

# Project Proposal

- **Title:** Time series analysis of COVID-19 related emotion detection
- **Team members:** Chia-Wei Lin
- **Collaborators:** None
- **Project Sharing:** None

## Summary of the selected research paper

- **Title:** Exploratory Analysis of COVID-19 Related Tweets in North America to Inform Public Health Institutes
- **Authors:** Hyeju Jang, Emily Rempel, Giuseppe Carenini, Naveed Janjua
- **Conference:** ACL 2020 Workshop on Natural Language Processing for COVID-19
- **Year:** Jul 2020
- **URL:** [https://openreview.net/forum?id=7blr\\_OIAxvC](https://openreview.net/forum?id=7blr_OIAxvC)

- **What does the paper do and why is it new or noteworthy?**

In this paper, authors presented the exploratory results of topic modeling and aspect-based sentiment analysis(ABSA) on COVID-19 related tweets in North America. They compare topic modeling results of Canada and the US, and show how public health intervention related topic changes over time.

This paper shows how Natural Language Processing techniques could detect meaningful information on public opinion, which could be applied to public health questions and provide practical strategies for supporting the public and navigating the pandemic.

- **Why did you choose this paper?**

The COVID-19 outbreak has impacted our physical and mental health for almost a year. An exploratory social-media related sentiment study with time series analysis like this paper could help governments, organization and researchers to better address public health issues for current and future pandemics.

- **If the paper uses a particular dataset (or datasets) clearly state what they are.**

The paper uses a public twitter dataset about the COVID19 pandemic, collected by [Chen et al. \(2020\)](#), selecting only North America located English tweets from January to May, resulting in 319,524 tweets in total.

- **If the paper uses a particular evaluation metric (or metrics) state it clearly. Also state any key scores that it achieves on these metrics.**

The paper does not mention any evaluation metrics.

- **If the paper has any important models or techniques, describe them.**

The ABSApp, a weakly-supervised ABSA system is used to train the unlabeled data.

- **If you like, you can reference other papers.**

- a) [Oyebode, O. et al. \(2020\). "COVID-19 Pandemic: Identifying Key Issues Using Social Media And Natural Language Processing." \[online\] Arxiv.org. Available at: <<https://arxiv.org/pdf/2008.10022>>](#) offers detailed information about data preprocessing techniques.

## Project description

- **What are the main goals of your project?**

Analyze COVID-19 related tweets using Natural Language Processing techniques, including capture public perception about the pandemic from twitter, categorize opinions into themes with sentiment polarity, determine the emotional intensity scores, investigate the change in emotion and the prevalence of the topics over time, and explore the relationship between public emotional reaction with the number of COVID cases.

- **What NLP task(s) do you plan to address?**

Information retrieval, sentiment/emotion analysis, and topic modelling

- **What data do you plan to use?**

[Covid-19 Twitter chatter dataset](#) or [Coronavirus \(COVID-19\) geo-tagged tweets dataset](#)  
[Google Research's Open COVID-19 Data project](#)

- **What methods do you plan to use?**

Latent Dirichlet Allocation (LDA) technique for topic modelling.

VADER, a lexicon-based tool for sentiment analysis

- [A paper](#) mentioned above has given preprocessing steps for data cleaning and preparation.
- [One paper](#) and [the other paper](#) uses deep LSTM for sentiment classification, which could be another way to approach the problem.
- [Another paper](#) developed a BERT-based model called CT-BERT specific for COVID-19

- **What baseline will you use? How will you evaluate your results?**

I am not sure about which would be a good baseline and how to evaluate the result on current project design. However, I can either compare my result with previous research papers (For example, in [this paper](#) the authors compare the AUC and F1 score between their LSTM model with a SVM model which previously used by others) or test the model dataset with ground truth emotions like Real World Worry Dataset (RWWD) introduced in [this paper](#).