



TAN YI HONG

RDS

017-4255258

yhtan-pm18@student.tarc.edu.my

Dr. Lim Siew Mooi

## PROJECT TITLE: SENTIMENT ANALYSIS OF COVID-19 VACCINATION IN THE GLOBAL WITH NAÏVE BAYES APPROACH

### ABSTRACT

The COVID-19 pandemic has emerged as one of the world's most serious threats, and it is still very much a concern. Around the same time, we are in the middle of the world's largest vaccine program to fight against the deadly virus. While the vaccine has given the battle against COVID-19 a new lease of life, it has also sparked a wave of anti-vaccine protests. It would be useful to use sentiment analysis on recent Twitter data by using Twitter API or Twint to crawl Tweets data about vaccination on Twitter to gauge public opinion on the COVID-19 vaccine. This research project will carry out sentiment analysis using the Naïve Bayes approach and know the opinions of the people around the world towards their perception on the COVID-19 vaccine and determine the percentage of their responses that are positive, neutral, or negative emotions regarding the vaccine for the final classification.

### PROBLEM

With the ongoing vaccination campaign around the world, government and scientist need to prove and show the success of the development of vaccine to strengthen the confidence of all people in order to fight against COVID-19. The debate about vaccine development, accessibility, effectiveness, and side effects is ongoing, and it pervades news reports and Twitter feeds on a daily basis. Our online visibility, on the other hand, is restricted to our own echo chambers. As a result, the aim of this project is to broaden the perspective on the global pandemic by using the power of Twitter data using Twint or Twitter API. Without a doubt, showing the successful effect of vaccination will bring positive impact for all people to be positive towards the vaccine and believe that vaccine is the life-saving key on the battle against COVID-19. By using sentiment analysis with Naïve Bayes approach, we can peer into an incredibly complex and wide-range conversation using natural language processing techniques (NLP). So, government and scientist can know the public opinion of the vaccine by looking at the visualized result of sentiment analysis using visualization tools such as Plotly, WordClouds, and etc in order to encourage people to trust the vaccine, and also promote the positive side of vaccine to change their mind by having campaign.

### SOLUTION

To acquire the public opinions towards the COVID-19 vaccine, we will be working on a project about the sentiment analysis of COVID-19 vaccination of all people around the world by collecting Tweets data about vaccination by using Twitter API or Twint. With the sentiment analysis with Naïve Bayes approach, we can classify the results of what people think of vaccine;

either positive opinion such as they are confident, believes, and having courage to the vaccination effect; or negative opinion such as they are not believing in it, and fear towards it. This sentiment analysis will be carried out by crawling the resources from the twitter with Twitter API or Twint to collect people's tweets about the vaccine related topic. The collected data will be processed and categorized into several classes using sentiment analysis with Naïve Bayes approach that can be categorize in positive, negative, and neutral based on the text in all of the tweets, and find out their expression such as grateful, happy, relief, sad, angry, fear, and no expression. The results of the sentiment analysis will help government and scientist to know the public opinion of the vaccine by looking at the visualized result with visualization tools from public resources such as WordClouds or Plotly to have a better information and insights.

#### **TARGET MARKET**

- Government
- Scientist
- Public community
- People that wish to use our machine

#### **COMPETITION/CONTRIBUTION**

This sentiment analysis research project will contribute as an important source of better insights about the public opinion towards the vaccination of COVID-19 in the current situation. This project will crawl data from Twitter using **Twint API**, and visualize all the collected data for analysis purpose. The machine learning model will help to classify the type of sentiment and the result will also be visualized for an overall image to everyone especially government and scientist to decide their next action to encourage people to be confident with the COVID-19 vaccination such as having campaign.

So, data visualization will be performed through the entire project flow by using varies charts, graphs, or tables by using Plotly, Bokeh, Wordclouds, and etc into a dashboard to visualize the data process, results of the analysis, and an overall insight of the sentiment analysis based on the COVID-19 vaccination to gives a better image to the government, scientist, and public community.

#### **Visualization tools that may be use :**

- Jupyter Notebook – Live code, calculations, narrative text, and visualisations can all be created with this open-source web application.

