Sentiment Analysis on Covid-19 Vaccination Reviews Using BERT and Comparative Study with LSTM, Vader, and Text blob Models - Research Report

MSc Research Project
Data Analytics

Sourav Ramalingam
Student ID: x20199911

School of Computing
National College of Ireland

Supervisor: Mr. Taimur Hafeez

National College of Ireland

MSc Project Submission Sheet

School of Computing

Student Name:	MR. Sourav Ramalingam		
Student ID:	x20199911		
Programme:	Data Analytics	**Year:**	2022
Module:	MSc Research Project……………………………………………………………………….		
Supervisor:	Taimur Hafeez…………………………………………………………………………………….		
Submission Due Date:	15/12/2022…………………………………………………………………………………………….		
Project Title:	Sentiment Analysis on Covid-19 Vaccination Reviews Using BERT and Comparative Study with LSTM, Vader, and Text blob Models.		
Word Count:	…………8107……………… **Page Count**…………………23……………………………..……..		

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.
ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature: Sourav Ramalingam……………………………………………………………………………

Date: 15/12/2022……………………………………………………………………………………………………

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST

Attach a completed copy of this sheet to each project (including multiple copies)	□
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	□
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	□

Assignments that are submitted to the Programme Coordinator Office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Sentiment Analysis on Covid-19 Vaccination Reviews Using BERT and Comparative Study with LSTM, Vader, and Text blob Models

Sourav Ramalingam
Student ID: x20199911

Table of Contents

Table of Figures

Abstract

The pandemic of 2019 is known to everyone, as it significantly impacted everyone's life. The pandemic took a toll on the economic as well as the mental situation of the people. In such a situation, it became difficult for the government in meeting everyone's needs. Social media paved the way for the government to analyse the situation based on the people's opinion and their mindset towards the pandemic. Once the vaccination was available in the market for public usage, it was also noticed that people had hesitations in getting vaccinated as many rumours were spreading around by word of mouth and through the reviews that were posted on social media platforms. As per the research, the Covid-19 vaccination proved to be effective in fighting against the virus. Anyways, the public had different opinions about the effects of vaccination. So, it has also become important for the government to analyse the people's opinions about the different vaccination, which could be done using the sentiment analysis approach applied to the data generated from social media platforms.

Keywords: BERT - Bidirectional encoder representation from transformers, LSTM - Long short-term memory, NLP - Natural language processing, Pre-trained Sentiment Analysis Models, Opinion Mining.

1 Introduction

The research that I have developed is all about finding the sentiment of the COVID-19 Vaccination review text using the different algorithms and models as the collective results from all the models would help to take finite sentiment results. Usually, the sentiment results could be classified [11] as positive, negative, neutral, sarcasm, etc. Anyways, for this research, the classifications are limited to positive, negative, and neutral. The primary model that is developed out of this research is using the Bidirectional Encoder Representations from Transformers (BERT) algorithm and other the models like LSTM, pre-trained models [12] are used for comparison so that the optimal one could be chosen for the further enhancements in the future phases. Alternatively, if all the selected models are working in the best way, then extracting an overall result by assigning a weightage for the results of the chosen models would result in a fusion model.

1.1 Background And Motivation:

The background of the research work is about the reviews that flooded the internet during and after the pandemic. Before the invention of the vaccine, the mindset of the people was different, as they thought the vaccine would help them from recovering or not getting the disease at all. After the vaccine got invented there existed other problems such as a shortage of vaccines, effects of the vaccine, types of vaccines, and deaths due to the vaccine [13]. The reviews in both the online and offline methods need to be analysed. The analysis needs to be done as there were several companies producing different vaccines, and it was important to identify the effects of all the vaccines that are available for public use. People had a lot to talk about the pandemic, ongoing economic crisis, mental health, awareness, and vaccine. The people's opinions mostly could be massively gathered from social media platforms. The reviews or the views were about everything related to the situation. These reviews could help the government to understand the impact of the situation and to eliminate the vaccines that cause real side effects. The sentimental analysis could be done on the reviews of the vaccination because it showcases the people’s feelings towards the pandemic and its outcome.

1.2 Research Question:

The research question mentioned below helps to keep the focus towards the technical aspects of the proposed research.

1. How far the Bidirectional Encoder Representations from Transformers (BERT) approach would be optimal for the sentiment analysis implementation?

2. Does this research helps to extract the results of the comparison between the custom-trained models like BERT and LSTM with the pre-trained models like Vader and Text Blob?

1.3 Aim Of the Research:

The main aim of the chosen research is to analyse the reviews of vaccination as it induced different feelings in different people. In technical aspects, the aim is to build a sentiment analysis system which would be efficient for finding the sentiment score, specific to the COVID-19 vaccination data. During the initial stage of vaccination there was reported no death due to the vaccine but as the days goes by there existed death due to vaccination. The deaths due to vaccinations are believed to be rumours. While some people are affected by that rumour. Such people fear getting the vaccination done. The government with the help of these reviews analyse their mind and find ways in creating awareness among people. To eradicate this very thought in people's minds the government conducted mass vaccination camps to indulge many people in vaccination. These mass vaccination camps proved to be a success in different parts of the world. The government could further target a large set of audience using technical advancements like deep learning techniques to ensure most of the public is getting vaccinated.

1.4 Scope Of Work:

The scope of the work is to develop a model that accepts the COVID-19 vaccination-related text and to identify whether the text is positive, negative, or neutral. In terms of training, the scope is to select the dataset which was already processed and available online so that I could make use of it for the model training. The scope could be wider to scrap the data from online and use it for model training, anyways due to the limited time available for the model development, the scope is limited to use that dataset that is readily available. In terms of models, developing a minimum viable product (MVP) version is set as the scope, as the MVP version would've limited features when compared to the production version which usually has all the required features implemented.

