# ANALYSING COVID-19 VACCINE-RELATED TWEETS: INSIGHTS INTO PUBLIC SENTIMENT ON SOCIAL MEDIA PLATFORMS

FINAL THESIS PRESENTATION

BY

SARAVANAN KANMANI

MASTER OF SCIENCE IN DATA SCIENCE

LIVERPOOL JOHN MOORE'S UNIVERSITY (IN ASSOCIATION WITH UPGRAD)

MARCH 2024

# COVID-19 – A Global Pandemic

**SARS-CoV2 (COVID-19)** virus, an infectious disease was first emerged in **Wuhan, China** in **December 2019**.

The virus spreads primarily through **infected persons** when they cough, sneeze, talk, sing or breathe. The virus rapidly outspread around the world due to its **highly contagious** nature.

COVID-19 cases and fatalities were increased rapidly every day. On **March 11, 2020**, the **World Health Organization (WHO)** officially declared COVID-19 a **pandemic**.

As of 13th March 2024, approximately **774 million** people have been affected by the virus and there have been **over 7 million deaths** worldwide.

# COVID-19 Vaccine and its challenges

The global fight against COVID-19 witnessed remarkable efforts by pharmaceutical companies, research institutes, and government organizations to **develop vaccines**.

Due to misinformation, rumours, safety and side effects concerns **people hesitated** to accept the vaccines.

Health organizations **faced challenges** in achieving widespread vaccine coverage to control the disease that slowed down vaccination campaigns.

So, the health authorities must understand **public sentiments about COVID-19** vaccines for effective planning.

# Sentiment analysis on social media

Sentiment analysis is an NLP technique used to find out **emotional tone** conveyed in a piece of text.

Social media is a valuable resource where people share their **individual thoughts** or comments openly.

Public discussed about the **COVID-19 and its vaccine** as well in the social media platform like X (formerly known as Twitter) in the form of tweets.

Analysing the tweets about COVID-19 vaccines from X platform can identify **public opinion** about the COVID-19 vaccines.

# Aim and Objectives of the study

## Aim

This study aimed to conduct sentiment analysis on COVID-19 vaccine related tweets from the X platform.

The outcome of this study could be beneficial to government organization, health authorities and policy makers to plan their vaccination campaign efficiently in similar situations.

Profit organization can also understand the sentiments of public to produce essential items.

Also, this study can provide a better approach to classify the sentiments

## Objectives

To analyse the COVID-19 vaccine-related tweets posted by public and find the most common sentiments expressed.

To find the recurrent keywords for each sentiment discussed in COVID-19 vaccine related tweets.

To classify the sentiments using suitable lexicon and deep learning model

To evaluate the performance of the classification model developed.

# Literature Review

Multiple researchers have explored different approaches to analyse public sentiments regarding COVID-19 and its vaccine.

**COVID-19 pandemic and vaccines**

- *Kwon et al.* research shows the effectiveness of **wearing masks** to control COVID-19 transmission
- Prevention of the disease relied on the **development and adoption of vaccine** as studied by *Chou and Budenz*.
- *Levin et al.* found vaccine BNT162b2 COVID-19 substantially **decreased** the virus in six months after the **second dose**.

**Sentiment Analysis**

- According to *Dhanalakshmi et al.*, opinions of stakeholders have a **greater impact** on decision making than facts.
- Sentiment analysis are classified into different levels as **Document**, **Sentence** and **Aspect** as per *Birjali et al*.
- There are three main techniques in sentiment analysis *Aqlan et al.* and they are **Lexicon-based**, **Machine learning** based and **hybrid** methods.

