Sentiment Analysis and Visualisation of global COVID-19 vaccination plan using Naïve Bayes algorithm

Tan Yi Hong
Bachelor of Computer Science (Honours) in Data Science
Supervisor: Dr. Lim Siew Mooi

TARC TUNKU ABDUL RAHMAN UNIVERSITY COLLEGE — BEYOND EDUCATION

Abstract

The COVID-19 pandemic has emerged as one of the world's most serious threats, and it is still very much a concern. This research project will carry out sentiment analysis using the Naïve Bayes approach and know the views of the people around the world towards their perception on the COVID-19 vaccine and determine the percentage of their responses regarding the vaccine for the final classification.

Problem Statement

With the ongoing vaccination campaign globally, governments and scientists need to prove and show the success of the development of vaccines to strengthen the confidence of all people to fight against COVID-19. Without a doubt, leading the successful effect of vaccination and informative insights by visualisation will positively impact all people to be positive towards the vaccine and believe that the vaccine is the life-saving key in the battle against COVID-19.

Objectives

- Crawl the resources from Twitter with Twint API to collect people's tweets about the vaccine-related topic.
- Classify the sentiment of the opinions on vaccination from the tweets collected.
- Develop a classification model with Naïve Bayes approach to applying in the real-world system.
- Visualize the data analysis and results to display valuable and informative insights.

Design and Methodology
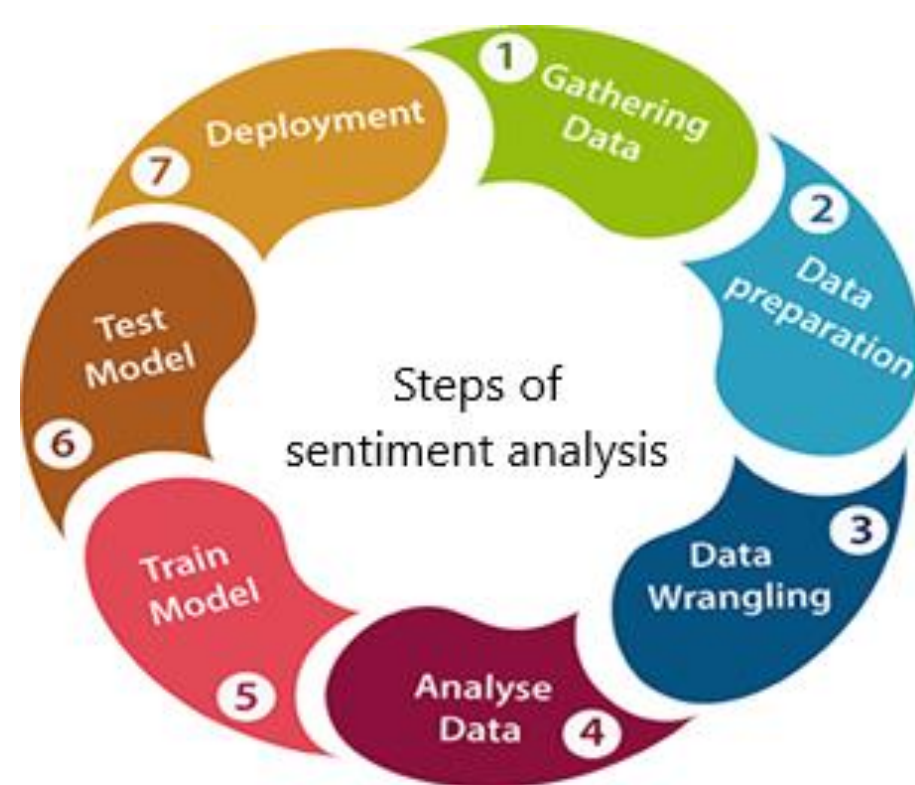
Gathering Data -> Data preparation -> Data wrangling -> Analyse data -> Model Training -> Model Testing -> Deployment

Machine learning has given computers the ability to learn on their own without having to be explicitly programmed. However, how does a machine learning system function? As a result, the machine learning life cycle can be used to explain it. A machine learning project's life cycle is a cyclic method for developing a practical machine learning project. The life cycle's primary goal is to find a solution to the problem or project.

Understanding the problem and knowing the problem's purpose are the most critical aspects of the entire procedure. As a result, we must first comprehend the problem before beginning the life cycle, as a positive outcome is contingent on a thorough comprehension of the situation. To address an issue, we develop a machine learning system called a "model" in the complete life cycle process, and this model is built by providing "training." However, we need data to train a model. Therefore the life cycle begins with data collection.

Construction and Testing

Research Method
- Data Collection - Twint API
- Text Cleaning - NLTK, spaCy, etc.
- Data Visualization - Plotly, Wordclouds
- Model GUI Design – Tkinter

Experiment Testing
- Disparity of preprocessed data and non-preprocessed data
- Minor differences in raw data can lead to incorrect assumptions
- Pre-processing the original dataset is crucial for this sentiment analysis research

Table 1 : Experiment Testing

Program Name :	Experiment Testing			
Test Date : 10/11/2021		Tester : Tan Yi Hong		
No.	Objective/Test Cases	Test Data	Expected Results	Actual Results	Remarks
1.	Importance of Data Preprocessing	Original dataset without preprocessed	The information and insights given by the Data Visualization on NLP does not contains meaning.	The word clouds and the graphs on NLP has fault insights.	

System Testing
- Determine the capability of sentiment detector GUI system
- Functional testing, interface testing, performance testing, and end user testing

Table 2 : System Testing

Program Name :	System Testing			
Test Date : 10/11/2021		Tester : Tan Yi Hong		
No.	Objective/ Test Cases	Test Data	Expected Results	Actual Results	Remarks
1.	Positive Sentiment Detector	A positive sentence : "I am feeling good right now."	The sentence overall sentiment will be rated as 'Positive'.	The sentence overall sentiment is rated as 'Positive'.	
2.	Neutral Sentiment Detector	A neutral sentence : "The book has 40 pages".	The sentence overall sentiment will be rated as 'Neutral'.	The sentence overall sentiment is rated as 'Neutral'.	
3.	Negative Sentiment Detector	A negative sentence : "This glasses is not useful at all!".	The sentence overall sentiment will be rated as 'Negative'.	The sentence overall sentiment is rated as 'Negative'.	
4.	System Recoverability	User press the 'Reset' button.	The text input field will be cleared and results field will be empty.	The text input field is cleared and results field becomes empty.	
5.	System Durability	User use it to test for 15 sentences	The system will be able to execute all sentences successful without having issues.	The system is able to execute all sentences successful without having issues.	
6.	System Performance	A random sentence as input.	The system will be able to detect the sentiment of the input sentence within 1 second.	The system detects the sentiment of the input sentence within 1 second.	

Contribution

This sentiment analysis research study contributes as a vital source of better insights into public opinion about COVID-19 vaccination. The machine-learning model will help to classify the types of sentiment, and the results will be visualised for everyone. And with the data visualization, overall insight of the sentiment analysis based on the COVID-19 vaccination will give a better image to the government, scientists, and general public.

Conclusion

While the vaccine has resurrected the fight against COVID-19, it has also triggered a wave of anti-vaccination rallies. So, it is important to gauge the public opinion towards the vaccination that has currently been given to almost all of the country in the world. Employing sentiment analysis on recent Twitter data by crawling Tweets data on vaccination on Twitter was beneficial in gauging public opinion on the COVID-19 vaccine.