1.5 Objective and Hypothesis:

The main objective of the research work is to create models that could classify the COVID-19 vaccination reviews into positive, negative, and neutral. The objective is also set to prove that RoBERTa, a BERT-based model is performing well when compared to the other models. To meet the objective, the plan is set to create multiple models so that the developed models could be compared and the best one would be chosen based on the evaluation results. Identify the related works that were done in the related area, so that the steps in the approach could be more finite and would help to achieve the intended research development.

Hypothesis: Choosing the BERT-based pre-trained algorithm for the sentiment analysis development would be more efficient when compared to deep learning models like LSTM, and sentiment analysis libraries like Vader, and Text Blob as the BERT model allows the addition of context-specific data layers.

1.6 Limitations:

The research work has a few limitations as I have set the scope to develop only the MVP version of the models. The primary limitation is the volume of data that is used for training the models. In our research, a couple of pre-trained models - Vader and Text Blob are used, which wouldn't get impacted by the limitation in the dataset size as it was already trained with the huge volume of data. Anyways, when it comes to the BERT and LSTM model, having a limited dataset would impact the accuracy of the model as it has limited knowledge about the domain data that it is trained with. The reason for having the limited dataset is due to the limited hardware and software configuration that is used for model training. The limitations would be overcome in future phases.

1.7 Stakeholders of the Research Work:

The stakeholders are the people who would be using the application for them to make the business decision. This research might have more than one stakeholder group. For example, let's take the European Medicines Agency (EMA) [14] which is an organization that closely watches the drugs that are available to the public across Europe. The drug usually follows a different life cycle like the development phase, testing phase, approval for public usage phase and then the monitoring phase, in which the side effects of the drugs would be monitored as there is a high possibility of having the side effects after using it over a period. So here the EMA agency would use the sentiment analysis model that I have built to keep track of people's opinions about any drug that is in the monitoring phase. On the other hand, even we have agencies like Eudora which is deployed to observe the adverse drug effect which is one of the main reasons for medical deaths. The government could also use the developed model for understanding the people's opinion towards getting vaccinated so that the campaign [15] content could be targeted to answer the people's questions about the vaccination effects.

2 Related Work

The background was done in different areas like opinion mining/sentiment analysis, technical fields like Natural language processing, implementing deep learning modelling techniques and so on so that I could identify if any gaps exist in the solutions that were already built, the approach that needs to be followed for the research development so that I would identify the purpose of the research that is developed out of this research.

2.1 Opinion Mining

The term opinion mining is all about extracting the opinion of people using the techniques like analysing social networks [1]. The traditional way of performing the sentiment analysis is just analysing the textual data, anyways in the modern approach opinion mining could be done from different data types like images, videos, expressions, etc. Different algorithms were identified as a suitable one for the implementation of the sentiment analysis concept. In the modern era, big data takes a major role as it could be able to process the huge volume of data that is generated from social media platforms. The concept of opinion mining could be used in a wide range of applications like predicting the stock market price, forecasting the winning political party, movie reviews, etc. As per the research [2], deep learning has shown a great result in performing sentiment analysing using the concept of artificial intelligence.

Sentiment analysis follows different approaches [3, 4] to classify the social media text. Some of the approaches are Machine Learning Approach, Hybrid Approach, Lexicon-Based approach, Graph-based approach and so on. On the other hand, text-based sentiment analysis could be applied at 2 different levels. The first one is the sentence level and the other one is the entity level. The primary limitation in performing the text-based sentiment analysis is the limitation in the text size also the short forms that are used in review texts. The lexicon-based approach is a widely used approach to find the polarity score of the textual message. The primary advantage of the lexicon approach is it doesn't require the model training phase [5]. Different platforms were created using the lexicon-based approach and it is also deployed to be consumed by social media platforms for getting the polarity score for the texts that are posted on the platform.

To Summarize, the research work that was done on the opinion mining/sentiment analysis area was reviewed so that I could understand all the possible options that are available in the opinion mining area which would help our implementation. The different approaches that could be followed in performing the sentiment analysis are discussed. The research papers also help to identify the problems that could raise in opinion mining. For example, the text length not being enough for finding the sentiment polarity helps to anticipate the situation when the data is pre-processed and taken into the model for the training process.

2.2 Natural Language Processing

Natural Language Processing (NLP) is the research area which enables the computer to understand human language [6]. The primary aim of NLP programming is to create a system that understands human language exactly how we understand the language. In this way, computers could be enabled to perform actions based on user input in the same language that human beings speak. For example, speech recognition software works based on NLP programming. Several research also took place for multilingual processing using NLP programming. Some of the problems that we face in natural language processing are the thought process of the human being, next is the representation of the text and the third one is the knowledge that the world would keep on updating. To handle all these problems, the NLP system would need to be progressed by adapting to enhanced approaches. For example, the initial NLP version analyses the text at the word level using the morphological pattern. Anyways, later it is moved to the sentence level analysis to understand the whole context including the grammar. The current NLP libraries work using the knowledge-based approach.

The pre-trained NLP based model BERT has been used for building the sentiment analysis tool which works on aspect-based [7]. The research aimed to develop a prediction code that predicts sentiment polarity. The major problem in the aspect-based approach is performing the sequence tagging, whereas in a few other approaches we face the classification program. The BERT pre-trained model is compared to word embedding layers which have specific tasks assigned to them. The usual problem in preparing sentiment analysis models is not having a sufficient dataset to build a sophisticated model. So to overcome this problem, new layers were introduced which hold the context-specific word embeddings that work with the pre-trained layers. Instead of building a task-specific network, finetuning the pre-trained network with the task-specific information would potentially help us to improve the performance of the model.

To Summarize the background study of the natural language processing topic, the base idea of Natural language processing is discussed along with the evolutions that happened in NLP, which helps to gain a clear picture of the problems that NLP could in making it work. Discussions were also made on the BERT pre-trained technique that uses the NLP logic as the base for performing the tasks like sentiment analysis, answering user questions based on the data trained and other text-based processing. From the study, instead of developing a sentiment analysis model from scratch using the training data is less efficient when compared to the approach of using pre-trained models like BERT and adding additional data layers would be more efficient and provide room for enhancements in the model.