# Literature Review – cont.

## X (Twitter) sentiment analysis

- Social media influence a crucial role in **shaping public perception** during health crisis such as global pandemics - *Malecki et al.*
- X platform filled with COVID-19 related tweets so analysing it can **identify patterns** in it - *Kaur et al.*

## Related research

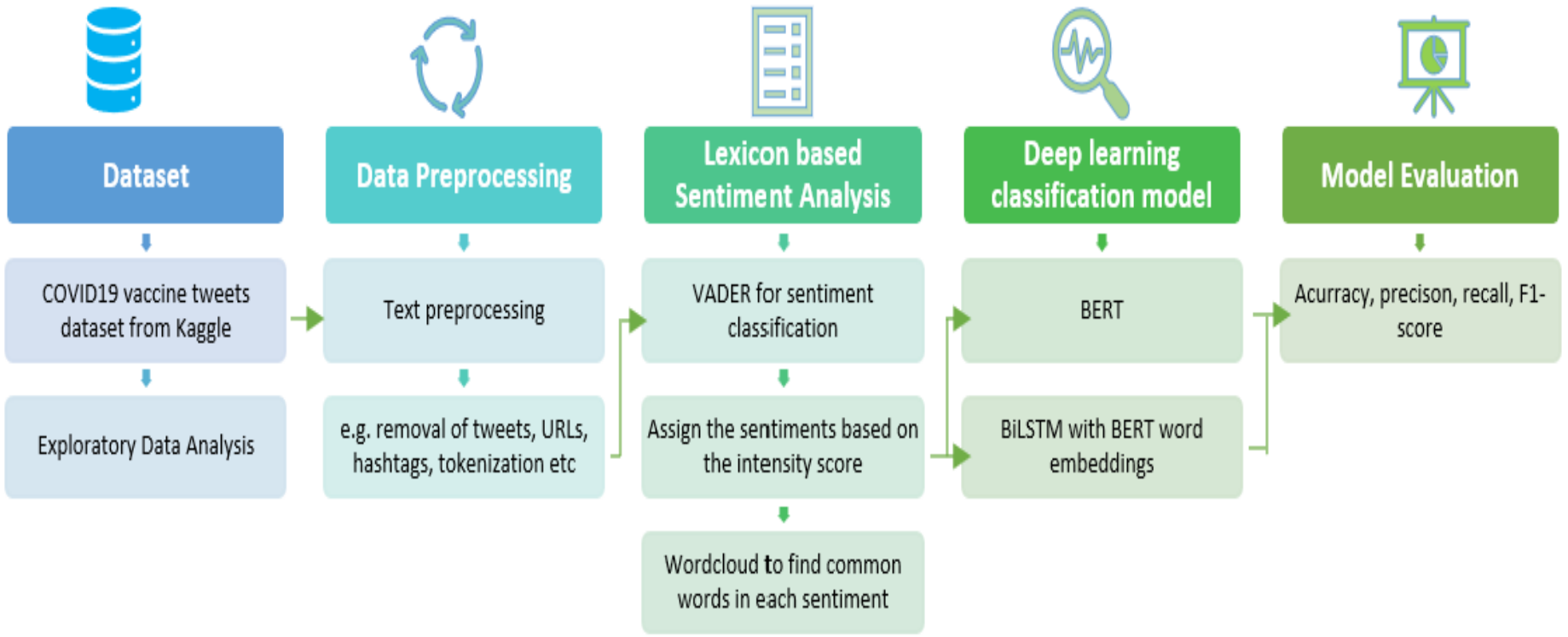
- *Ainapure et al.* investigated the impact of **deep learning models with lexicon-based** approaches to assess the sentiments.
- *Aslan et al.* identified CNNs and optimization techniques **improved** sentiment classification **accuracy**.
- BERT with CNN models achieved **state-of-the-art results** in sentiment analysis - *Joloudari et al.*
- TextBlob enables sentiment analysis **without** requiring **extensive training** data - *Kathiravan et al.*
- *Ferdous et al.* proposed neural network that use text vectorization that compared with **LSTM** and **Bi-LSTM** models.
- **BERT** model **outperformed** other three traditional models like LR, SVM and LSTM - *Chintalapudi et al.*

