Twitter Stance Analysis towards COVID-19 Vaccination Using Machine Learning Classifiers

by

Fardin Rahman (170021068)
Lamim Ibtisam Khalid (170021052)
Muntequa Imtiaz Siraji (170021038)

A Thesis Submitted to the Academic Faculty in Partial Fulfillment of the Requirements for the Degree of

BACHELOR OF SCIENCE IN ELECTRICAL AND ELECTRONIC ENGINEERING

Department of Electrical and Electronic Engineering
Islamic University of Technology (IUT)
Gazipur, Bangladesh

10 May 2022

Declaration of Authorship

We, declare this thesis titled, "*Twitter Stance Analysis towards COVID-19 Vaccination Using Machine Learning Classifiers*" and the works presented in it are our own. We verify that:

- This work has been done for the partial fulfillment of the Bachelor of Science in Electrical and Electronic Engineering degree at Islamic University of Technology (IUT).
- Any part of this has not been submitted elsewhere for the award of any Degree or Diploma.
- We have always clearly attributed the sources when we have consulted the published work of others.

Authors:

Fardin

Fardin Rahman
ID: 170021068

Lamim

Lamim Ibtisam Khalid
ID: 170021052

Muntequa Imtiaz Siraji
ID: 170021038

Approved by:

12 May, 2022

Mr. Safayat Bin Hakim

Supervisor and Assistant Professor,
Department of Electrical and Electronic Engineering (EEE)
Islamic University of Technology (IUT),
Boardbazar, Gazipur-1704.

Date: 10.05.22

Table of Contents

List of Tables

List of Figures

List of Acronyms

API	Application Programming Interface
BERT	Bidirectional Encoder Representation from Transformers
CNN	Convolutional Neural Network
FN	False Negative
FP	False Positive
HTML	Hypertext Markup Language
KNN	K-Nearest Neighbor
LDA	Latent Dirichlet Allocation
LR	Logistic Regression
LSTM	Long Short-Term Memory
ML	Machine Learning
NB	Naïve Bayes
NLP	Natural Language Processing
NLTK	Natural Language Toolkit
NPI	Non-Pharmaceutical Intervention
NRC	Natural Research Council
RF	Random Forest
RNN	Recurrent Neural Network
SVM	Support Vector Machine
SWN	Sentiment WordNet
TF-IDF	Term Frequency-Inverse Document Frequency
TN	True Negative
TP	True Positive
URL	Uniform Resource Locator
VSM	Vector Space Model
WHO	World Health Organization
XGB	XGBoost

Acknowledgements

This thesis has been prepared under the guidance of Mr. Safayat bin Hakim, supervisor for our EEE 4800 (Project and Thesis) course at Islamic University of Technology. We would like to offer him our most sincere gratitude for guiding us with his knowledge, experience and care.

We would also like to thank all our teachers, classmates and staff members for their assistance and providing valuable information which aided in the completion of this thesis.

2022

Abstract

The COVID-19 pandemic has impacted the world as a whole in ways which were unimaginable before. From the economical and medical impacts to geopolitical views and influences, COVID-19 has changed the world as we see it. Since the introduction of different vaccines to prevent COVID-19, people's opinions have been divided regarding it. The social media platform, Twitter, provides a noteworthy platform for voicing opinions in support of and against the vaccines which results in long debates and discussion and often spreading of misinformation. In this paper, a dataset has been manually collected from twitter using the Twitter API and tweets were manually annotated into three distinct categories – *provac*, *antivac* and *other*. Six machine learning algorithms were used to train and test on the annotated data and the best classifier for this case was identified. Using the best classifier, the whole dataset was automatically annotated and stance towards the COVID-19 vaccine was analyzed. Further analysis was done to identify changes in trends of people's opinions over time. The results indicate that, with proper implementation of the ML algorithms, it is possible to identify and predict people's stances towards the COVID-19 vaccine and similar approach can be used in analyzing stance towards other vaccines and treatments of various diseases.

Keywords— COVID-19, Vaccines, Stance Analysis, Machine Learning Algorithms

Chapter 1

Introduction

Ever since World Health Organization (WHO) declared the novel COVID-19 as a global pandemic on March 11th of 2020, many preventive measures (e.g., banning all sorts of gatherings, citywide lockdown, self-isolation etc.) were initiated to reduce the transmission of the acute virus. Four months prior to that in December 2019, some individuals were affected with severe respiratory infections originating from a fish and seafood market in Wuhan, China and the disease would later be identified as SARS-CoV-2 or coronavirus. Due to the protective measures, people eventually refrained from physically going to their workplace and started spending more time on internet for online communication and job purposes. However, numerous issues started to become afloat on the internet like discussions of a Chinese virus, biological weapon, running out of toilet tissues along with the general frustration and sufferings due to Non-Pharmaceutical Interventions (NPIs) namely quarantine, social distancing etc. Aside from these topics, there were instances of spreading false information about the pandemic resulting in different levels of depression and stress in many individuals [1]. Till to this date, there have been over 370 million cases with over 5 and a half million deaths worldwide listed, added on top of the casualties caused by common respiratory infection (flu) [2]. Nevertheless, from the beginning of 2021, vaccine started to roll out as a shield to protect individuals from being affected by the virus. Different countries eventually commenced their vaccination programs from different points of time complying with their healthcare standard and regulations, However, people were still divided between whether to get vaccinated or not, resulting in many discussions and arguments across all social media, micro-blogging sites, and online forums. With the help of computer science and availability of data, all these disputes can be studied in-depth to initiate a proper response in the case of any uncertain circumstance.

In order to understand public perception about any topic, there is no other platform on the internet but social media where millions of people share their ideas and opinions with the outer world every day. Twitter is one of the most popular social media platforms with more than 330 million active users [3], where people have the opportunity to tweet while others have the option to re-tweet, like and comment. These texts contain an individual's perception and feelings regarding public events, news, stories etc. and the data can be utilized for social or economic experiments. The experiments are performed for enhancing industrial efficiency, policy development and revealing insights which can be useful in many different domains. It is to be noted that people's behavior is often influenced by social media content, from positive to negative bias and the threat of the infectious disease can make people think in unreasonable ways. In addition, twitter data can be analyzed to track real-time public panic, dynamic data trends and evolution status of a disease [4-6]. Government officials of different countries also utilize this facility to share up-to-date instructions and bring awareness by introducing future plans to people.

1.1 Introduction to Stance Analysis

One of the types of research based on twitter data is analyzing people's attitude and behavior towards a specific entity known as stance analysis which belongs to sentiment analysis in the field of Natural Language Processing (NLP). It mainly involves building an automated process of determining a text's polarity and subjective content from the written language contributing to the understanding of human perception. As mentioned earlier, the twitter user demographic is huge and it would be impossible to interpret a general overview of people's reaction manually from the enormous number of tweets being posted every day. On the other hand, due to increasing demand of sentiment analysis models built for different applications of science, education, and other literature, the performance also needs to be evaluated in a conscious manner. The reason for the unreliable results is mainly due to the inconsistency that can be observed in a sentence as social media users tend not to follow grammatical rules or narratives properly. Hence, a reliable and robust automated process will have the capability to predict different outcomes for large amounts of data with relative accuracy. This type of analysis can greatly help healthcare professionals and governments to take justified actions towards containing the pandemic along with formulation of strategies for future outbreaks [7].

1.2 Challenges of Stance Analysis

Similar to any other text analysis work, there are some challenges which make stance analysis difficult. These challenges can give a wrong impression on the accuracy of the machine learning models. The main challenges of stance analysis are:

- **Sarcasm detection**: Sarcasm is the use of words in an ironic way to convey a different, often opposite, meaning than what is actually being said. The main objective of sarcasm is to mock and make fun of something without being direct about it. In text format, it becomes difficult to identify whether someone is being sarcastic or not without knowing anything about their situation or expression. As a result, even humans make errors to identify sarcasm. This can lead to mislabeling text into the wrong category. Moreover, if a lexicon-based approach is used to train the machine learning models, unless the majority of the dataset consists of sarcasm, the models will not be able to identify whether a particular word or phrase is being used sarcastically or not.
- **Emojis**: Emojis are a way to convey feelings without using words. But these emojis can often appear as spam or be used in the wrong context. Moreover, the diversity of emojis means that they could be considered as a language on their own.
- **Negation**: Negation is the use of words like not, never, cannot, etc. which changes the meaning of the word it relates to into the opposite meaning. Without taking this into consideration, machine learning models can easily mislabel text. So, during the training phase, it is important to take negation words into consideration to shift the polarity of whatever word they relate to in order to get the accurate prediction.
- **Idioms and metaphors**: People often use idioms and metaphors to better explain their thoughts on social media. These idioms and metaphors tend to use words in an unintended way where the meaning is more abstract rather than literal. Comparisons are made to imaginary situations in order to better convey

a thought. The models are not able to identify when a phrase is being used literally and when the same phrase is being used metaphorically. This can also end up mislabeling the text.

- **Videos and images**: In the case of social media, a lot of the time, people share videos or images which convey how they feel or what they are concerned about. Sometimes, text from images can be extracted in order to understand the content of the message better but text-less images as well as videos cannot be analyzed by machine learning models which are trained for text analysis.
- **Human labelling bias**: In order to train a model properly, it is important to have good training data. One of the best ways to annotate text information is to annotate it manually. This ensures that the varying ideas behind the text are properly understood and labelled into the right category. But it is not always possible to perfectly determine what category a certain piece of text belongs to. This is especially possible when annotating data on an unfamiliar topic or something which is more subjective in nature. Disagreement in how data should be annotated can lead to the models learning poorly. Usually manually annotated data is cross-checked among others in order to verify the integrity and to avoid any form of bias during the annotation process.

1.3 Problem Definition

Stance analysis is basically a classification task as it assigns different classes to different texts and the approaches can be categorized into three main divisions, lexicon/dictionary-based, statistical/machine learning-based and hybrid model-based [8]. Lexicon-based analysis depends on the aggregate value generated from a predefined effect value of the words by identifying each word in a sentence. The predefined effect value is contained in an existing lexicon such as SentiWordNet, NRC Word-Emotion Association lexicon etc. Though this value is quite stable, it does not always remain accurate as many sentences contain patterns such as denial, word turns etc. To avoid this validity issue, there have been development of human-annotated specialized dictionaries which happened to be quite cumbersome to build. In the meantime, Machine Learning (ML) algorithms and deep learning neural networks use a specific dataset or a portion of dataset for training purpose and then the models try to predict different stances for other datasets or the other portion of dataset. This approach tends to perform well when the research is focused to a specific domain of interest and not to a wide variety of fields. Then comes the hybrid approach which utilizes both the aforementioned approaches by inheriting lexicon-based models' stability and the ML-based models' accuracy on larger data.