- Google Colab - Open-source browser that can create live code, equations, narrative text and visualizations.
- WordClouds – Graphical representations of word frequencies
- Matplotlib – Plotting library for the Python programming
- Seaborn – Data visualization library that draw informative statistical graphs
- Bokeh – Create visualization on modern web browsers such as plots or graphs into a complex dashboard
- Plotly – Provides graphing, statistics tools, analytics for Python programming

Throughout the sentiment analysis about COVID-19 vaccination, the result model of sentiment analysis and classifier using Naïve Bayes approach will be built to carry out the analysis of tweets by the people, and predict the outcome of those tweets crawled by Twint API. Then, all the data will be visualized by visualization tools such as WordClouds and Plotly to gain the insights about the results of sentiment analysis. After the sentiment analysis model is trained, this project will produce a machine that can be deploy in the real application with GUI. The GUI tools that may be use is the Tkinter from the python GUI sources.

## MILESTONES

### PROJECT 1

Tasks	Dates	Remarks
Introduction	29/3/2021	1. To confirm project scope 2. Includes: <ul style="list-style-type: none"> <li>• Introduction</li> <li>• Project specification</li> <li>• Research area</li> </ul>

Research Background	12/4/2021	1. To carry out background research, including literature review or preliminary research to gather requirements. 2. Includes: <ul style="list-style-type: none"> <li>• Theoretical considerations and Literature review</li> <li>• Other related work background</li> <li>• Preliminary research or feasibility study results</li> </ul>
Methodology and Requirements Analysis	25/6/2021	1. To present the results of the gathered requirements (if any) 2. Include: <ul style="list-style-type: none"> <li>• Research Approaches (e.g. questionnaire, experiments)</li> <li>• Research or development model (e.g. prototyping, machine learning, SDLC, agile)</li> <li>• Requirements Analysis</li> </ul>
Research Design	5/7/2021	1. To present the design diagrams, e.g. Machine learning steps 2. Include <ul style="list-style-type: none"> <li>• Collection of data</li> <li>• Data insights</li> </ul>
Submission Project 1 Portfolio	19/7/2021	Submit all updated work, including Proposal, Introduction, Research Background, Methodology and Requirements Analysis, Research Design, diagrams, etc. (in either softcopy or hardcopy)



**TAN YI HONG**  
RDS

017-4255258

yhtan-pm18@student.tarc.edu.my

Dr. Lim Siew Mooi

## **PROJECT 2**

<b>Tasks</b>	<b>Dates</b>	<b>Remarks</b>
Working preview	8/11/2021	Working preview
Result preview	15/11/2021	Final working result preview
Submission of draft Thesis	29/11/2021	Compilation of all contents, which may include: <ul style="list-style-type: none"> <li>• Future improvements or work</li> <li>• Conclusions of the entire research</li> </ul>
Submission of Final Thesis (Individual)	6/12/2021	