2.3 Deep Learning Approach

The deep learning approach has developed a lot in applying the sentiment analysis technique with the progress in big data analytics and NLP programming [8]. In developing the sentiment analysis model that works on cryptocurrency data, the implementation of long short-term memory (LSTM) has shown a phenomenal result. With the results obtained, it could be stated that recurrent neural network programming works best for solving sentiment analysis problems. The research also states that the deep learning algorithms outperformed the machine learning-based approach algorithms like the Support vector machine. The fully connected layers in LSTM, the activation functions to output the predictions and the embedding feature vector components are used to generate the sentiment polarity. Deep neural algorithms have shown great achievements in generating successful results in sentiment analysis. The LSTM networks could be excellent in performing the learning operation from the dataset that is provided for model training as the network layers in the LSTM model works like a human brain. Anyways, the LSTM approach has limitations in terms of the model getting trained with common sense facts to perform the intended task. Aspect-based sentiment analysis is suggested in the search [26] by using the deep neural sequential model. The extension of the LSTM model is proposed to have a hybrid version of the sentiment analysis model which is called Sentic-LSTM. The evaluation was done on the developed hybrid model and the results were outperforming when compared to the model just built using the LSTM algorithm.

To summarize, I have identified that deep learning algorithms also could be considered for performing sentiment analysis. Business requirements could be developed using multiple algorithms so that the results could be compared and select the best-fit model for the production environment. To perform the same in this research, the deep learning model LSTM is chosen as a potential candidate for developing the COVID-19 vaccination sentiment analysis and comparing it with the BERT model that was identified previously. Some other research also has shown tremendous performance even with the LSTM model but with adding more layers to have a hybrid version of the LSTM model. The more layers that are added to handle the context-specific data help in generating the LSTM model with could perform better.

2.4 Sentiment Analysis Tools

The Valence Aware Dictionary for sEntiment Reasoner (VADER) is developed by Gilbert and works on the base of rules to perform sentiment analysis. The VADER approach has shown benefits when compared to the other traditional approaches like LIWC [21], which is based on Lexicons [9]. The Vader model is aware of the expressions that are posted on social media platforms, and on the other hand, the model is generalized to handle cross-domain data. Even though the VADER is a rule-based analyser, it used the lexicons when providing the sentiment score for the social media review text. The analyser could classify the review text into positive, negative, or neutral scores along with the compound value, which states the overall sentiment result. The tool internally uses the NLTK [24] package. Text Blob is developed using python coding and it has a basic API structure to access the endpoint which accepts the input value and provides the sentiment result after performing the required NLP tasks. It has a few limitations in using the Text Blob as it just provides only the subjectivity and polarity of the text [10]. On top of that, the text blob is not designed the process the emojis which could also be used as a source for finding the sentiment score. Also, it has limitations in analysing emotions.

To conclude, analysing the pre-trained model tools in the area of sentiment analysis helps to understand the need for a domain-specific model that is trained with specialized data.

Literature Review Papers	Summary Of the Paper
A survey on sentiment analysis and opinion mining for social multimedia (2018)	The idea of extracting people's opinions and attitudes from social media is mentioned in this research. It also highlights the conventional way of applying sentiment analysis with the modern approach. The comparison helps to identify the enhancements required in the sentiment analysis field.
Deep learning for sentiment analysis: successful approaches and future challenges (2015)	This research explains how far deep learning programming is emerging in the field of applying sentiment analysis. It is stated that the state-of-art is tremendously improved in performing the intended task in sentiment analysis. The challenges in extracting the essential features are also highlighted in the research.
Like It or Not: A Survey of Twitter Sentiment Analysis Methods (2017)	This research highlights the importance of using Twitter data for performing the analysis as people large share their opinion on the platforms like Twitter.
A Survey and Comparative Study of Tweet Sentiment Analysis via Semi-Supervised Learning (2016)	This research explains the semi-supervised learning approach in performing the sentiment analysis which uses unlabelled data instead of labelled one for the training process, in which the data is extracted from Twitter.
Lexicon-Based Methods for Sentiment Analysis (2011)	This research highlights using the Lexicon-based approach for performing the sentiment analysis in which orientations are calculated based on the whole document or set of sentences considering the lexicons.
Natural Language Processing (2020)	This research highlights the importance of computers having knowledge of human languages like English and the applications in which NLP is more useful.
LSTM Based Sentiment Analysis for Cryptocurrency Prediction (2021)	As per the research, The LSTM model which is a recurrent neural network has shown a lot of progress in sentiment analysis with related fields like big data analytics and Natural Language Processing code module.
Twitter Sentiment Analysis Using Natural Language Toolkit and VADER Sentiment (2019)	In this research, it is mentioned using the sentiment analysis toolkit VADER in twitter data to automatically perform the sentiment polarity classification along with the toolkit advantages.
Exploiting BERT for End-to-End Aspect-based Sentiment Analysis (2019)	This research highlights the BERT-based architecture for developing the sentiment analysis solution, which is the pre-trained language model that allows the embedding of contextual layers. It is stated that with the BERT approach, the model could outperform state-of-the-art. As per the research, the aspect-based approach is followed in performing the intended task.

Table 01 – Summary of Literature Review

3 Research Methodology

The Cross-Industry Standard Process for Data Mining (CRISP-DM) [16] Methodology has been followed for the implementation of the research work. The chosen methodology has different phases which suit well for the chosen research work. In the business understanding phase, the business need of analysing the COVID-19 vaccination was formulated. In the data understanding and gathering phase, the source of data where the COVID-19 vaccination could be extracted has been decided. The social media platforms like Twitter were chosen as source platforms where user reviews about the vaccination could be extracted. In the modelling phase, the actual implementation takes place. For this research, I have chosen algorithms and pre-trained models of different types to perform the sentiment analysis as it is not ideal approach to choose a single approach without making any comparison and stating it is performing well. In the evaluation phase, the chosen algorithm results and pre-trained model results are compared to identify which one works best. At last, the methodology has the deployment phase, in which the developed models would be deployed into a server for making it accessible. Anyways, due to the academic nature of the research work, the deployment phase is eliminated.