1.4 Comparison to Sentiment Analysis

Stance analysis differs from sentiment analysis. While they are both subtasks of opinion mining, sentiment analysis focuses on whether a statement is positive, negative or neutral [8]. On the other hand, stance analysis gives an idea of whether a piece of text agrees or disagrees with a particular viewpoint regardless of how positively or negatively it is worded. Essentially, stance analysis cares more about the meaning behind the text in order to identify which side the author supports in a discussion while sentiment analysis only cares about whether the text conveys a positive or negative feeling towards the discussion. This means that in the case of stance analysis, sentiment alone cannot explain which side the author supports without knowing the exact topic the sentiment is towards.

1.5 Thesis Contribution

The aim of this paper is to detect and analyze people's stance towards COVID-19 vaccination by following the ML-based approach and thus classifying an arbitrary individual to one of three classes namely in favor (*provac*), against (*antivac*) and impartial (*other*) to vaccine. Though there is an increase in interest to get vaccinated in a short span of time, negative impression about vaccine can still be observed on social media platforms. The most cost-effective measure to avoid any disease is to take vaccines stated by WHO, reported in 2019 about the threats to global health [9]. Furthermore, the relative change in perception that may have occurred after the second wave compared to the first wave of COVID-19, vaccination process especially booster dose can follow a different course than before. Details of each category along with analyzing process and findings are discussed in the following sections.

1.6 Structure of the thesis

In the thesis, at first, the relevant research works are briefly discussed in Chapter 2. In this chapter, literature review is done on data pre-processing techniques, data extraction methods, performances of different machine learning models as well as exploration of some stance analysis challenges. In Chapter 3, the methodology that we followed to conduct our research is discussed in detail. The data collection process, data pre-processing, feature extraction and model training is elaborated. The results from the research are discussed in Chapter 4 as well as provides an analysis of the overall data. Finally, Chapter 5 contains the conclusion and future work intentions.

Chapter 2

Related Work

Extensive work has been done on the processes of stance analysis which tackle different challenges in the research including the pre-processing, feature extraction and analysis processes. These research works are briefly discussed in the following sections.

2.1 Related Work in Data Pre-processing

There are many different parts and aspects of stance analysis process that can be explored and scrutinized in depth. For instance, Jianqiang et al. [10] made an extensive analysis on text pre-processing techniques effect on the performance of sentiment classification and the results show that removing URLs, stop words and numbers does not have impact on the performance of classifiers significantly whereas negation replacement and acronym expansion can improve the accuracy. The negation context is elaborately studied by Gupta et al. [11] where they presented a hybrid approach of generating the feature vectors using SWN-based lexicons with Support Vector Machine (SVM) classifier and they achieved about 6 percent improvement over traditional methods by only handling negations and negation exceptions without any use of other features. Moreover, Kouloumpis et al. [12] investigated the usefulness of the existing lexicon-based resources related to sentiment and emotions along with informative features and in the end, they concluded that micro-blogging features (e.g., intensifiers, positive/negative/neutral emoticons and abbreviations) were comparatively more useful than existing sentiment lexicons. More in-depth study about feature extraction was performed by Xuan et al. [13] who classified different category of features (text, user etc.) by feature engineering for stance classification and showed state-of-the-art performance with simple Logistic Regression (LR) model using those attributes.

2.2 Related Work in Data Extraction

Within this domain, some studies [14-19] were dedicated towards building specialized lexicon/corpus to classify texts with better precision. Saif et al. [14] built a new twitter evaluation dataset named "STS-Gold", where their focus was on entity-based manual annotation. Rice et al. [15] have researched on developing a semi-supervised method for building a sentiment dictionary to analyze text where generalized dictionaries tend to perform poorly. Besides, Bandhakavi et al. [16] have explored the impact of an emotion corpus based on psychology mapping between emotions and sentiments from computational perspective. Furthermore, there have been development of dataset in other language than English regarding the pandemic, such as "ArCovidVac" by Mubarak et al. [17] who have fine-tuned the data into different layers such as importance-based, information-based, and stance-based. Dimitrov et al. [18] also formulated a corpus-"TweetsCOV19" which was annotated based on different semantics and includes metadata, entities, sentiments etc. to explore in depth about online discourse of the outbreak. Manual annotation has been also used by Hayawi et al. [19] to build a dataset called "ANTi-Vax" for detecting misinformation regarding COVID-19 vaccine and classification analysis was performed as well using ML models.

2.3 Related Work in Unsupervised Learning

Another category of research works [20-23] mainly focuses on identifying themes, patterns, or topics from different datasets and showing them using various tools and methods in an elegant way. For example, Xue et al. [20] used unsupervised Latent Dirichlet Allocation (LDA) ML method for analyzing tweets and identifying topics, themes etc. which showed that people's fear for the unknown coronavirus was prominent but symptoms or treatments related information was not frequently discussed. In a study by Boon-Itt et al. [21], LDA was employed for exploring different trends and concerns of the COVID-19 pandemic and the results were further analyzed into three aspects (spread and symptoms, outlook of the people, and recovery from the pandemic). Another research from Dubey [22] was focused on different countries' citizen and their reaction to COVID-19 outbreak where after applying NRC Emotion lexicon, the results concluded that apart from the minor presence of negative feelings (disgust and sadness), positive approach was prevalent across the world. In one study from Cotfas et al. [23], a stance analysis has been conducted on general people's perception to wear masks by using annotated tweets from one year period and the findings have suggested most percentage being in favor of mask-wearing in public.

2.4 Related Work in Sentiment Analysis

In order to perform sentiment and polarity analysis, some researchers [24-26] employed a basic method of using python's built-in textblob library and assigning polarity scores to labels (e.g., positive, negative and neutral). In particular, Manguri et al. [24] aimed to perform such analysis by extracting tweets from twitter using hashtags \#coronavirus, \#COVID-19 and they found that neutral class of tweets was significantly higher than the other two classes, positive and negative with objective information being predominant. Another example comes from Nemes et al. [25] where they have compared the performance of textblob with Recurrent Neural Network (RNN) for classifying emotions on tweets regarding the pandemic and their results showed that the method tends to overestimate the neutral tweets which fails to be a sustainable solution. Meanwhile, Saleh et al. [26] have examined social distancing related discussions by using LDA for topic modelling with python's built-in textblob library and the findings showed the users being devoted to the regulation in the initial stages of the pandemic.

Apart from this category of literature, there are many studies [27-33] that were dedicated for building ML models or neural networks to detect sentiments with certain accuracy. With the aim of deducing the public sentiments on coronavirus into positive and negative class, Machuca et al. [27] have used 1.6 million tweets to train their LR model for the binary classification and obtained an accuracy of 78.5 percent. Another research study from Yadav et al. [28] have utilized a dataset named Sentiment140 from Kaggle to train three ML algorithms namely, LR, Naive Bayes (NB), SVM and achieved 83.71 percent accuracy with linearSVC. The Sentiment140 dataset along with Emotional Tweets dataset were used for training Deep Long Short-Term Memory (LSTM) classifiers by Imran et al. [29] where they analyzed people's reaction from different cultures regarding COVID-19 pandemic and achieved state-of-the-art accuracy score for detecting polarity and different emotions. Neural network was further discussed and improvised in the research of Behera et al. [30] where they went for a hybrid approach comprised of Convolutional Neural Network (CNN) and LSTM for classifying sentiment of customer reviews. The capability of ML models has been examined in a work of Sireesha et al. [31] where they have employed two ML classifiers (LR and NB) and the results yielded 87 percent and 81 percent accuracy respectively. Alenezi et al. [32] experimented on social media misinformation and proposed three detection models, LSTM networks, multi-channel convolutional neural network (MC-CNN) and k-nearest neighbors (KNN) by evaluating their performance. In terms of other NLP models, Bidirectional Encoder Representations from Transformers (BERT) and the bi-LSTM were used to identify anti-vaccination tweets by To et al. [33] and BERT outperformed other classical (SVM and NB) models.

2.5 Related Work Exploring Challenges

Some of the notable works [34-41] largely utilized annotated datasets for conducting stance analysis in the midst of global vaccination campaign against coronavirus. Muller et al. [34] addressed the problem of concept drift, which is the sudden shift of discussed topics, when training ML models and the results also indicated the occurrence of misclassification of data unless the problem being taken into account. Villavicencio et al. [35] have performed a twitter sentiment analysis towards COVID-19 vaccines in Philippines where they have trained NB algorithm using manually annotated 993 tweets for classifying three stances (positive, negative, neutral) with 81.77 percent accuracy. Ebeling et al. [36] have investigated political influence on COVID-19 vaccination stance in Brazil and they concluded that a strong bias remains in people when they choose which side to support. Bi et al. [37] have attempted to construct an opinion mining model using which they studied Chinese population's behavior in disputes regarding vaccination with the phenomenon of truth decay and the results show a linear relationship between the negative and neutral attitudes. Another group of researchers from Italy, Giovanni et al. [38], have presented a semi-automated approach based on refined hashtags to label large amount of data and performed stance analysis by employing BERT-based binary-classifier. Moreover, Sang et al. [39] made a study about transfer learning where they have used social distancing measure dataset to train ML models which then used to predict other preventive measures. Hesitancy to get vaccine has been thoroughly examined by Cotfas et al. [40-41] for the first month of vaccination by employing several ML algorithms and deep learning models and illustrated the concurrent events connection with the topic.

2.6 Chapter Summary

Stance analysis is an important research field and many researches has been conducted in this field but still some challenges remain to be further addressed. Different challenges in the research including the pre-processing, feature extraction and analysis processes were discussed by referencing various researchers. Similarity and drawbacks of different researches were also pointed out as well as the successive work that attempted to solve them.

Chapter 3

Methodology

We have presented our framework which explains the entire data analysis process from beginning to end. We collected data from Twitter which were then labelled manually and cleaned for use by the machine learning models. The data was split into training set and test set which was then used to estimate the accuracy of the models in predicting the stance of the users. A detailed description of each step of the process is given in the following subsections and a workflow diagram is shown in Figure 3.1.