## APPENDICES

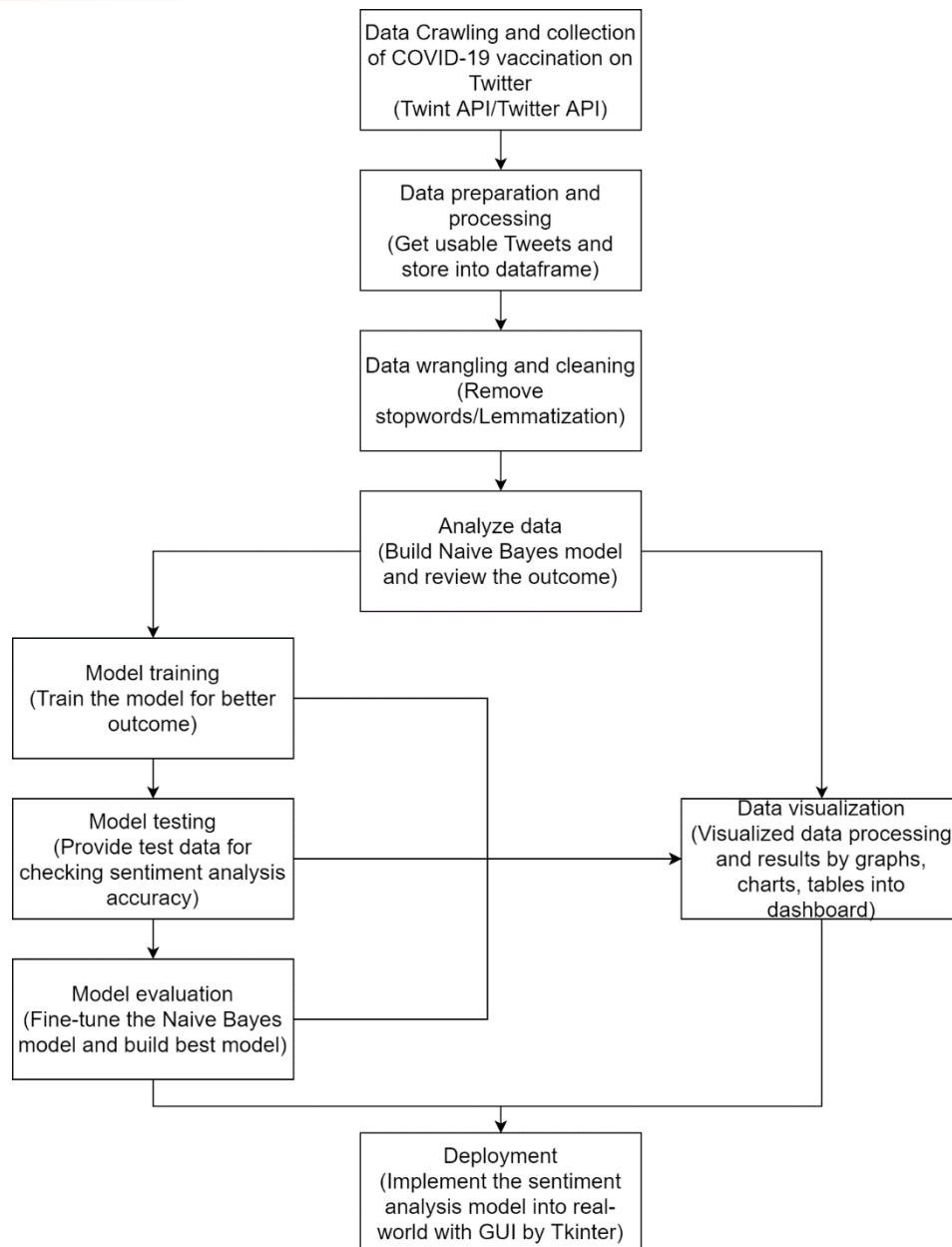


Figure 1 : Overall structure chart of the sentiment analysis

### Machine learning life cycle for sentiment analysis with Naïve Bayes approach:

- 1. Gathering Data** - The goal of this step is to identify and obtain all data-related problems. It is to collect data, and integrate data obtained from Twitter using Twint API/Twitter API to get a coherent set of data called as dataset.
- 2. Data preparation** - This step is to put the data collected into a suitable place and prepare it to use in our sentiment analysis model training. The first process is data exploration followed by data pre-processing, and stored into dataframe.
- 3. Data wrangling** - After process the data collected, this step helps to clean and convert raw data into usable format such as removing stopwords from texts and lemmatize the sentences. This step is important to clean the issues of the data such as invalid data, unreadable text.
- 4. Analyze data** - After cleaning data, this step is to build a Naïve Bayes model to analyze the data using various analytical techniques and review the outcome of the sentiment analysis. This is to take the data and use machine learning algorithms to build the model.
- 5. Train the model** - This step is to train the model to improve its performance for better outcome of the sentiment result.
- 6. Test the model** - In this step, the model is checked for the accuracy of sentiment by providing a test dataset to it.
- 7. Deployment** - After the above steps is complete, the final Naïve Bayes model is deployed in the real-world system with GUI using Tkinter.

**Distribution of the project :**

	Tan Teoh Xin Ee	Tan Yi Hong
Data collection	✓	✓
Data preparation and processing	✓	✓
Data wrangling	✓	✓
Analysis data	✓	✓
Train model	✓	
Test model	✓	
Model Evaluation	✓	✓
Data visualization		✓
Model deployment	✓	✓

**Table 1 : Distribution table for overall project**





TAN YI HONG

RDS

017-4255258

yhtan-pm18@student.tarc.edu.my

Dr. Lim Siew Mooi

## REFERENCES

1. K Bannister 2018, Understanding Sentiment Analysis, viewed on 6 Feb 2021.  
<<https://www.brandwatch.com/blog/understanding-sentiment-analysis/>>
2. M Rizvi 2020, 6 Essential Data Visualization Python Libraries, viewed on 6 Feb 2021.  
<<https://www.analyticsvidhya.com/blog/2020/03/6-data-visualization-python-libraries/>>
3. K Leung 2021, COVID-19 Vaccine – What’s the Public sentiment? , viewed on 10 Feb 2021.  
<<https://towardsdatascience.com/covid-19-vaccine-whats-the-public-sentiment-7149c9b42b99>>