# Problem Statement

- More people were getting **infected** with COVID-19 and **dying** from it. But, anti-vaccination groups on social media **influenced** others to **reject vaccines**.
- Many researchers analysed COVID-19 vaccine related tweets to understand the human behaviour. Most of the researchers used **either lexicon or deep learning** approach.
- Some of the researchers used only two classes (i.e. positive and negative) of sentiment with limited number of tweets or used only ML algorithms.
- To overcome some of the existing problems, this study used **both lexicon and deep learning** for sentiment classification and prediction. Also, the tweets are classified into **three classes** (i.e. positive, negative and neutral)

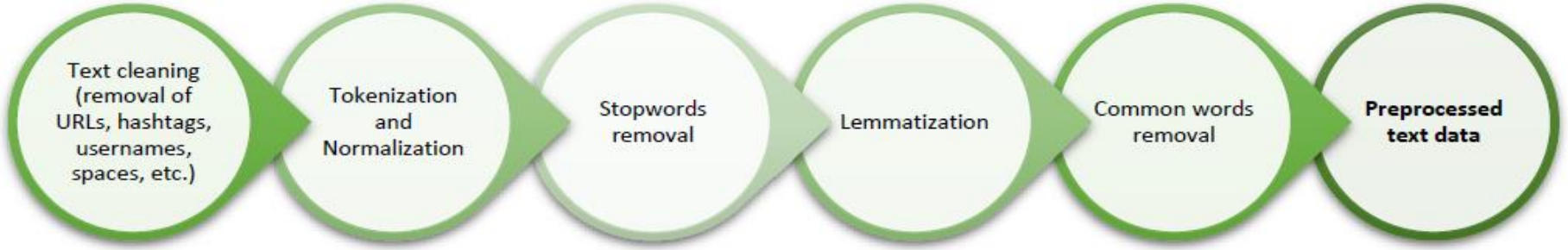
# Methodology

Overview  of the research methodology is shown below.