Once the data is pre-processed, the chosen models would be implemented with the processed data. Implementation of the model is done by feeding the processed data into the application. In this topic, the following model implementations are discussed: BERT - bidirectional encoder representation from transformers, LSTM - Long short-term memory, Text Blob, and Vader. In implementing the models, I have performed the below tasks:

- As a first step, I have chosen a dataset which has the COVID-19 vaccination-related tweets that were gathered from a different location and have the required fields like a rating column that is used to find the sentiment score and the review which is used to train the model with the word bag under the different sentiments.
- I have ensured that I have sufficient data volume for the model to get trained. The model trained with fewer data would obviously have lesser accuracy, due to the limited knowledge about the domain data.
- Next, I developed the pre-processing code [22] module using python which would be used for the custom model training which is LSTM and BERT models.
- Created an LSTM model by inputting the pre-processed data and the model weight is stored for running the samples in the future.
- Created a BERT model with the same pre-processed data and the BERT model weight is also stored for future test runs.
- Imported the Vader Sentiment package and Text Blob package for finding the sentiment score from the pre-trained models for the sample text that I provide for custom models as well.
- Developed a Python Flask [17] Web Application to interact with the developed model.

3.1 Dataset:

The dataset chosen has the Twitter data that has the COVID-19 vaccination which is used for model training using the supervised learning approach. The chosen dataset has a sufficient volume of the dataset for the model training.

Link 01- https://ieee-dataport.org/open-access/coronavirus-covid-19-tweets-dataset

Name of the Dataset: COV19Tweets Dataset
Total Number of tweets: 2,104,250,235 tweets
Coverage: Global
Language: English (EN)

3.2 Data Pre-Processing:

The below-mentioned pre-processing steps are applied to the chosen dataset to eliminate the unnecessary words and characters from the data so that the model training would be more optimal. Different data format exists which could be used for model training. For the structured data, the essential features are explicitly available, whereas, for the textual data, the features are explicitly available. Instead, we need to apply some processes to identify the essential features and extract the same. The sentence in the training data is called the document and the collection that is formed from the document is known as the word corpus.

+ Removed the retweets from the text.
+ Removed the URLs from the text.
+ Remove the @user mentioning from the text.
+ Remove the hashtags from the text.
+ Perform ASCII encoding to normalize the special characters in the text.
+ Remove the numerical values from the text.
+ Change all the text to lower-case for uniformity in the text.
+ Remove the punctuation from the text.
+ Lemmatize the words in the text.
+ Remove the stop-words from the text.
+ Do Part-of-speech (POS) tagging in the text.

After the pre-processing, the processed text needs to convert into numerical values using the below functions:

+ Texts_to_sequences: Transforms the texts into an Integer sequence. The text to sequences in the Keras tokenizer performs the encoding of the sentences in the training data and form the sequences out of it.
+ Pad_sequences: Perform padding to have the sequences of the same length. This function transforms the list of data in the sequences/integers into an array with the 2D NumPy shape so that the sequence length of the list is set with the longest sequence

length. The maximum length of the sequence could be set when invoking the method. In this case, additional sequence values would get discarded.

+ The sequenced data is then provided to the model training.

3.3 BERT:

In the bidirectional encoder representation from transformers, the input data and the output data are connected to each other. When the model is fed with sentences it gives a single or list of vector outputs that can be used for our further reference. In the Bidirectional method, the text is analysed from right to left and left to right. Other models analyse the text in one direction only that is from left to right. Natural language processing involves the prediction of text. Some models are trained in a unidirectional way such as right or left, in such cases, Bidirectional encoder representation from transformers proves to be successful as it can predict both directions. This is the main place where the BERT model stands out from the other models. Transformers act as the essential model for Bidirectional encoder representation from the transformer's method.

To understand the working approach of the BERT Model, let's analyse a sample review which was mentioned in a sarcastic note and verify whether the review is positive or negative. Assume, the review is a mixed emotion like the "vaccination to be an ultimate failure", other models may read this sentence as positive with the word "ultimate" but bidirectional encoder representation from the transformer method analyse the words based on the context of the sentence and gives the output for this sentence as negative. The main problem of sentimental analysis is sarcasm, which is solved by this model language, as the method is analysed by the contextual meaning of the sentence/paragraph rather than the words alone. The application of the bidirectional encoder representation from transformer models is done by using the Tokenizers, which are used to process the data. Bidirectional encoder representation from transformers is used as an embedded layer. Fine-tuning the BERT and the core of the model.

To be more specific about the BERT algorithm that is used for implementation, the Roberta BERT [18] model which stands for Robustly Optimized BERT pre-training Approach is used in the implementation. The base BERT model uses static masking whereas Roberta uses dynamic masking which enables the model to be more robust. When compared to BERT, the Roberta model uses a large volume of datasets for the model training.