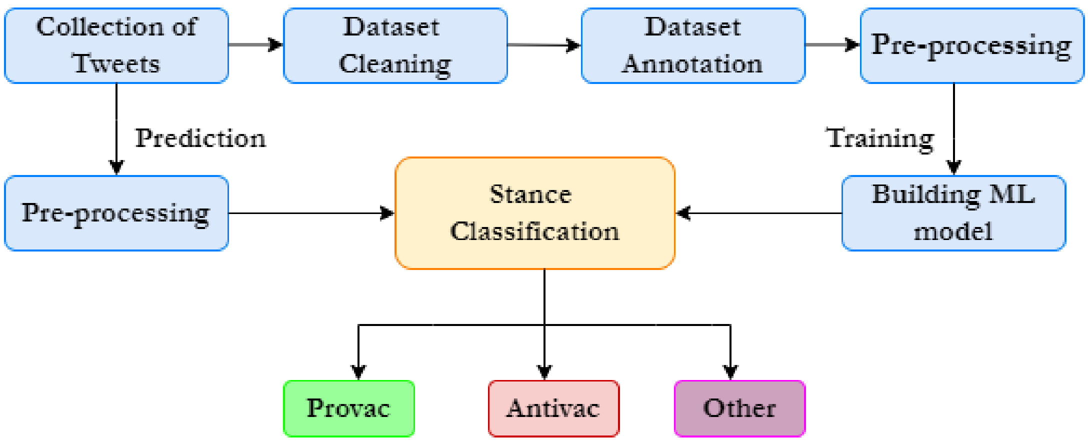

Figure 3.1: Workflow Diagram

3.1 Data Collection

Data collection is the process of gathering relevant data for use in research. This is the first step for stance analysis. Data in the form of text messages is collected from various social media platforms such as Twitter, Facebook, YouTube, different forums and blogs, etc. [42-43]. For the purpose of our research, we extracted tweets from

Twitter. A tweet is a message shared publicly by users on Twitter and each tweet can have a maximum of 280 characters. Tweets can also hold image and video information along with the usual text information. We have used the Twitter API to extract these tweets. The API allows us to extract a wide range of information including the original tweet, time of tweet, username, user location, followers, likes, retweets and other information which have been made public by the user. The Twitter API only allows limited access to the data stream and so only a small amount of the daily tweets can be extracted. Only tweets in the English language containing keywords including ‘vaccine’, ‘covid’, ‘pandemic’, ‘antivaxxer’, ‘vaccinated’ and ‘covid19’ were collected along with the date and time of the tweets. 73290 tweets were initially scraped from 5 October, 2021 to 3 November, 2021 and saved in a .csv file, shown in Figure 3.2. A very limited number of tweets were collected from between 1 November, 2021 and 3 November, 2021 but they were still included in the dataset. Afterwards, retweets and duplicate tweets were removed to finally arrive at a dataset of 51761 tweets. The final set of tweets were then partially cleaned to remove any kind of noise from the text such as broken links and errors in decoding.

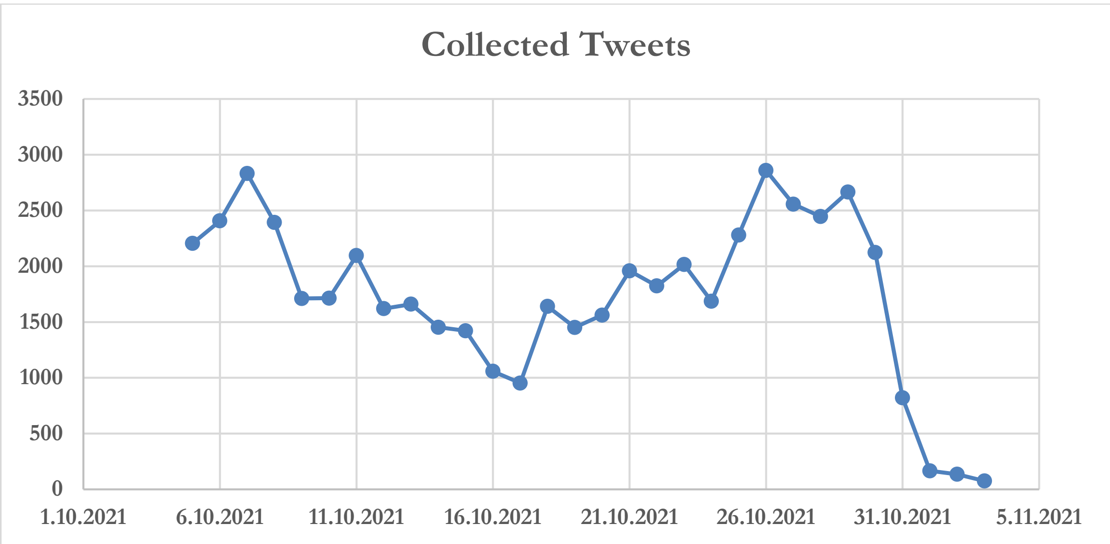

Figure 3.2: Number of Collected Tweets

3.2 Data Annotation

Of the 51790 tweets collected, 5986 tweets were manually labelled into three categories which were *provac*, *antivac* and *other* by four researchers.

- **Provac**: This class was given to those tweets which made it obvious that the user was vaccinated or at least in support of the vaccine drive. Tweets sharing information about which dose number the person is at or showing disdain towards people who are against vaccination also fall under this category.
- **Antivac**: This class was given to those tweets that showed hate towards vaccines or vaccinated people. Those unwilling to receive the vaccine in its current state was also given this label. Besides, there were instances of spreading misinformation of vaccines or supporting natural immunity against COVID-19 which were considered to be part of *antivac* label.
- **Other**: This label was for the tweets which showed the user to be unbiased towards both vaccinated and unvaccinated people. Moreover, the tweets which, regardless of the vaccination status of the user, showed sympathy or support to the opposite class were also given the *other* label. Irrelevant tweets were given this label in order to differentiate from the tweets which were completely irrelevant to vaccination stance.

Figure 3.3: Stance Distribution for Training

Initially, *provac* category had 570, *antivac* category had 534 and *other* category had 4366 tweets. Due to the high number of tweets labelled as *other*, only a limited number of them were included in the final training dataset in order to avoid learning bias. The number of tweets in *other* category were then 500 and the overall training dataset in stance numbers is shown in Figure 3.3. Examples of tweets belonging to each category are given in Table 3.1.

Table 3.1: Example of Labelled Tweets

Tweet	Label
' Already got mine a few weeks ago. Super sad that Im still not magnetic and therefore, still losing my keys. \#GetVaccinated'	*Provac*
The vaccine will make you sick, not protect you or others. \#covid \#vaccine	*Antivac*
Does anyone else think that \#coachbeard is the most profound voice of our time or is that just me? \#momlyfe \#2021 \#pandemic'	*Other*

3.3 Data Pre-processing

Tweets are structured a bit differently than normal text information as different elements such as images, emojis, gifs, hashtags, acronyms, colloquialism etc. are used alongside text. Tweets can usually be expected to contain a lot of information which machine learning algorithms cannot benefit from. Rather, certain elements in a tweet can reduce the performance of the model. These elements need to be cleaned out of the tweets before the data becomes usable. We took the following pre-processing steps:

- Removal of usernames: Any characters after '@' which denote usernames were removed. Usernames are too diverse which increases the number of features in the model and do not provide any information on the stance of a person.
- Removal of weblinks: URLs were removed by searching for characters after https. These are removed because the links lead to websites which cannot directly be analyzed by the learning models for any form of sentiment.

- Removal of special characters: Any character outside of A-Z and a-z were removed since they do not normally provide any useful information.
- Removal of white space: Any extra white spaces were replaced to be only one white space.
- Case conversion: All letters were changed to lower case. This was done in order to reduce variations of the same words in lower and upper case.
- Extension of contracted words: Contracted words like can’t, couldn’t, shouldn’t, won’t, etc. were extended in order to be consistent with text where the word is already written in the extended form.
- Removal of stop words: Stop words are words like a, an, the, and, or, of, should, could, etc. which do not contain any information on the polarity of the information. These were removed using NLTK in Python.
- Removal of emojis: While emojis can provide useful information about the contents of a tweet, they do not provide any information on the stance a person takes. For this reason, emojis were also removed.
- Removal of miscellaneous encoding elements: During the extraction process, some tweets ended up with HTML elements like ‘\&’ as a part of the decoding process. These were replaced with the appropriate characters e.g., ‘\&’ was replaced with ‘and’.
- Lemmatization: Words can have different inflected forms such as ‘to go’ can have forms like ‘going’, ‘gone’, ‘went’. These all actually have the same meaning. In order to simplify the learning process, these words need to be changed back to their root form. Lemmatization was done using the WordNetLemmatizer library from NLTK.

After cleaning the tweets, they were saved in a separate column in the same csv file. A few examples of the cleaned data are shown in Table 3.2.

Table 3.2: Example of Cleaned Tweets

Original Tweet	Clean Tweet
Texas Gov. Greg Abbott bans any \#COVID-19 \#vaccine mandates | \#Texas Governor shows utter	texas gov greg abbott ban covid vaccine mandate texas governor show utter

disrespect for \#publichealth. \#covid \#pandemic	disrespect publichealth covid pandemic (*provac*)
A different \#vaccine "may also be considered based on vaccine supply and access considerations". \#COVID19 \#Pandemic \#Vaccines \#ThirdDose \#WHO \#SAGE '	different vaccine may also considered based vaccine supply access consideration covid pandemic vaccine thirddose sage (*provac*)
The dark winter is coming for the vaccinated. Pandemic of the vaccinated is real. \#DarkWinter \#Pandemic '	dark winter coming vaccinated pandemic vaccinated real darkwinter pandemic (*antivac*)
VanDyke Im not vaccinated not so much as a sniffle since before the \#pandemic started. Im an essential worker too. Do your research then make a decision on whats best for you.'	im not vaccinated not much sniffle since pandemic started im essential worker research make decision whats best (*antivac*)
\#HumanitarianResponse: Teams in \#WestBengal distributed 564 food kits across Malda and North 24 Parganas in the state. The communities in these districts have been severely affected by the \#pandemic and the subsequent \#Cyclones that struck the eastern coast of India.	humanitarianresponse team westbengal distributed food kit across malda north parganas state community district severely affected pandemic subsequent cyclone struck eastern coast india (*other*)

3.4 Feature Extraction

Any form of text-based opinion mining work relies on the use of words in the text. In the case of stance analysis, after having annotated the data into different categories, machine learning models are used to determine what category any new text belongs to. This can be accomplished by comparing the frequency of the words with that in the annotated dataset. Certain words may end up being used more in case of a particular category than another. TF-IDF (Term Frequency-Inverse Document

Frequency) feature extraction method was used to assign a score to each word in a tweet based on the frequency of the word in the tweets across the entire dataset [44]. This score represents the weight of each word within the text. A higher score signifies that the word plays a bigger role in determining a piece of text as belonging to that category. The score of each word in the unclassified text is then used to determine the overall stance of the text. More frequent words which tend to belong in all the categories and do not have any significant impact on the stance of the text. These words get a low score and affect the overall stance prediction less. This process turns the text information into a Vector Space Model (VSM) which is then used to train the different prediction models. This principle can be extended to more than just singular words by considering 2,3 or more words as a single feature. This can help to classify text based on phrases. When only a single word is considered, it is called a unigram. For two words it is called bigram and so on. In our research we have considered unigrams, bigrams and trigrams in order to train the models. A limit of 2000 feature words was set during the learning process to eliminate words which appeared too frequently.