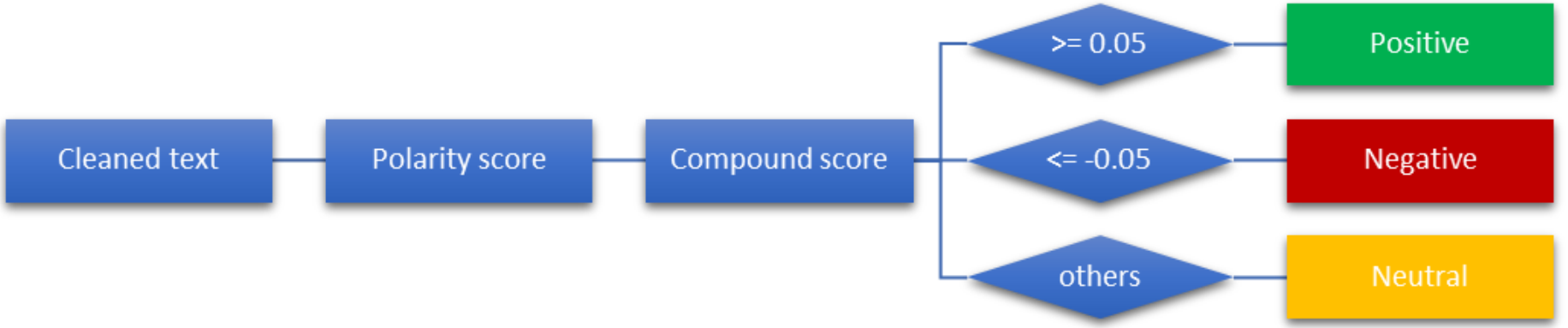


# Methodology cont. 1

**Data preprocessing**

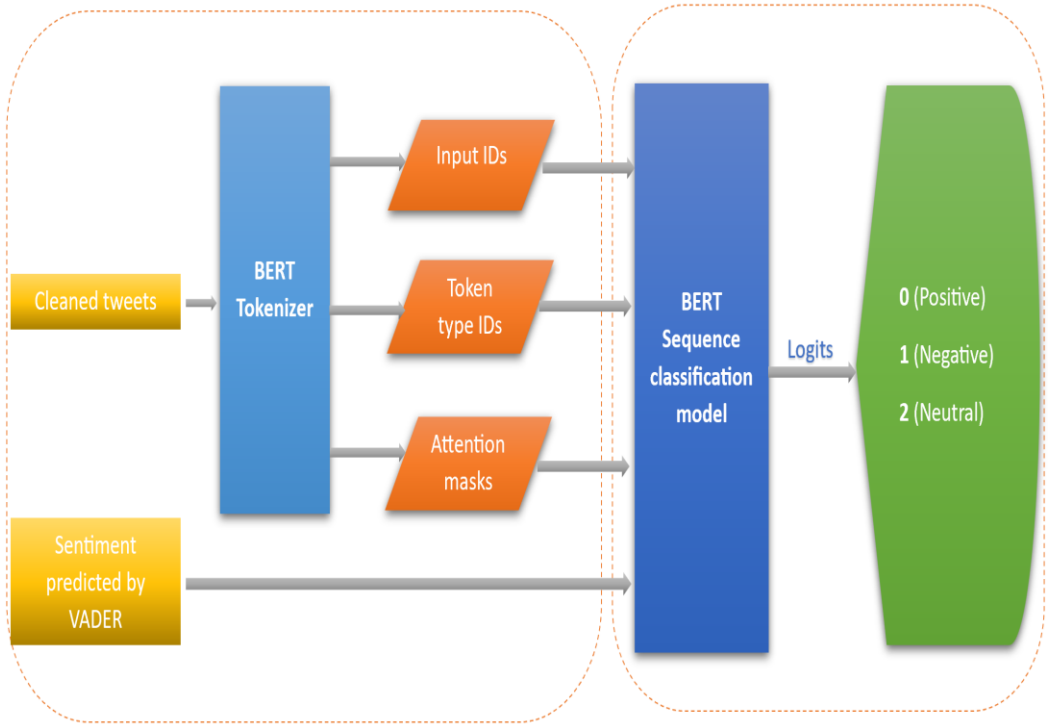



**VADER sentiment classification**

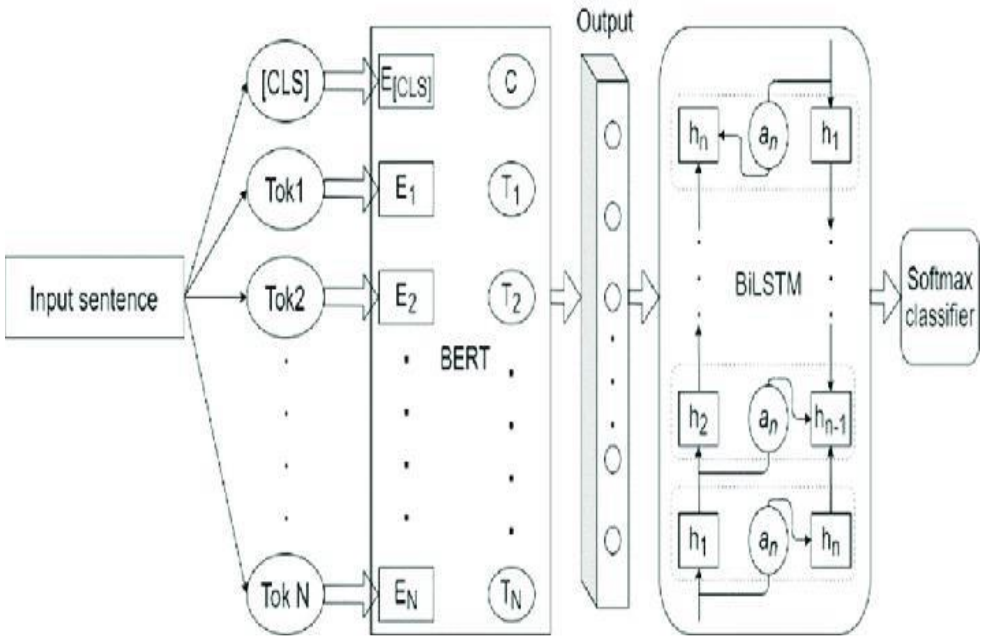


# Methodology cont. 2

## BERT Model




## Bi-LSTM with BERT word embeddings

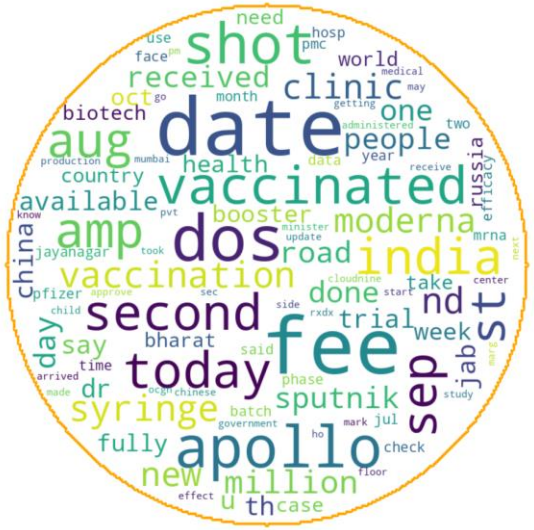