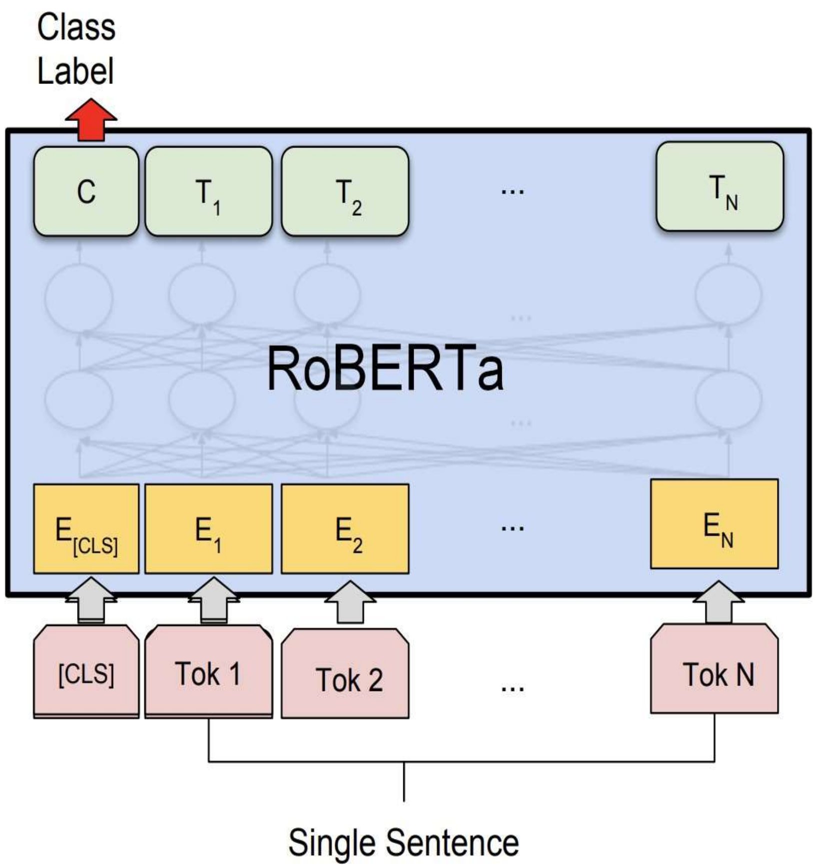


Figure 1: RoBERTa - A BERT Based Model [27]

3.4 LSTM - Long Short-Term Memory:

Unlike other methods, the LSTM method doesn't process only image formats. It can process the raw data in more than one format such as voice format, handwriting format, and robotic format [23]. The four methods which comprise the LSTM models are cell, input gate, output gate, and forget gate. The cell helps in the flow of value and information (stored in it) through the three gates. In the forget gate if the binary value of the processed input is 0, the information or the input is forgotten and not passed on to the other gate. Whereas when the binary value is more than 0 the information or the input is sent to the gates for further analysis. The information which is not needed is removed from this gate and not stored thereby decreasing the storage space occupied by the model. The inputs which are fed from the former gate are remembered and solved together to get the desired output. Several steps of analysis based on the pre-trained model for the specific tasks are done and then the output is obtained. The output gate acts as a screen for the output sent from the previous gates and the input feeder for the next gate if further analysis is to be done.

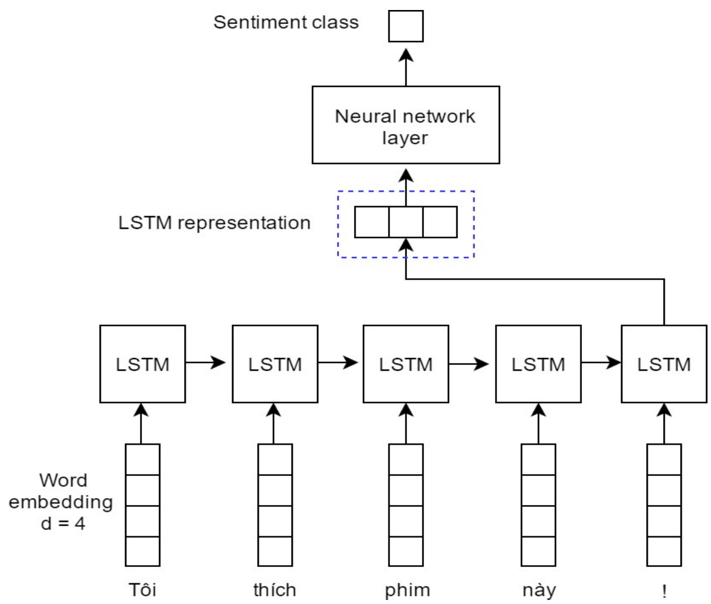


Figure 2: LSTM - Sentiment Analysis [28]

When compared [19] with the BERT - bidirectional encoder representation from the transformers model, this method doesn't consider the vanishing gradient. This affects the accuracy of the output when a surplus amount of raw data is analysed. The hardware and storage space becomes a problem when the model is trained with utmost accuracy. Some information is required to be remembered for the next step of processing; in such cases it takes a longer time to train such models.

3.5 Vader Sentiment:

This method deals with the scoring process of the sentiment used. It not only classifies the sentiment as positive or negative but even tells the score of positivity or negativity in it. The sentimental score is the form of output obtained from it. All the words in a paragraph or phrase sum up to the score of the output. It not only understands the positive or negative words in the text, but it can also understand the emotion of the adverb or noun proceeding or descending it. The words such as sad and disappointed come under the category of negative words, but in the context of "not sad" the sentimental analysis language treats it as positive words.

Vader language exceeds the human accuracy level as is it sensitive to polarity and the intensity of the positivity or negativity of the word. The output comes as a sore. The score values range between -4 to +4. The value -4 indicates negative whereas the +4 indicates positive. The value 0 indicates the neutral emotion of the text. Vader language doesn't require any pre-training as it can understand the sentiment of the paragraph or sentence. The punctuation marks such as full stops and commas can also be understood by the Vader method. Hence Vader's method proves to be successful in analysis the of social media reviews.