3.5 Machine Learning Models

Six machine learning models were used for the purpose of the research. They were Logistic Regression, Support Vector Machine, Random Forest, K-Nearest Neighbor, XGBoost and Naïve Bayes. The dataset was split into training and test dataset. 80 percent of the data was used for training and the rest 20 percent was used for testing by each model. In the learning models, some parameters for the algorithm were tweaked using a technique called hyperparameter tuning. The grid search method (GridSearchCV) was used to optimize the parameters. This method involves running the model multiple times using different parameter values within a specified range to find the best values for the parameters [45].

3.5.1 Logistic Regression

This algorithm is based on the concept of probability and works using what is known as 'logistic function' as the cost function. This cost function is bounded within 0 and 1 by the hypothesis of the algorithm and the formula is as follows:

$$\theta(x) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 X)}}$$

In classification problem, a threshold probability value is set to distinguish one class from another [46]. In the case of multiclass regression, the principle is the same except that there are M possible outcomes rather than just 2.

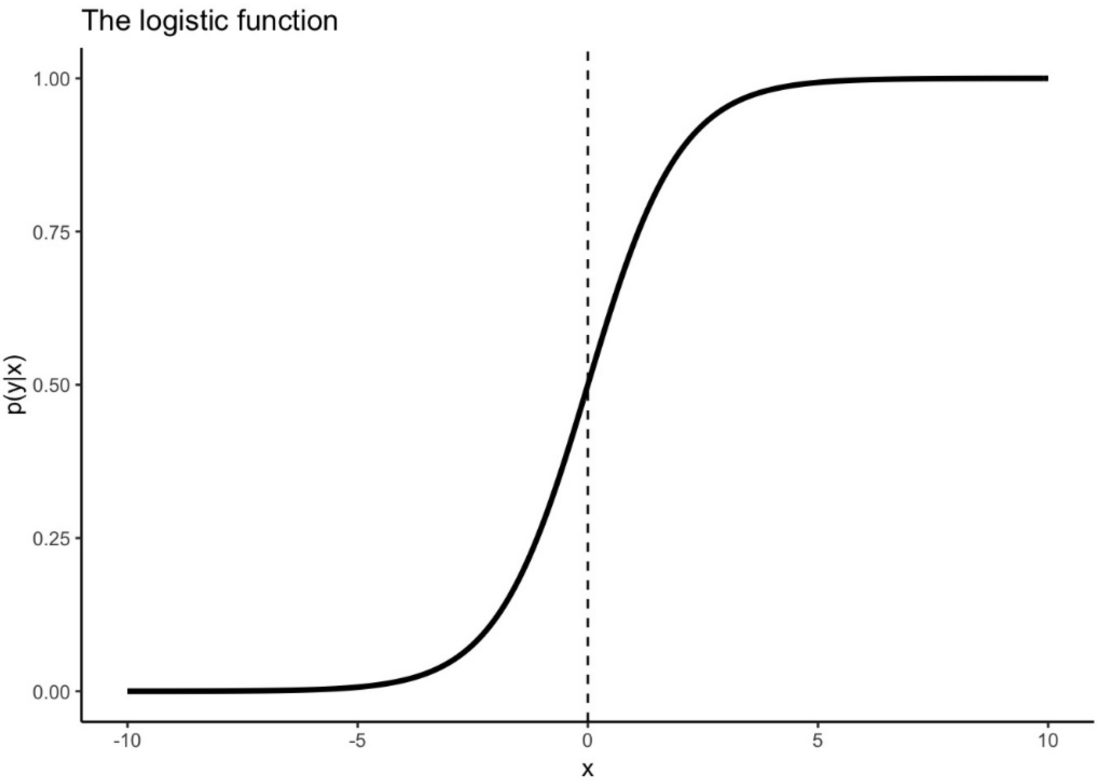

Figure 3.4: Logistic Regression

3.5.2 *Support Vector Machine*

Short for SVM, this algorithm works by finding a hyper-plane that separates a group of points in a feature-based N-dimensional space. The hyper-plane needs to maximize the margin between the class of data to future-proof the classification process. Here, support vectors are the values which fall closer to the hyper-plane and thus influencing the position and orientation the margin. Among various types, equation for linear SVM is as follows:

$$w^T * x - b = 0$$

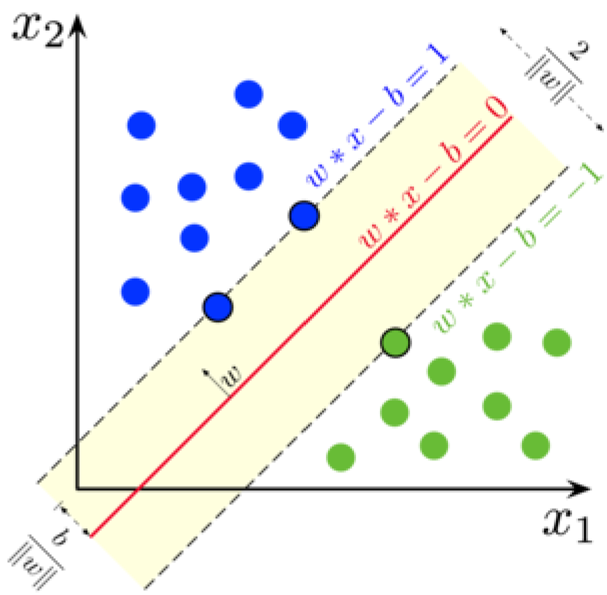

Figure 3.5: Support Vector Machine

In the Figure 3.5, the red line represents the maximum margin hyperplane which is equidistant from both classes (represented in blue and green). The dotted lines represent skewed hyperplanes which can be adjusted to favor either class.

One advantage of SVM is that it tends to prevent the over-fitting problem compared to other algorithms [34].

3.5.3 *Random Forest*

Random forest uses the decision tree concept and stretching it further by introducing a large number of individual trees. Each of these trees' outcome is given a weight and the aggregate value works as the basis to determine the predicted class using voting. The trees are need to be less correlated to each other and Bootstrap Aggregation (Bagging) is the most used way of to accomplish this aim. Formula for bagging is as follows:

$$f = \frac{1}{B}\sum_{b=1}^{B} f_b(x')$$

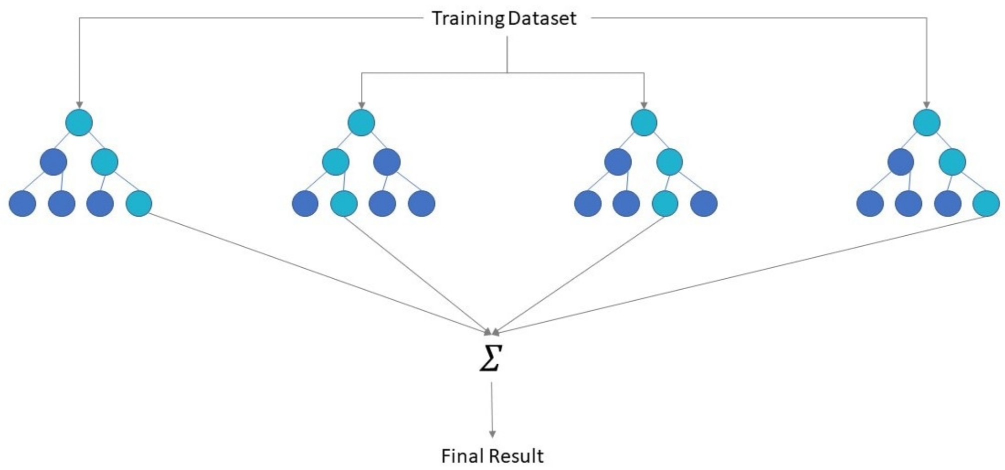

Figure 3.6: Random Forest

As an ensemble classifier, this algorithm provides better accuracy compared to other classification methods [34].

3.5.4 K-Nearest Neighbor

This algorithm works by finding the data points that are in close proximity to each other. Generally Euclidean distance is used to determine the closeness of a new point from k numbered reference points. Here, the k number defines how much stability the algorithm will preserve in accurate predictions. In a feature space, a point to be assigned to the class of closest neighbors is:

$$C_n(x) = Y_1$$

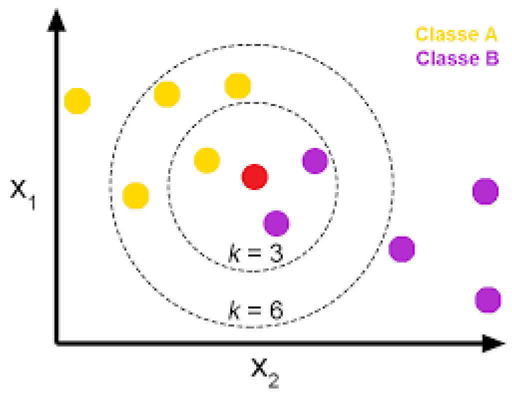

Figure 3.7: K-Nearest Neighbor

From Figure 3.7, it can be keen that for a value of k = 3, there are 2 data points in class B while only 1 in class A. As a result, the value of the red data point will be predicted to be class B. On the other hand, with k =6, there are 4 points in class A and 2 points in class B. This will result in the red point being predicted as class A.

It is relatively easier to build since there is no need for additional assumptions or tuning to be made.

3.5.5 *Naïve Bayes*

It is a classifier that is solely based on the Bayes theorem and applying probabilistic models. For an event's occurrence, how likely to occur another event is the main concept of Bayes theorem. The crucial assumption here is that the features are independent to each other for which an outcome is predicted. The classifier formula is as follows:

$$y = argmax_y p(y) \prod_{i=1}^{n} P(x_i)|y$$

For n-gram based tasks in NLP, the multinomial Naïve Bayes can be used to detect frequencies of those n-grams [34]. The classifier formula for multinomial Naïve Bayes is:

$$P(x_i|y) = \frac{(\sum_{i=1}^{n} x_i)!}{\prod_{i=1}^{n} x_i\,!} \prod_{i=1}^{n} {P_{ki}}^{xi}$$

When expressed as a log function, this classifier acts as a linear classifier.

$$logP(x_i|y) \propto \log\left(P(y) \prod_{i=1}^{n} {P_{ki}}^{xi}\right)$$

If a feature never occurs within a category in the training samples, in that case the probability estimate will be zero. This happens since the number of occurrences of a feature's value is proportional to the probability estimate. This becomes a problem as it renders the other probabilities useless. In order to avoid this issue, a small correction can be done to avoid any feature from being zero. This method is called Laplace smoothing when the count is set to a minimum of one. Another way to tackle this issue is discussed by Rennie et al. [47] where they suggested to use TF-IDF in the place of frequencies associated with raw term and document length normalization.

3.5.6 XGBoost

This is one of the ensemble algorithms which is based on decision tree and gradient boosting framework. In recent times, it is performing better than neural networks in small to medium sized dataset. Whereas previous algorithms are mathematical and conceptual, XGBoost is more of a customized and tuned version by incorporating several methods [48].

