-Khoi Vo*[1,3], Nhat-Quynh Le-Pham*[1,3], Hai-Dang Nguyen[1,2,3], Minh-Triet Tran[1,2,3]
[1]University of Science, VNU-HCM, [2]John von Neumann Institute, VNU-HCM
[3]Vietnam National University, Ho Chi Minh city, Vietnam
ttan20@apcs.fitus.edu.vn,nntan20@apcs.fitus.edu.vn,vdkhoi20@clc.fitus.edu.vn
lpnquynh20@apcs.fitus.edu.vn,nhdang@selab.hcmus.edu.vn,tmtriet@fit.hcmus.edu.vn

## ABSTRACT

Corona Virus and Conspiracies Multimedia Analysis Task is the task in MediaEval 2021 Challenge that concentrates on conspiracy theories that assume some kind of nefarious actions related to COVID-19. Our HCMUS team performs different approaches based on multiple pretrained models and many techniques to deal with 2 subtasks. Based on our experiments, we submit 5 runs for subtask 1 and 1 run for subtask 2. Run 1 and 2 both introduces BERT[5] pretrained model but the difference between them is that we add a sentimental analysis to extract semantic feature before training in the first run. In run 3 and 4, we propose a naive bayes classifier[4] and a LSTM[8] model to diversify our methods. Run 5 ultilize an ensemble of machine learning and deep learning models - multimodal approach for text-based analysis[3]. Finally, in the only run in subtask 2, we conduct a simple naive bayes algorithm to classify those theories. In the final result, our method achieves 0.5987 in task 1, 0.3136 in task 2.

## 1 INTRODUCTION

The COVID-19 pandemic has severely affected people worldwide, and consequently, it has dominated world news for months. Thus, it has been the topic of a massive amount of misinformation, which was most likely amplified by the fact that many details about the virus were unknown at the start of the pandemic. In the Multimedia Evaluation Challenge 2021 (MediaEval2021), the purpose of Corona Virus and Conspiracies Multimedia Analysis Task is to develop methods capable of detecting such misinformation. By this way, this task aid in preventing misinformation outspread causing social anxiety and vaccination doubts. We propose different methods which are mainly based on deep learning model to solve the problem in various aspects which would be described in the later sections.

## 2 DATASET AND RELATED WORK

### 2.1 Dataset

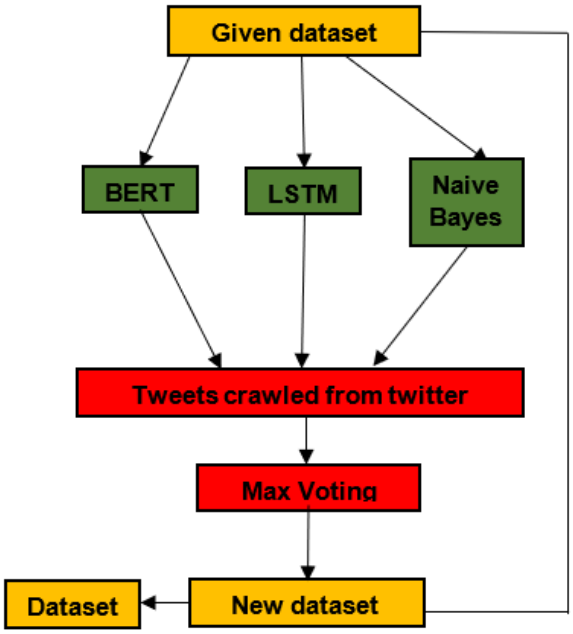
In subtask 1, we have recieved two datasets in total:

* dev1-task1.csv: unbalance dataset consisting of 500 tweets in which (Non-Discuss-Promote) respectively is (340,76,84)
* dev-task1.csv: unbalance dataset consisting of 1011 tweets in which (Non-Discuss-Promote) respectively is (414,186,411)


*MediaEval'21, 13-15 December 2021, Online*

### 2.2 Data Centric Approach

Text-Based Misinformation Detection exists similar objectives to the text classification task. Hence, we take advantage of pretrained NLP models and fine-tune them for this task. However, the validation result is biased towards non-conspiracy class since given dataset is small and unbalance. Therefore, we adapt those models to generate new data by crawling data from Twitter and assigning a label for a tweet if it gets the most voting which increase the effectiveness and balance on the dataset as well.




## 3 METHOD

### 3.1 Data Processing

From the pure train data, we need to preprocess the tweets to make it easier for our model to learn. The first step is replace those words in short form into its original form - I'm to I am. The second step is to remove the stopwords - about, above, ...; lemmatize the family words - roofing, roofers,... into roof. Finally, we also try to remove any other meaningless feature in the tweets such as the "https", "(amp)" and the emoji to get a perfect tweet for training.

### 3.2 Run 01 - Subtask1

Firstly, we use sentiment analysis method to categorize all tweets into two classes - optimism and anger. Based on the observation, non conspiracy tweets contain a higher rate in optimism while discuss/promote tweets dominate the anger rate. Therefore, we decide to pick out the tweets with the opt rate greater than 0.8 and anger rate less than 0.2 in the test set and directly label them as non conspiracy. The remained tweets are predicted by BERT[1] - a pre-trained of deep bidirectional transformer for epochs = 20, batch-size = 4 with adam optimization.