3.6 Text Blob:

The text blob [25] method is python-based language. It is used in sentimental analysis and word correction predominantly. Polarity and subjectivity are the two forms of output. The extreme degree of the output such as -1 and +1 comes under the category of polarity whereas the values lying between 0 to 1 comes under the category of subjectivity. Subjectivity is the opinion or review, and it can differ from the original facts. Personal opinions and judgments come under the category of subjectivity. Its analysis the words it knows and leaves the words it doesn't know as it is a pre-trained model. The text blob method used the entire words and process the analyses for each word. The final output is the average score of all the words. It can detect emojis, smileys, etc so it is used in the process of fine training. When the output has a higher range of subjectivity it means the personal context is more than the factual context. Intensity is one other feature that changes the subjectivity value. For example, "you are beautiful" has a subjectivity of 0.1 whereas "you are very beautiful" has a subjectivity of 0.5. This change in the range of values is because of the use of adverbs. The feature intensity considers the adverbs in analysing the text. Application import and writing of the program which calculates the subjectivity and polarity of the paragraph or phrase are the important steps involved in the text blob process. The Text Blob internally consumes the NLTK core module to perform the tasks like part-of-speech tagging, extracting the nouns from the phrase and so on. Even for performing the sentiment analysis, it uses the NLTK features to process the data.

3.7 Web User Interface Development:

The developed model needs to be accessible for performing the validation and making use of it. On top of that, for enabling the best user to experience the interaction with the application needs to be easy and smooth. As a part of this research, I have built the sentiment analysis model which could be validated with the COVID-19 vaccination-related text. To enable easy interaction for providing the input text, a python flask web application was created as a part of the research. The user could provide the text in the input box, the provided text is taken into the application backend by providing it to the developed models and obtained results are presented to the user in the same interface. On the backend, the application programming interface (API) is developed so that the model could be reused as a service which is consumable by other applications.

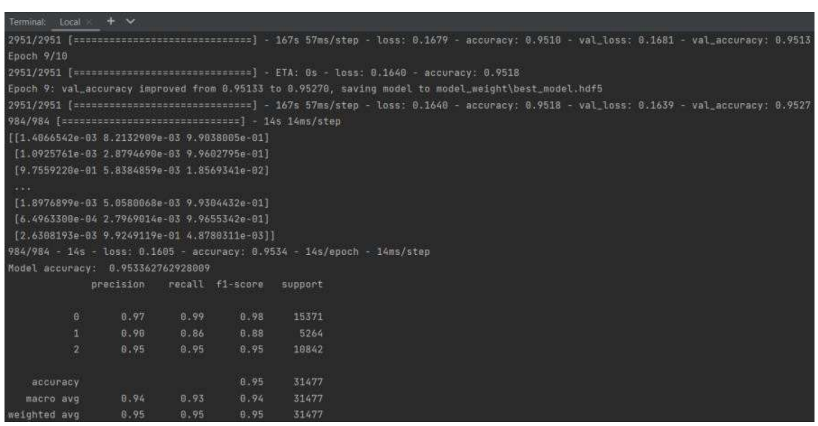
The Python flask is a microweb framework which is used to develop lightweight web applications using the python language. The learning and implementation of the Python flask are flexible in a way, even the new developers could build the web application with the API access in a quick way. For the development of the Python Flask application, the basic requirement is having knowledge of the programming components like HTML and Python.

4 Evaluation

The developed models need to be evaluated with the evaluation metrics [20] to identify which model works best. The evaluation is also required to identify if any fine tunings are required for the model to improve accuracy, recall, precision, etc. The model evaluation is usually conducted based on certain metrics to understand how the machine learning or deep learning models are performing so that the identification could be done in terms of the strengths of the model as well as the weakness of the model. Usually, the evaluation takes phases in 2 different situations. One is during the initial model development, in which the evaluations would be the supporting point to decide the best optimal model if more models are chosen. Or also with the evaluation results, the decisions could be taken for eliminating the model for production deployment. The second situation in which evaluation would be done is during the monitoring phase which usually happens post deploying the model into the production environment. The evaluation is important in the monitoring phase as well, due to the volume of data that follows into the model for training using the data streaming process. Also, the test data that has been used in the development phase would've influenced the model accuracy. So, when it comes to the monitoring phase, the real data could be used for evaluation, which would provide more appropriate evaluation results.

4.1 Training Results

On the training process completion of the BERT model, different evaluation factors are displayed. The model is trained with 10 epochs and after completing all the epochs the model weight is stored for future use. The accuracy of the model denotes the performance measure that represents the ratio of the predictions that were correct with the total number of observations. Our model states the accuracy value is 0.9533, which is 95% accuracy. The term precision refers to the ratio of correct predictions to the total number of predictions. The term recall refers to the number of correct predictions on any specific label. For example, positive label prediction results. The F1 Score refers to the weighted average of Precision and Recall. With all the above metrics, the model could be identified on how good it is predicting.



```
Terminal: Local
2951/2951 [==============================] - 167s 57ms/step - loss: 0.1679 - accuracy: 0.9510 - val_loss: 0.1681 - val_accuracy: 0.9513
Epoch 9/10
2951/2951 [==============================] - ETA: 0s - loss: 0.1640 - accuracy: 0.9518
Epoch 9: val_accuracy improved from 0.95133 to 0.95270, saving model to model_weight\best_model.hdf5
2951/2951 [==============================] - 167s 57ms/step - loss: 0.1640 - accuracy: 0.9518 - val_loss: 0.1639 - val_accuracy: 0.9527
984/984 [==============================] - 14s 14ms/step
[[1.4066542e-03 8.2132909e-03 9.9038005e-01]
 [1.0925761e-03 2.8794690e-03 9.9602795e-01]
 [9.7559220e-01 5.8384859e-03 1.8569341e-02]
 ...
 [1.8976899e-03 5.0580068e-03 9.9304432e-01]
 [6.4963300e-04 2.7969014e-03 9.9655342e-01]
 [2.6308193e-03 9.9249119e-01 4.8780311e-03]]
984/984 - 14s - loss: 0.1605 - accuracy: 0.9534 - 14s/epoch - 14ms/step
Model accuracy:  0.953362762928009
              precision    recall  f1-score   support

           0       0.97      0.99      0.98     15371
           1       0.90      0.86      0.88      5264
           2       0.95      0.95      0.95     10842

    accuracy                           0.95     31477
   macro avg       0.94      0.93      0.94     31477
weighted avg       0.95      0.95      0.95     31477
```

Figure 3: BERT Model Training Result

The LSTM has shown an accuracy rate of 90%. Out of the comparison, the BERT model works well when compared to the LSTM. In reality, the comparison with the pre-trained models couldn’t be done as I haven’t trained the model, it has been just consumed in the research project so a comparison could be done in terms of the results for the sample input that would be provided for the models.