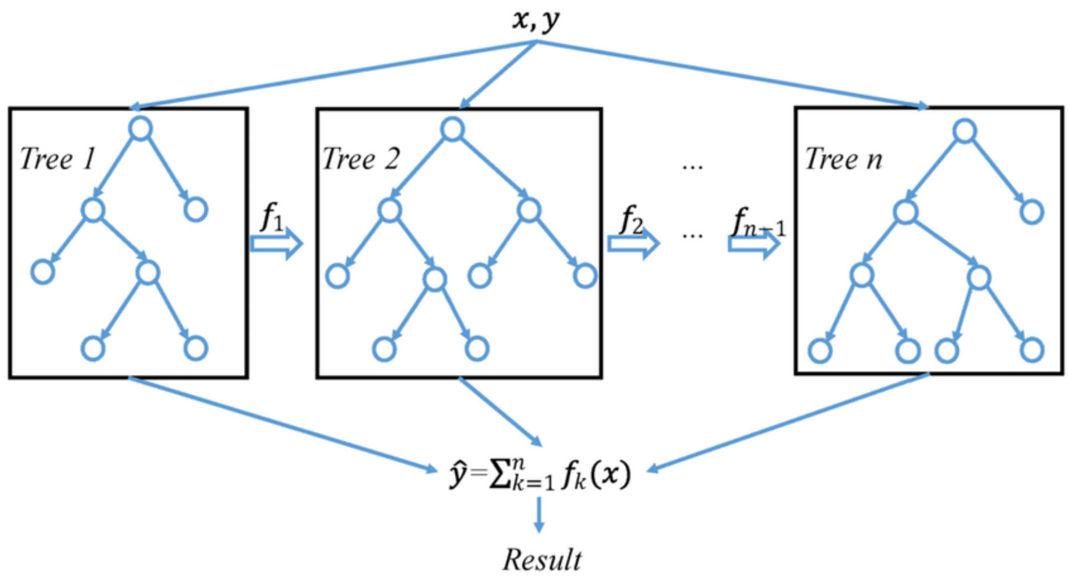

Figure 3.8: XGBoost

3.6 Evaluation Metrics

In order to evaluate the performance of the learning models on the dataset, four metrics are used in our research namely accuracy, precision, recall and f1-score. These are explained in detail.

- **Accuracy**: It is measured by the ratio of observations that are correctly predicted to the total observations. Higher accuracy does not always show the true picture since datasets tend to have imbalances in them [49].

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN}$$

- **Precision**: It indicates the ratio of correctly predicted positive observations to the total positive observations. When it is necessary to detect a True Positive (TP) with higher accuracy, the precision score should be higher [50].

$$Precision = \frac{TP}{TP + FP}$$

- **Recall**: It indicates the ratio of correctly predicted positive observations to the total observations. When it is necessary to detect a False Negative (FN) with higher accuracy, the recall value should be higher [50].

$$Recall = \frac{TP}{TP + FN}$$

- **F1-score**: It indicates the weighted average of precision and recall. This is measured to strike a balance between precision and recall scores.

$$F1 - score = 2 * \frac{Precision * Recall}{Precision + Recall}$$

3.7 Chapter Summary

A large dataset was required in order to conduct stance analysis. The dataset was extracted from twitter and individual tweets were manually labelled and cleaned before feeding the dataset to six machine learning models to train and test. The best classifier among the six was identified and used for prediction of the larger dataset. From this the stance of users were predicted and analyzed for trends and patterns.

Chapter 4

Experimental Results and Discussion

4.1 Performance of ML Models

In this section, we present the outcomes of our research. Unigrams, bigrams and trigrams were considered for feature extraction in case of stopword removal as well as leaving them in the text.

4.1.1 With Stopword Removal

Table 4.1: N-gram based Classifiers Performance with Stopword Removal

N-gram	Classifier	Accuracy	Precision	Recall	F1-score
unigram	LR	0.71038	0.71047	0.71038	0.71042
	SVC	0.70492	0.70597	0.70492	0.70515
	RF	0.70765	0.70787	0.70765	0.70770
	KNN	0.6011	0.6303	0.6011	0.5912
	XGB	0.68579	0.690165	0.68579	0.68738
	NB	0.6284	0.68263	0.6284	0.61809
bigram	LR	0.5950	0.6005	0.5950	0.5955
	SVC	0.5732	0.5777	0.5732	0.5735
	RF	0.5296	0.5685	0.5296	0.5284
	KNN	0.3022	0.0913	0.3022	0.1402
	XGB	0.5109	0.5552	0.5109	0.5067
	NB	0.5607	0.5812	0.5607	0.5561
trigram	LR	0.3988	0.4732	0.3988	0.3259
	SVC	0.4673	0.5244	0.4673	0.4309
	RF	0.3863	0.5603	0.3863	0.3482
	KNN	0.2710	0.2631	0.2710	0.1509
	XGB	0.3583	0.5839	0.3583	0.2149
	NB	0.3987	0.4632	0.3987	0.3269

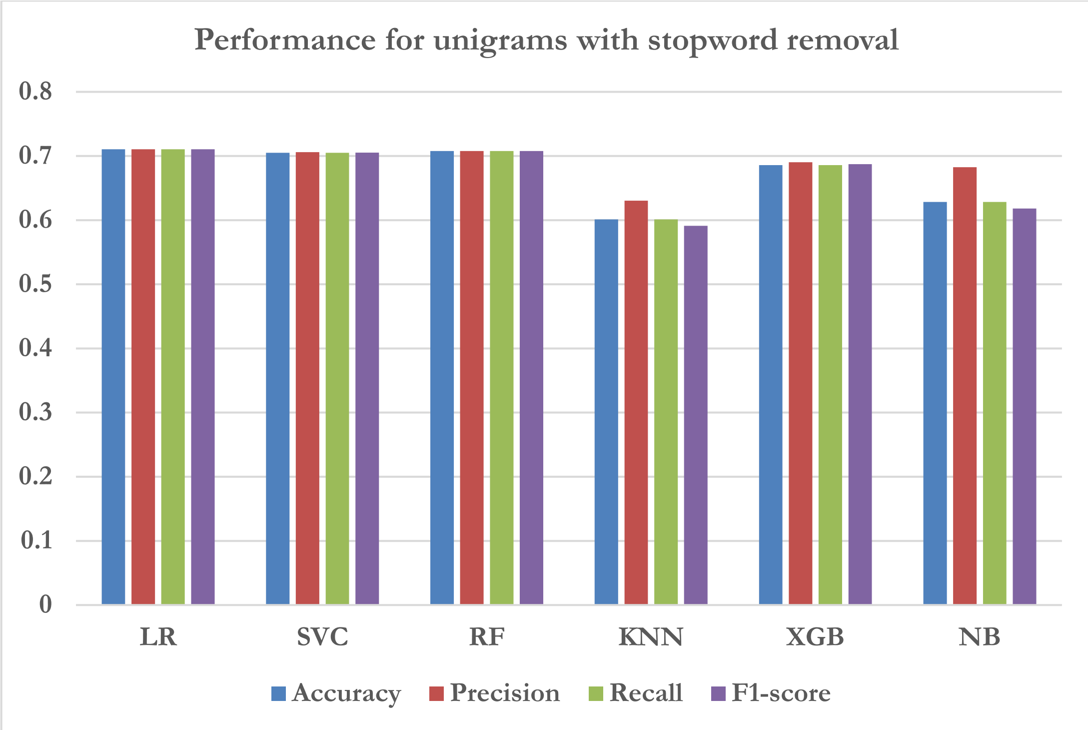

Figure 4.1: Performance for unigrams with stopword removal

Considering unigrams, from Figure 4.1 it can be seen that LR is the best performing algorithm with an accuracy of 71.038% and an F-1 score of 71.042%. SVC and RF also give comparable accuracies of 70.492% and 70.765% respectively but the other algorithms fall short with KNN being the worst performing one with an accuracy of 60.11%.

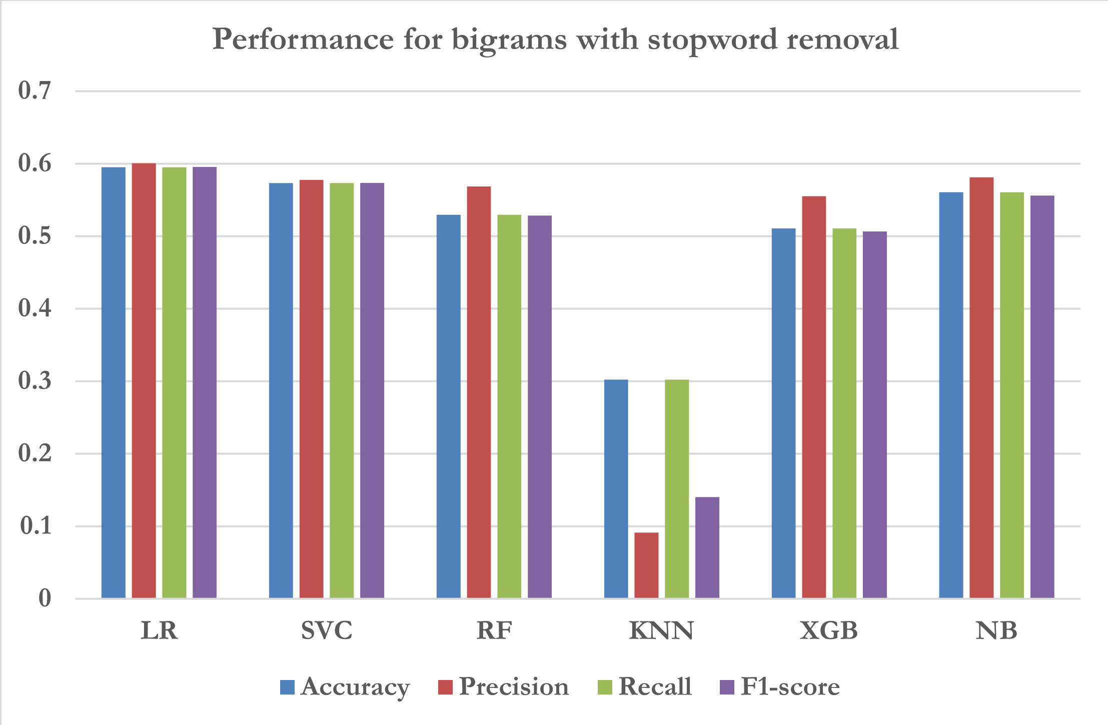

Figure 4.2: Performance for bigrams with stopword removal

From Figure 4.2 of bigrams performance, LR still performs better than the others with an accuracy of 57.32% but in this case, KNN has the worst accuracy at 30.22% and F-1 score of 14.32%. Overall, the performance for bigrams is much worse than when using unigrams for feature selection.