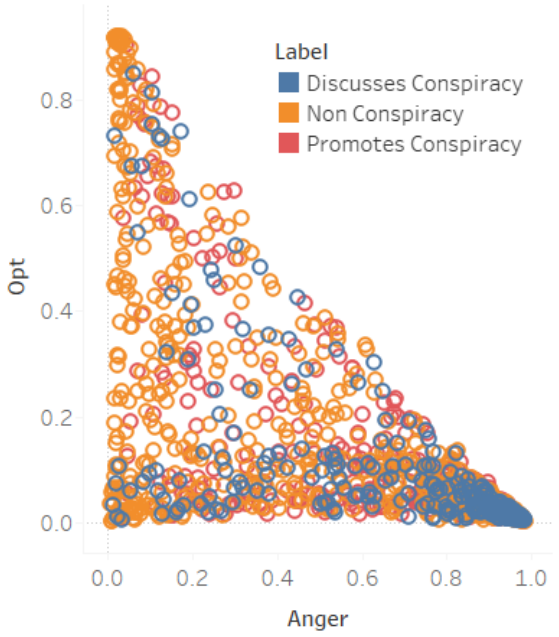



**Figure 1: Tweet sentiment analysis**

### 3.3 Run 02 - Subtask1

Different from Run 01, we try keeping the stopwords and just cleaning the tweets as well as replacing all the shorten terms into full written terms in order to remain the original structure of the sentence for the best performance of Transformer model[9]. The training process is still conducted on pre-trained BERT model with augmented dataset (batch_size=16, epochs=10).

### 3.4 Run 03 - Subtask1

We use tf-idf vectorizer to extract feature from the text. After trying both Logistic Regression and Naive Bayes[4], the latter algorithm perform better. The result of this run is our baseline score.

### 3.5 Run 04 - Subtask1

We use pretrained glove to transform each word in the sentence into an array of 300 numbers represent the "meaning" of the word. Finally we built a 2D LSTM[8] with adam optimization - batch_size = 64 , epoch = 4.

### 3.6 Run 05 - Subtask1

We combine all the results run by those deep-learning and machine-learning algorithm and label the tweet by its highest-voted class[3]. There are rare cases that the tweet get equal votes in different classes, we decide to label it the result given by BERT model.

### 3.7 Run 01 - Subtask2

Similar to the method in run 3, we used tf-idf vectorizer to extract feature from the text. Base on our observation, test set is extremely unbalanced regrading to multilabel problem, so we try to resolve it by downsampling the data - keeping only the dominant sentences in the biased class. In order to handle multilabel problem, we utilize three different methods: Binary Relevance[7], Classifier Chain[6] and Label Powerset[2] combined with Naive-Bayes and Logistic Regression. According to our experiment, Binary Relevance with Logistic Regression gives the best result.

## 4 EXPERIMENTS AND RESULTS

Table 1 shows the results of our runs in term of matthews correlation coefficient score.

| Team-run | Task-1 | Task-2 |
|---|---|---|
| SelabHCMUSJunior BERT/run1 | 0.5581 | - |
| SelabHCMUSJunior BERT/run2 | 0.5106 | - |
| SelabHCMUSJunior Naive-Bayes/run3 | 0.4469 | 0.3136 |
| SelabHCMUSJunior LSTM/run4 | 0.2570 | - |
| SelabHCMUSJunior Multi-model/run5 | **0.5987** | - |

**Table 1: HCMUS Team Submission results for Corona Virus and Conspiracies Multimedia Analysis Task**

## 5 CONCLUSION AND FUTURE WORKS

In summary, we identify challenges of the dataset and propose different approaches to address the issues. We conclude that classifying a tweet promotes/supports or discusses sentiment task is heavily biased towards the writers attitude, therefore making it difficult for NLP model to learn the true label. In recent study, we can only extract basic state of sentiment of a tweet such as sad or optimism, so we aim to tackle the challenge in a higher level in the future.

## ACKNOWLEDGMENTS

This research was funded by SeLab-HCMUS and VNUHCM-University Of Science.

## REFERENCES

[1] Stefano Bocconi Andrey Malakhov, Alessandro Patruno. 2020. Fake News Classification with BERT.
[2] Preeti GuptaEmail authorTarun K. SharmaDeepti Mehrotra. 2018. Label Powerset Based Multi-label Classification for Mobile Applications. In *Soft Computing: Theories and Applications*.

[3] Mihir P Mehta Chahat Raj. 2020. MediaEval 2020: An Ensemble-based Multimodal Approach for Coronavirus and 5G Conspiracy Tweet Detection.
[4] Di Li Haiyi Zhang. 2007. Naïve Bayes Text Classifier. arXiv:2007 IEEE International Conference on Granular Computing (GRC 2007)
[5] Kenton Lee Kristina Toutanova Jacob Devlin, Ming-Wei Chang. 2019. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding.
[6] Geoff Holmes Eibe Frank Jesse Read, Bernhard Pfahringer. Classifier Chains for Multi-label Classification.
[7] Xu-Ying Liu Xin Geng Min-Ling Zhang, Yu-Kun Li. 2018. Binary relevance for multi-label learning: an overview. In *Frontiers of Computer Science*.
[8] Rouhollah Rahmani Seyed Mahdi Rezaeinia, Ali Ghodsi. 2017. Improving the Accuracy of Pre-trained Word Embeddings for Sentiment Analysis.
[9] Victor Sanh Julien Chaumond Clement Delangue Anthony Moi Pierric Cistac Tim Rault Rémi Louf Morgan Funtowicz Joe Davison Sam Shleifer Patrick von Platen Clara Ma Yacine Jernite Julien Plu Canwen Xu Teven Le Scao Sylvain Gugger Mariama Drame Quentin Lhoest Alexander M. Rush Thomas Wolf, Lysandre Debut. 2020. HuggingFace's Transformers: State-of-the-art Natural Language Processing. (2020).