Obtaining the model that results with the best accuracy rate and minimizing the time that the model takes for getting trained are influenced by the major process called data pre-processing. The primary reason for this is, in the data pre-processing the volume of the data is reduced by eliminating all the unnecessary information which is not required for identifying the sentiment score. In the research, all the possible data pre-processing steps were followed before the data is provided to the model for getting it trained. The model results are compromising the data set that I have chosen for the model development. The fact is that the quality of the dataset is also the major factor that would have an impact on the accuracy if the quality were not up to the level. In the data pre-processing step, it is also essential to verify that none of the essential information is lost in the process. For example, some of the models are capable to analyse the emojis which is added as a part of the text. The emojis have meaning and it is one of the expressing it. The process needs to ensure this information is not lost.

4.2 Experiment / Case Study 1

Multiple experiments were conducted once the model is built. The first experiment is to obtain the neutral polarity for the review text “I have got COVID-19 even after getting vaccinated.” that is provided as input to the model.

Figure 4: Experiment - Neutral Review

Result: As expected, the models are predicting the sentence as a neutral statement. With the developed user interface, the results could be easily viewed from the interface.

4.3 Experiment / Case Study 2

The second experiment is conducted by providing a positive review "I believe in goodness. I believe vaccines would save my life." of the application models and observed the below results.

Figure 5: Experiment - Positive Review

Result: The LSTM Model, Vader and Text blob are predicting the sentence as a positive statement. Anyways the BERT classify the text has a Neutral statement.

4.4 Experiment / Case Study 3

The third experiment is conducted by providing a negative review "the vaccine has side effects and it's not advisable; the pharma company failed to deliver a quality vaccine." of the application models and observed the below results.

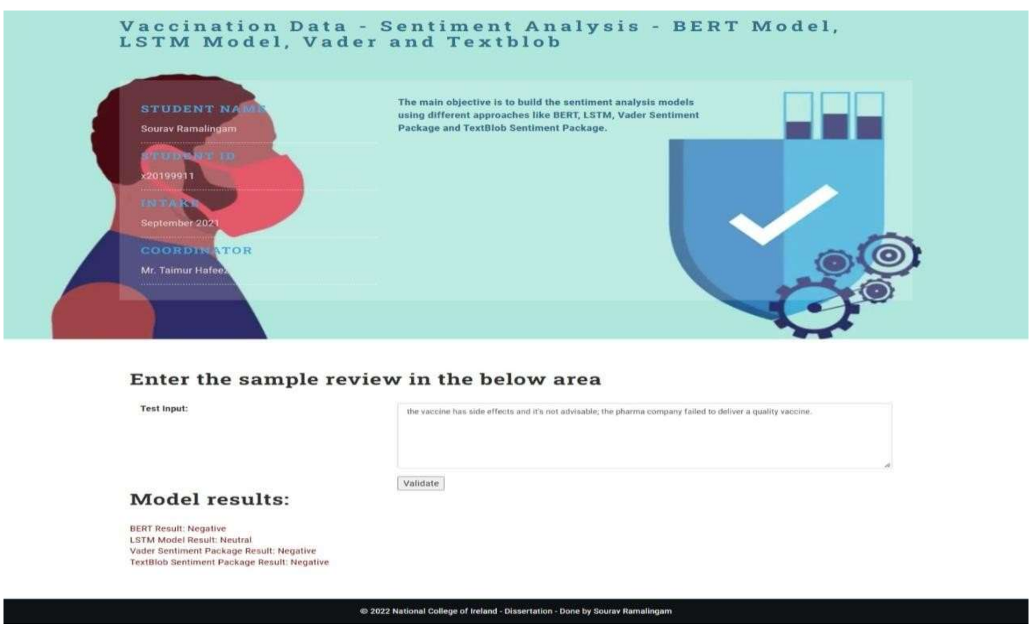


Figure 6 : Experiment - Negative Review

Result: The BERT model, Vader, and Text Blob are predicting the sentence as a negative statement, whereas the LSTM model has predicted has Neutral Statement.

5 Conclusion and Future Work

To conclude, out of this research, I have developed a complete solution that performs the sentiment analysis on the domain-specific which is COVID-19 Vaccination data. As I have obtained the aimed result from the research work, I would say that the methodology that I have chosen for the development was optimal. In terms of the results that have been obtained during the evaluation phase, the results were promising in terms of accuracy. For any research, the performance of the model is important, as we couldn't increase the scale of the model if it is performing poorly. The end-to-end process that I followed in the research developed helped to obtain knowledge like working with industrial standards.

As a part of future work, the model would be extended by training with more volume of dataset gathered from different sources, instead of just having only the Twitter feeds for model training. On top of that, more algorithms could be developed and used for comparing the results to identify which models work best for the chosen dataset and domain. Once the model is enhanced in the further phases, the model could be then integrated with the social sites or any blobs when medical reviews are usually written, so that model could be ensured that it works in a real-time environment with a wide range of data as well. On top of that, the model training and deployment could be done in the Graphics processing unit - GPU-based system, as the current development took place in the Central Processing Unit - CPU-based computers which have very limited resources.

References

1. Li, Z., Fan, Y., Jiang, B., Lei, T. and Liu, W., 2019. A survey on sentiment analysis and opinion mining for social multimedia. Multimedia Tools and Applications, 78(6), pp.6939-6967.