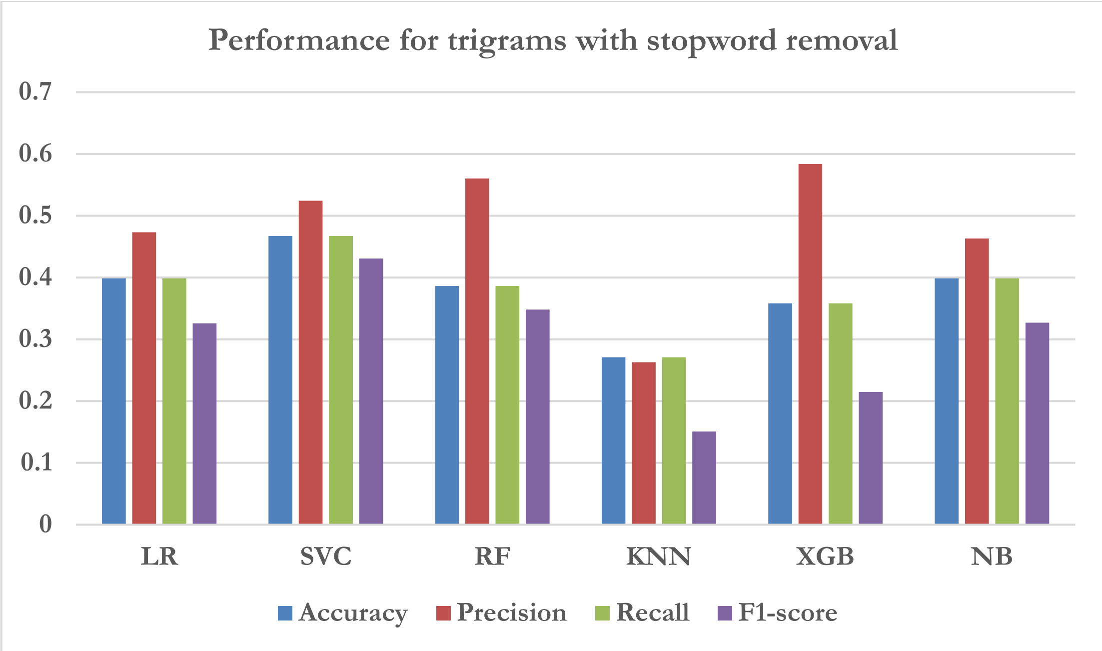

Figure 4.3: Performance for trigrams with stopword removal

Finally, in the case of trigrams, SVC performs the best with 46.73% accuracy and an F-1 score of 43.09% which can be seen in Figure 4.3. KNN still underperforms with an accuracy of 27.10%. Considering trigrams, the performance is still much worse than unigram as well as bigram feature selection.

4.1.2 Without Stopword Removal

Table 4.2: N-gram based Classifiers Performance without Stopword Removal

N-gram	Classifier	Accuracy	Precision	Recall	F1-score
unigram	LR	0.69977	0.69987	0.69977	0.69035
	SVC	0.7201	0.7189	0.7201	0.7348
	RF	0.6817	0.6877	0.6817	0.6690
	KNN	0.6704	0.6778	0.6704	0.6656
	XGB	0.6511	0.6533	0.6511	0.6512
	NB	0.6978	0.7166	0.6978	0.6937
bigram	LR	0.6293	0.6292	0.6293	0.6292
	SVC	0.6168	0.6170	0.6168	0.6161
	RF	0.6293	0.6467	0.6293	0.6287
	KNN	0.2741	0.075	0.2741	0.1180
	XGB	0.6324	0.6428	0.6324	0.6325
	NB	0.6137	0.6144	0.6137	0.6132
trigram	LR	0.4829	0.5154	0.4829	0.4814
	SVC	0.5265	0.5593	0.5265	0.5284
	RF	0.4984	0.5832	0.4984	0.4992
	KNN	0.2741	0.4115	0.2741	0.1463
	XGB	0.4393	0.5429	0.4393	0.4088
	NB	0.4829	0.5176	0.4829	0.4830

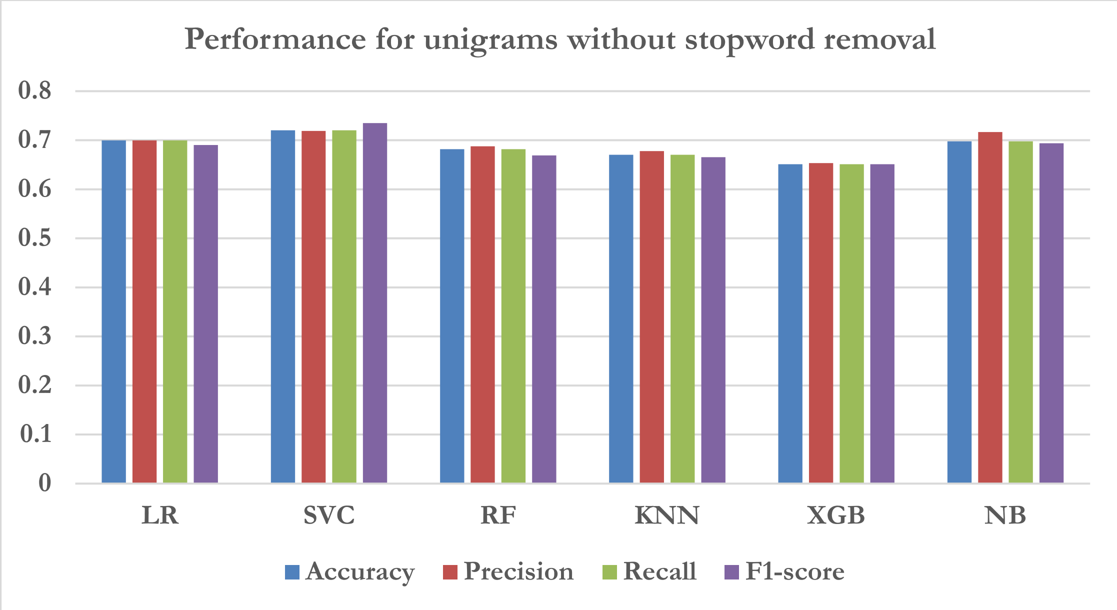

Figure 4.4: Performance for unigrams without stopword removal

The graph shown in Figure 4.4 indicates that when stopwords are not removed, for unigram feature selection, SVC has the best performance with an accuracy of 72.01% and an F-1 score of 73.48%.

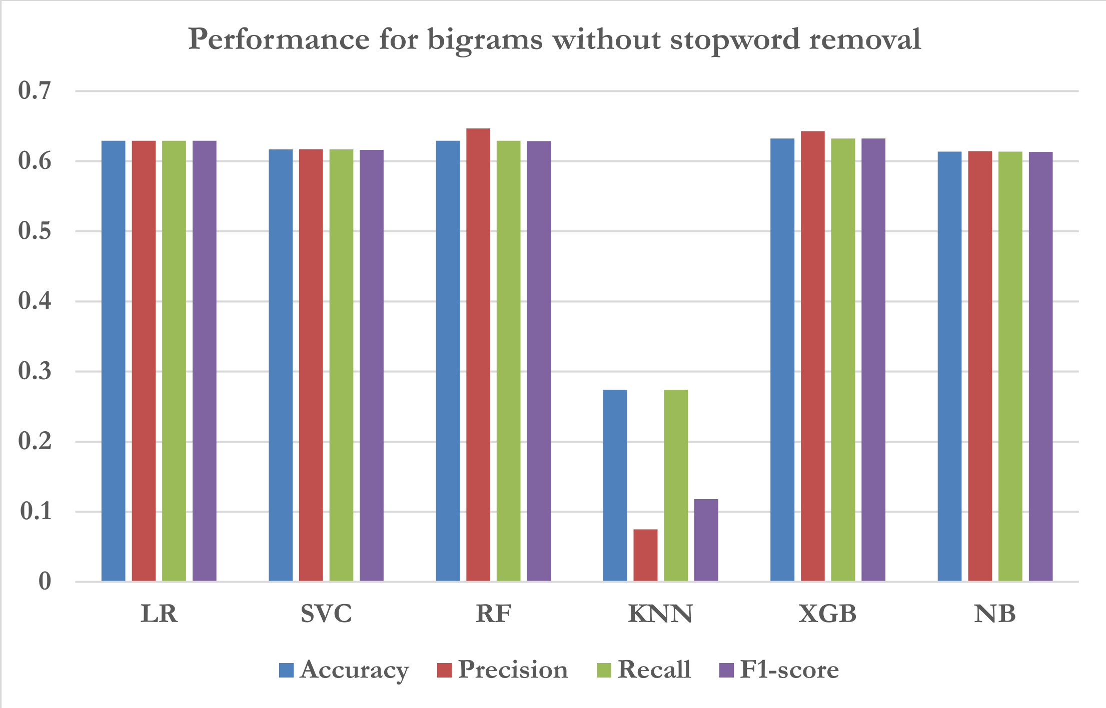

Figure 4.5: Performance for bigrams without stopword removal

Both RF and XGBoost have similar accuracies of 62.93% in the case of bigrams and it is illustrated in Figure in 4.5. KNN has the worst performance with an accuracy of 27.41%.

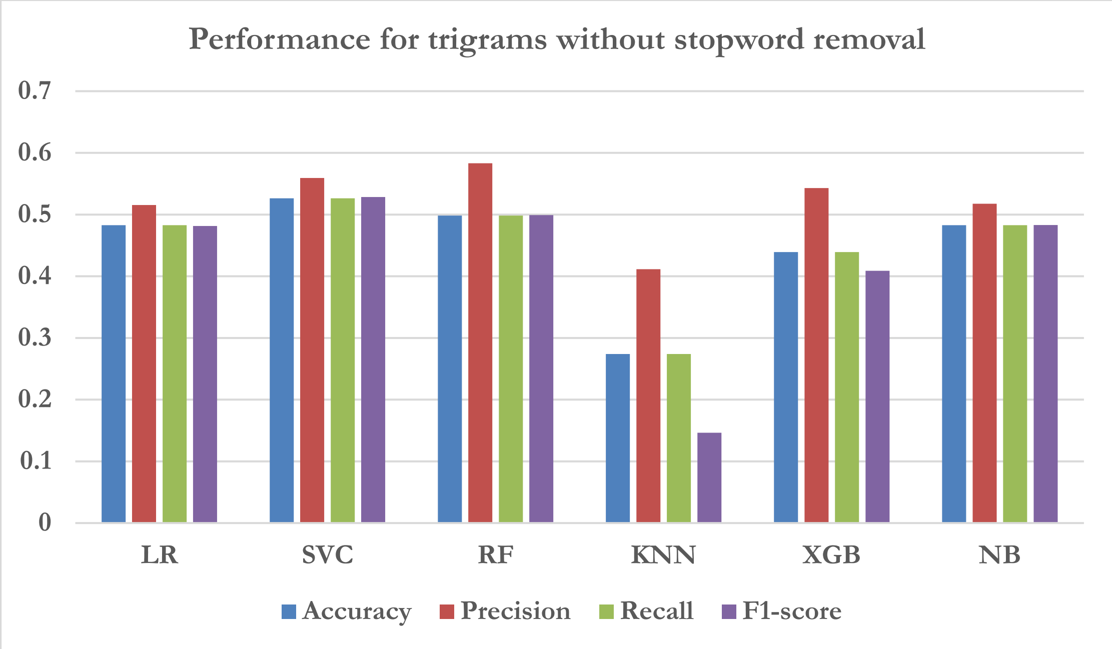

Figure 4.6: Performance for trigrams without stopword removal

SVC has the highest accuracy of 52.65% with KNN as the worst classifier with only 27.41% accuracy for trigrams shown in Figure 4.6.