# Results & Discussion

Wordcloud of sentiments predicted using VADER

Positive


Neutral




Negative

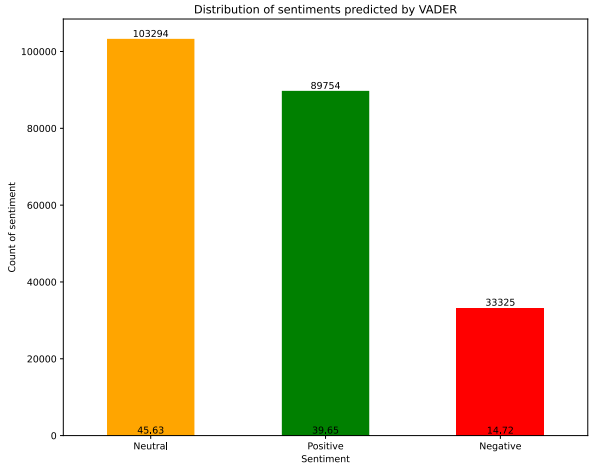
# Results & Discussion – cont. 1

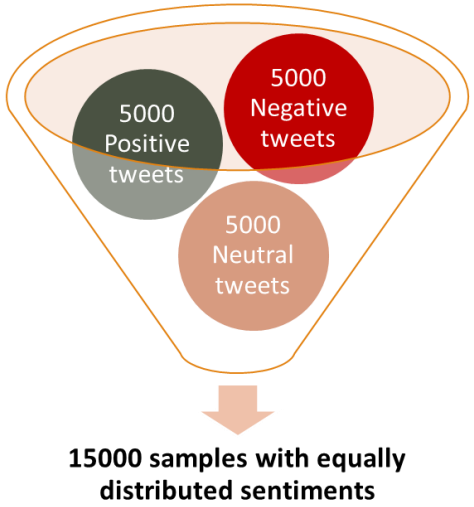
**Bootstrap resampling technique**

Sentiment classification by VADER provided imbalanced dataset.