2. Tang D, Qin B, Liu T (2015) Deep learning for sentiment analysis: successful approaches and future challenges. Wiley Interdiscip Rev Data Mining Knowledge Discovery 5(6):292–303

3. Giachanou A, Crestani F (2016) Like it or not: a survey of twitter sentiment analysis methods. ACM Comput Surv 49(2):28

4. Silva NFFD, Coletta LFS, Hruschka ER (2016) A survey and comparative study of tweet sentiment analysis via semi-supervised learning. ACM Comput Surv 49(1):15

5. Taboada M, Brooke J, Tofiloski M et al (2011) Lexicon-based methods for sentiment analysis. Comput Linguistics 37(2):267–307

6. Chowdhary, K., 2020. Natural language processing. Fundamentals of artificial intelligence, pp.603-649.

7. Li, X., Bing, L., Zhang, W. and Lam, W., 2019. Exploiting BERT for end-to-end aspect-based sentiment analysis. arXiv preprint arXiv:1910.00883.

8. Huang, X., Zhang, W., Tang, X., Zhang, M., Surbiryala, J., Iosifidis, V., Liu, Z. and Zhang, J., 2021, April. Lstm based sentiment analysis for cryptocurrency prediction. In International Conference on Database Systems for Advanced Applications (pp. 617-621). Springer, Cham.

9. Elbagir, S. and Yang, J., 2019, March. Twitter sentiment analysis using natural language toolkit and VADER sentiment. In Proceedings of the international multiconference of engineers and computer scientists (Vol. 122, p. 16).

10. Gujjar, J.P. and Kumar, H.P., 2021. Sentiment analysis: Textblob for decision making. Int. J. Sci. Res. Eng. Trends, 7(2), pp.1097-1099.

11. Brauwers, G. and Frasincar, F., 2022. A survey on aspect-based sentiment classification. ACM Computing Surveys, 55(4), pp.1-37.

12. Tian, H., Gao, C., Xiao, X., Liu, H., He, B., Wu, H., Wang, H. and Wu, F., 2020. SKEP: Sentiment knowledge enhanced pre-training for sentiment analysis. arXiv preprint arXiv:2005.05635.

13. Alhazmi, A., Alamer, E., Daws, D., Hakami, M., Darraj, M., Abdelwahab, S., Maghfuri, A. and Algaissi, A., 2021. Evaluation of side effects associated with COVID-19 vaccines in Saudi Arabia. Vaccines, 9(6), p.674.

14. Levey, A.S., Gansevoort, R.T., Coresh, J., Inker, L.A., Heerspink, H.L., Grams, M.E., Greene, T., Tighiouart, H., Matsushita, K., Ballew, S.H. and Sang, Y., 2020. Change in albuminuria and GFR as end points for clinical trials in early stages of CKD: a scientific workshop sponsored by the National Kidney Foundation in collaboration with the US Food and Drug Administration and European Medicines Agency. American Journal of Kidney Diseases, 75(1), pp.84-104.

15. Ahmad, R. and Hillman, S., 2021. Laboring to communicate: Use of migrant languages in COVID-19 awareness campaign in Qatar. Multilingua, 40(3), pp.303-337.

16. Schröer, C., Kruse, F. and Gómez, J.M., 2021. A systematic literature review on applying CRISP-DM process model. Procedia Computer Science, 181, pp.526-534.

17. Mufid, M.R., Basofi, A., Al Rasyid, M.U.H. and Rochimansyah, I.F., 2019, September. Design an mvc model using python for flask framework development. In 2019 International Electronics Symposium (IES) (pp. 214-219). IEEE.

18. Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L. and Stoyanov, V., 2019. Roberta: A robustly optimized bert pretraining approach. arXiv preprint arXiv:1907.11692.

19. Ezen-Can, A., 2020. A Comparison of LSTM and BERT for Small Corpus. arXiv preprint arXiv:2009.05451.

20. Poecze, F., Ebster, C. and Strauss, C., 2018. Social media metrics and sentiment analysis to evaluate the effectiveness of social media posts. Procedia computer science, 130, pp.660-666.

21. Carvalho, F., Rodrigues, R.G., Santos, G., Cruz, P., Ferrari, L. and Guedes, G.P., 2019, July. Evaluating the Brazilian Portuguese version of the 2015 LIWC Lexicon with sentiment analysis in social networks. In Anais do VIII Brazilian Workshop on Social Network Analysis and Mining (pp. 24-34). SBC.

22. Pradha, S., Halgamuge, M.N. and Vinh, N.T.Q., 2019, October. Effective text data preprocessing technique for sentiment analysis in social media data. In 2019 11th international conference on knowledge and systems engineering (KSE) (pp. 1-8). IEEE.

23. Behera, R.K., Jena, M., Rath, S.K. and Misra, S., 2021. Co-LSTM: Convolutional LSTM model for sentiment analysis in social big data. Information Processing & Management, 58(1), p.102435.

24. Yao, J., 2019, April. Automated sentiment analysis of text data with NLTK. In Journal of Physics: Conference Series (Vol. 1187, No. 5, p. 052020). IOP Publishing.

25. Diyasa, I.G.S.M., Mandenni, N.M.I.M., Fachrurrozi, M.I., Pradika, S.I., Manab, K.R.N. and Sasmita, N.R., 2021, May. Twitter Sentiment Analysis as an Evaluation and Service Base On Python Textblob. In IOP Conference Series: Materials Science and Engineering (Vol. 1125, No. 1, p. 012034). IOP Publishing.

26. Ma, Y., Peng, H., Khan, T., Cambria, E. and Hussain, A., 2018. Sentic LSTM: a hybrid network for targeted aspect-based sentiment analysis. Cognitive Computation, 10(4), pp.639-650.

27. Image Reference: https://production-media.paperswithcode.com/models/roberta-classification.png-0000000936-4dce6670.png

28. Image Reference: https://www.researchgate.net/profile/Huy-Tien-Nguyen/publication/321259272/figure/fig2/AS:572716866433034@1513557749934/Illustration-of-our-LSTM-model-for-sentiment-classification-Each-word-is-transfered-to-a.png