4.2 Trend in Stance Analysis

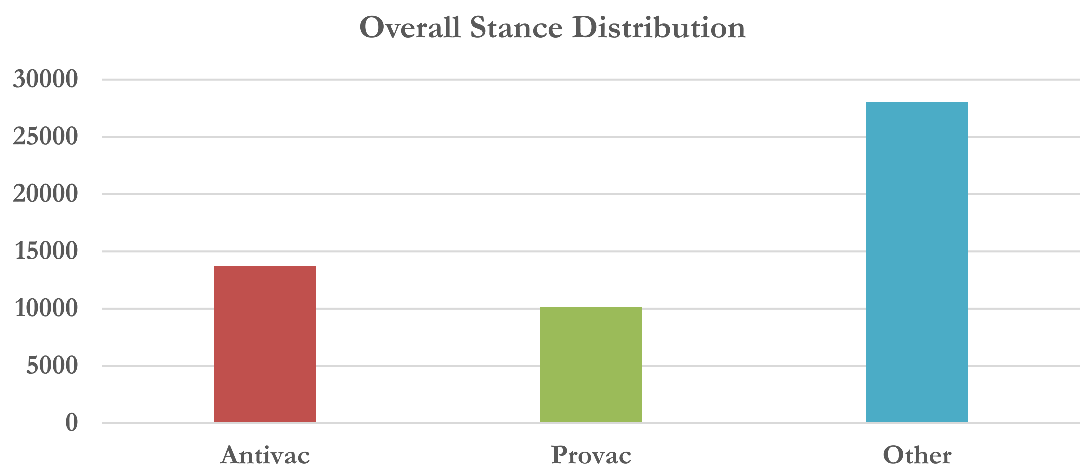

Figure 4.7: Stance Distribution in Collected Tweets

The distribution of tweets among the categories is shown in Figure 4.7. For the purpose of analysis of the data, we selected the best performing model which was support vector machine using unigram without stopword removal. The remaining 49575 tweets were thus automatically annotated by the model. We found that 13667 tweets were labelled as *antivac*, 10134 tweets were labelled as *provac* and 27989 tweets were labelled as *other*. Moreover, this result was analyzed as a frequency of tweets per day throughout the total time that we have considered. The given trend line in Figure 4.8 shows that there is not much change in the frequency of *antivac* and *provac* tweets throughout the month but there are 2 peaks for the *other* category highlighted in blue, one on 7 October, 2021 and another on 26 October, 2021. These 2 peaks indicate that there must have been some kind of event which led to increased discussions. In order to find out the cause of this increase in discussion, we evaluated the tweets on that day and cross-referenced it with world news available to us through various news articles available on that day. These news articles were primarily found through news portals available on the internet including sources such as www.cbsnews.com, www.who.int, Google News, etc.

On 7 October, the reason for the sudden increase in discussion was due to the approval of the new malaria vaccine for broader use on children by WHO. This corresponds to news headlines on that day mentioned in 'WHO recommends groundbreaking malaria vaccine for children at risk'(https://www.who.int/news/item/06-10-2021-who-recommends-groundbreaking-malaria-vaccine-for-children-at-risk). Again, on 26 October, discussion arose due the recommendation of Pfizer's vaccine for children aged 5 to 11 years old as mentioned in 'FDA advisers back Pfizer's COVID-19 vaccine for kids 5 to 11 years old - CBS News' (https://www.cbsnews.com/news/covid-19-vaccine-kids-ages-5-11-fda/).

Aside from these 2 peaks, there were minor changes in frequency in tweets belonging to *antivac* and *provac* but they did not correspond to any notable news. One noteworthy observation was that tweets indicating *antivac* stance were consistently but only slightly more than *provac* tweets.

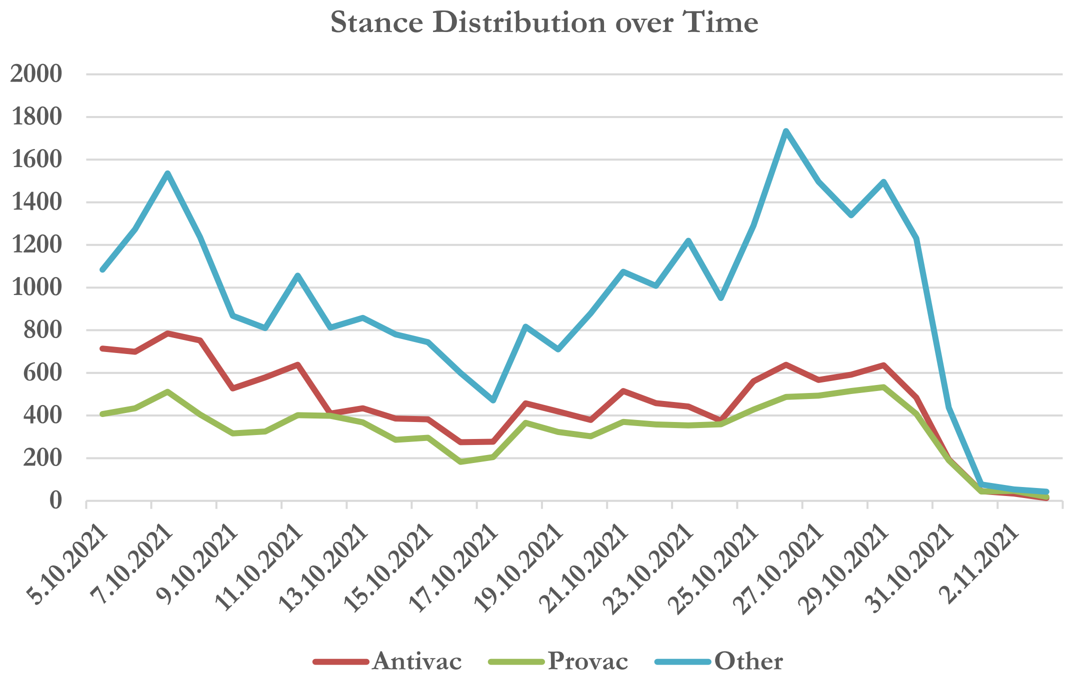

Figure 4.8: Trends in Stance

4.3 Comparison with Existing Literature

Villavicencio et al. [35] made an analysis using manual labelling technique to train the Naïve Bayes classifier where their results showed majority of the tweets fell under positive category. Their training ratio was also biased towards positive class and there was no *other* category which is the reason for giving comparatively higher accuracy. Since this research is dedicated to identify any tweet, which many times fall under *other* class, the accuracy is not as high as some other works. The *neutral* stance is a subset of *other* category in the current study.

Similar research was done by Cotfas et al [40-41] where they analyzed tweets from between 8 December, 2020 and 7 January, 2021 and found a correlation between the sudden rise in *antivac* stance on particular days due to different global events. Their research mainly focused on the *antivac* stance of the people and found that people were more inclined to be against all the newly developed vaccines. In our research, however, it can be seen that after almost a year since most of the vaccines were approved, people's debate regarding their own stance towards the vaccines did not change much. Instead, most people appeared to be discussing closely related issues which were more political in nature such as vaccine mandates, vaccine passports, international travel restrictions, vaccines for children, etc. In these discussions, people did not directly show any personal opinion about the vaccines but rather their effects on various issues.

4.4 Chapter Summary

In this chapter, the results of accuracy, precision, recall and F1-score for six machine learning models for with and without stopword removal are denoted. Unigrams, bigrams and trigrams were considered for feature extraction. From analyzing the results, the best classifier was identified to be Logistic Regression in our case and was used to predict stance on a larger number of tweets. The predicted data was analyzed further to identify trends throughout the month to find correlations between different spikes in specific stances relating to real world events and news.

Chapter 5

Conclusion and Future Work

5.1 Conclusion

There is no doubt that the effect from COVID-19 on public health and economy is catastrophic. Social media like Twitter engages people in various discussions which can broaden the view of an individual. This paper focuses on the effectiveness of machine learning algorithms when determining the stance of a person or organization towards the coronavirus vaccine from their public opinions. We manually collected the dataset in order to gain an insight on how people felt about the vaccine during a time when vaccines for the novel coronavirus have become easily accessible. We also manually labelled the dataset since the existing automated labelers only look for specific words or purely positive/negative words. Thus, it would not be able to differentiate the stance of people who spoke similarly of either opinion. Moreover, such labelers do not do a good enough job at understanding context yet. We used six pre-existing machine learning algorithms to compare their effectiveness. In future works, deep learning algorithms can be considered with an expanded dataset.

5.2 Future Work

Due to hardware limitations advanced algorithms such as BERT could not be used in this research. Algorithms such as BERT are very suitable for text-based analysis. We plan to implement such high-level algorithms in the future to further expand our work with better accuracy. Also labelling an even larger dataset would result in a better understanding of the stance towards the vaccine which would be our next goal. To further expand this research, we plan to collect data from different times in the pandemic to identify and analyze how the stance and trends change throughout the pandemic We also mean to apply similar approach to identify stance towards other vaccines and medical treatments to identify trends and patterns. This topic of research has a huge scope in predictive analysis of people's stance and with time we view to develop a proper model to conduct such analysis.

References

[1] Hadlington, Lee, Lydia J. Harkin, Daria Kuss, Kristina Newman, and Francesca C. Ryding. "Perceptions of fake news, misinformation, and disinformation amid the COVID-19 pandemic: A qualitative exploration." Psychology of Popular Media (2022).

[2] COVID-19 statistics. Available: https://covid19.who.int. [Accessed: 2022-02-02]

[3] Twitter statistics. Available: https://en.wikipedia.org/wiki/Twitter. [Accessed: 2022-02-02]

[4] Chew, Cynthia, and Gunther Eysenbach. "Pandemics in the age of Twitter: content analysis of Tweets during the 2009 H1N1 outbreak." PloS one 5, no. 11 (2010): e14118.

[5] Jones, James Holland, and Marcel Salathé. "Early assessment of anxiety and behavioral response to novel swine-origin influenza A (H1N1)." PLoS one 4, no. 12 (2009): e8032.