Bootstrap resampling with sample replacement technique has been used for the following reasons:

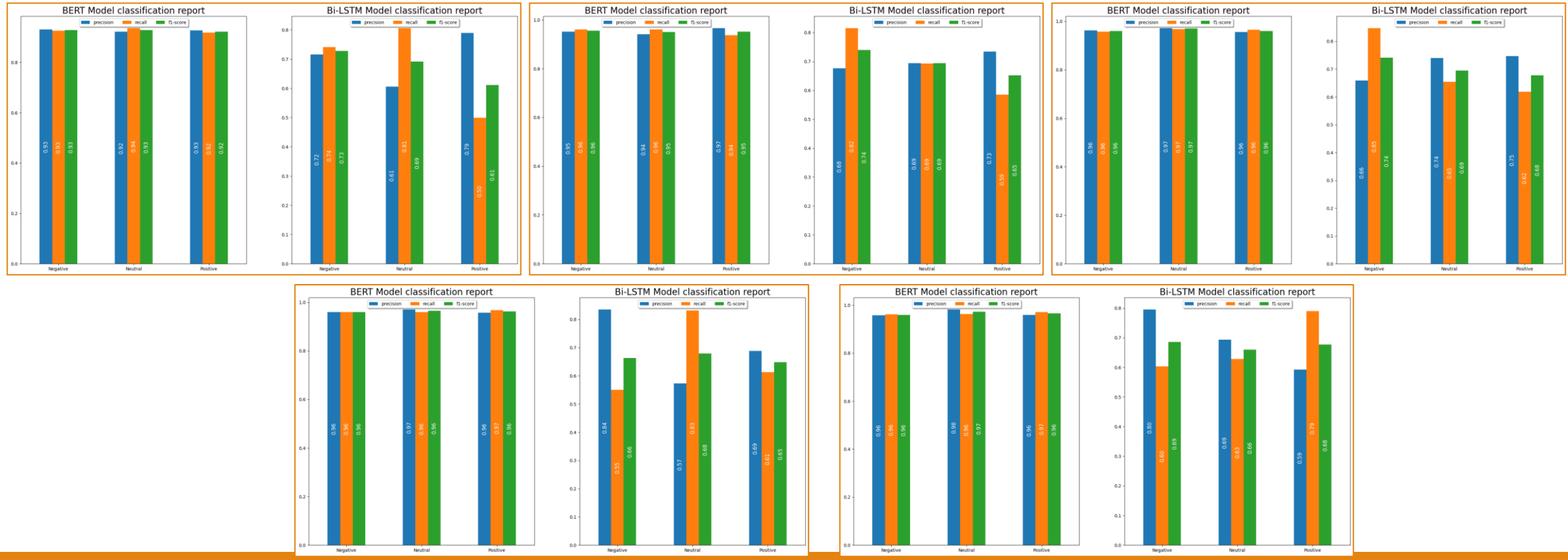
- To handle the imbalance dataset
- To accommodate computation cost and time required to train deep learning models.



# Results & Discussion – cont. 2

Both the models BERT and Bi-LSTM with BERT has been executed for five times for different set of bootstrapped samples.



# Results & Discussion – cont. 3

| Sample set | BERT test accuracy | Bi-LSTM with BERT test accuracy |
|---|---|---|
| 1 | 90.22% | 66.78% |
| 2 | 93.89% | 70% |
| 3 | 94.33% | 67.81% |
| 4 | 95.29% | 69.22% |
| 5 | 95.74% | 68.78% |

| Overall test accuracy | |
|---|---|
| **BERT** | **Bi-LSTM with BERT** |
| 93.9% | 68.52% |

Average of all the iterations test accuracies has been calculated to get the overall test accuracy of the models.

BERT sequence classification model outperformed the Bi-LSTM with BERT word embeddings model in terms of accuracy, precision, recall and F1-score

Bi-LSTM with BERT model's validation loss was very close to the training loss.

# Conclusion and future works

## Conclusion

- Hybrid approach of using both lexicon and deep learning models to classify the sentiments.
- Major keywords found across the sentiments could be helpful for relevant stakeholders.
- Limitations: less sample data for deep learning, only historical data used and only English language tweets are considered.

## Future works

- Sarcasm and misinformation should be eliminated before classification.
- Real time tweets can be included to find the latest discussion topics.
- Other language tweets can also be considered and train for large set of data.

# Thank you