[6] Kim, Yunhwan, and Jang Hyun Kim. "Using photos for public health communication: A computational analysis of the Centers for Disease Control and Prevention Instagram photos and public responses." Health Informatics Journal 26, no. 3 (2020): 2159-2180.

[7] Singh, Rameshwer, Rajeshwar Singh, and Ajay Bhatia. "Sentiment analysis using Machine Learning technique to predict outbreaks and epidemics." Int. J. Adv. Sci. Res 3, no. 2 (2018): 19-24.

[8] Cambria, Erik, Dipankar Das, Sivaji Bandyopadhyay, and Antonio Feraco. "Affective computing and sentiment analysis." In A practical guide to sentiment analysis, pp. 1-10. Springer, Cham, 2017.

[9] Ten health issues WHO will tackle this year. Available: https://www.who.int/news-room/spotlight/ten-threats-to-global-health-in-2019. Accessed: 2022-04-19.

[10] Jianqiang, Zhao, and Gui Xiaolin. "Comparison research on text pre-processing methods on twitter sentiment analysis." IEEE Access 5 (2017): 2870-2879.

[11] Gupta, Itisha, and Nisheeth Joshi. "Enhanced twitter sentiment analysis using hybrid approach and by accounting local contextual semantic." Journal of intelligent systems 29, no. 1 (2020): 1611-1625.

[12] Kouloumpis, Efthymios, Theresa Wilson, and Johanna Moore. "Twitter sentiment analysis: The good the bad and the omg!." In Proceedings of the international AAAI conference on web and social media, vol. 5, no. 1, pp. 538-541. 2011.

[13] Xuan, Kaizhou, and Rui Xia. "Rumor stance classification via machine learning with text, user and propagation features." In 2019 International Conference on Data Mining Workshops (ICDMW), pp. 560-566. IEEE, 2019.

[14] Saif, Hassan, Miriam Fernandez, Yulan He, and Harith Alani. "Evaluation datasets for Twitter sentiment analysis: a survey and a new dataset, the STS-Gold." (2013).

[15] Rice, Douglas R., and Christopher Zorn. "Corpus-based dictionaries for sentiment analysis of specialized vocabularies." Political Science Research and Methods 9, no. 1 (2021): 20-35.

[16] Bandhakavi, Anil, Nirmalie Wiratunga, and Stewart Massie. "Emotion-aware polarity lexicons for Twitter sentiment analysis." Expert Systems 38, no. 7 (2021): e12332.

[17] Mubarak, Hamdy, Sabit Hassan, Shammur Absar Chowdhury, and Firoj Alam. "ArCovidVac: Analyzing Arabic Tweets About COVID-19 Vaccination." arXiv preprint arXiv:2201.06496 (2022).

[18] Dimitrov, Dimitar, Erdal Baran, Pavlos Fafalios, Ran Yu, Xiaofei Zhu, Matthäus Zloch, and Stefan Dietze. "Tweetscov19-a knowledge base of semantically annotated tweets about the covid-19 pandemic." In Proceedings of the 29th ACM International Conference on Information & Knowledge Management, pp. 2991-2998. 2020.

[19] Hayawi, Kadhim, Sakib Shahriar, Mohamed Adel Serhani, Ikbal Taleb, and Sujith Samuel Mathew. "ANTi-Vax: a novel Twitter dataset for COVID-19 vaccine misinformation detection." Public health 203 (2022): 23-30.

[20] Xue, Jia, Junxiang Chen, Chen Chen, Chengda Zheng, Sijia Li, and Tingshao Zhu. "Public discourse and sentiment during the COVID 19 pandemic: Using Latent Dirichlet Allocation for topic modeling on Twitter." PloS one 15, no. 9 (2020): e0239441.

[21] Boon-Itt, Sakun, and Yukolpat Skunkan. "Public perception of the COVID-19 pandemic on Twitter: sentiment analysis and topic modeling study." JMIR Public Health and Surveillance 6, no. 4 (2020): e21978.

[22] Dubey, Akash Dutt. "Twitter sentiment analysis during COVID-19 outbreak." Available at SSRN 3572023 (2020).

[23] Cotfas, Liviu-Adrian, Camelia Delcea, Rareș Gherai, and Ioan Roxin. "Unmasking People's Opinions behind Mask-Wearing during COVID-19 Pandemic—A Twitter Stance Analysis." Symmetry 13, no. 11 (2021): 1995.

[24] Manguri, Kamaran H., Rebaz N. Ramadhan, and Pshko R. Mohammed Amin. "Twitter sentiment analysis on worldwide COVID-19 outbreaks." Kurdistan Journal of Applied Research (2020): 54-65.

[25] Nemes, László, and Attila Kiss. "Social media sentiment analysis based on COVID-19." Journal of Information and Telecommunication 5, no. 1 (2021): 1-15.

[26] Saleh, Sameh N., Christoph U. Lehmann, Samuel A. McDonald, Mujeeb A. Basit, and Richard J. Medford. "Understanding public perception of coronavirus disease 2019 (COVID-19) social distancing on Twitter." Infection Control & Hospital Epidemiology 42, no. 2 (2021): 131-138.

[27] Machuca, Cristian R., Cristian Gallardo, and Renato M. Toasa. "Twitter sentiment analysis on coronavirus: Machine learning approach." In Journal of Physics: Conference Series, vol. 1828, no. 1, p. 012104. IOP Publishing, 2021.

[28] Yadav, Nikhil, Omkar Kudale, Aditi Rao, Srishti Gupta, and Ajitkumar Shitole. "Twitter sentiment analysis using supervised machine learning." In Intelligent Data Communication Technologies and Internet of Things, pp. 631-642. Springer, Singapore, 2021.

[29] Imran, Ali Shariq, Sher Muhammad Daudpota, Zenun Kastrati, and Rakhi Batra. "Cross-cultural polarity and emotion detection using sentiment analysis and deep learning on COVID-19 related tweets." Ieee Access 8 (2020): 181074-181090.

[30] Behera, Ranjan Kumar, Monalisa Jena, Santanu Kumar Rath, and Sanjay Misra. "Co-LSTM: Convolutional LSTM model for sentiment analysis in social big data." Information Processing & Management 58, no. 1 (2021): 102435.

[31] Sireesha, Kusumanchi Naga, and Padala Srinivasa Reddy. "COVID19 Sentiment Analysis using Machine Learning Classification Algorithms." (2021).

[32] Alenezi, Mohammed N., and Zainab M. Alqenaei. "Machine learning in detecting COVID-19 misinformation on twitter." Future Internet 13, no. 10 (2021): 244.

[33] To, Quyen G., Kien G. To, Van-Anh N. Huynh, Nhung TQ Nguyen, Diep TN Ngo, Stephanie J. Alley, Anh NQ Tran et al. "Applying machine learning to identify

anti-vaccination tweets during the COVID-19 pandemic." International journal of environmental research and public health 18, no. 8 (2021): 4069.

[34] Müller, Martin, and Marcel Salathé. "Addressing machine learning concept drift reveals declining vaccine sentiment during the COVID-19 pandemic." arXiv preprint arXiv:2012.02197 (2020).

[35] Villavicencio, Charlyn, Julio Jerison Macrohon, X. Alphonse Inbaraj, Jyh-Horng Jeng, and Jer-Guang Hsieh. "Twitter sentiment analysis towards covid-19 vaccines in the Philippines using naïve bayes." Information 12, no. 5 (2021): 204.

[36] Ebeling, Régis, Carlos Abel Córdova Sáenz, Jeferson Nobre, and Karin Becker. "Analysis of the influence of political polarization in the vaccination stance: the Brazilian COVID-19 scenario." arXiv preprint arXiv:2110.03382 (2021).

[37] Bi, Datian, Jingyuan Kong, Xue Zhang, and Junli Yang. "Analysis on health information acquisition of social network users by opinion mining: case analysis based on the discussion on COVID-19 vaccinations." Journal of Healthcare Engineering 2021 (2021).

[38] Di Giovanni, Marco, Lorenzo Corti, Silvio Pavanetto, Francesco Pierri, Andrea Tocchetti, and Marco Giovanni Brambilla. "A Content-based Approach for the Analysis and Classification of Vaccine-related Stances on Twitter: the Italian Scenario." In Information Credibility and Alternative Realities in Troubled Democracies@ ICWSM 2021, pp. 1-6. 2021.

[39] Sang, E. Tjong Kim, Marijn Schraagen, Shihan Wang, and Mehdi Dastani. "Transfer Learning for Stance Analysis in COVID-19 Tweets." CLIN31: Computational Linguistics in The Netherlands (2021).

[40] Cotfas, Liviu-Adrian, Camelia Delcea, and Rareș Gherai. "COVID-19 vaccine hesitancy in the month following the start of the vaccination process." International Journal of Environmental Research and Public Health 18, no. 19 (2021): 10438.

[41] Cotfas, Liviu-Adrian, Camelia Delcea, Ioan Roxin, Corina Ioanăş, Dana Simona Gherai, and Federico Tajariol. "The longest month: Analyzing covid-19 vaccination opinions dynamics from tweets in the month following the first vaccine announcement." IEEE Access 9 (2021): 33203-33223.

[42] Bozkır, A. Selman, S. Güzin Mazman, and Ebru Akçapınar Sezer. "Identification of user patterns in social networks by data mining techniques: Facebook case." In International symposium on information management in a changing world, pp. 145-153. Springer, Berlin, Heidelberg, 2010.

[43] Lin, Jimmy, and Dmitriy Ryaboy. "Scaling big data mining infrastructure: the twitter experience." Acm SIGKDD Explorations Newsletter 14, no. 2 (2013): 6-19.

[44] Géron, A., “Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow: Concepts, tools, and techniques to build intelligent systems,” O'Reilly Media, 2019.

[45] Pedregosa, Fabian, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel et al. "Scikit-learn: Machine learning in Python." the Journal of machine Learning research 12 (2011): 2825-2830.

[46] Hosmer Jr, David W., Stanley Lemeshow, and Rodney X. Sturdivant. Applied logistic regression. Vol. 398. John Wiley & Sons, 2013.

[47] Rennie J, Shih L, Teevan J, Karger D (2003). “Tackling the poor assumptions of naïve Bayes classifiers.” ICML

[48] Chen, Tianqi, and Carlos Guestrin. "Xgboost: A scalable tree boosting system." In Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining, pp. 785-794. 2016.

[49] Hossin, Mohammad, and Md Nasir Sulaiman. "A review on evaluation metrics for data classification evaluations." International journal of data mining & knowledge management process 5, no. 2 (2015): 1.

[50] Gunawardana, Asela, and Guy Shani. "A survey of accuracy evaluation metrics of recommendation tasks." Journal of Machine Learning Research 10, no. 12 (2009).