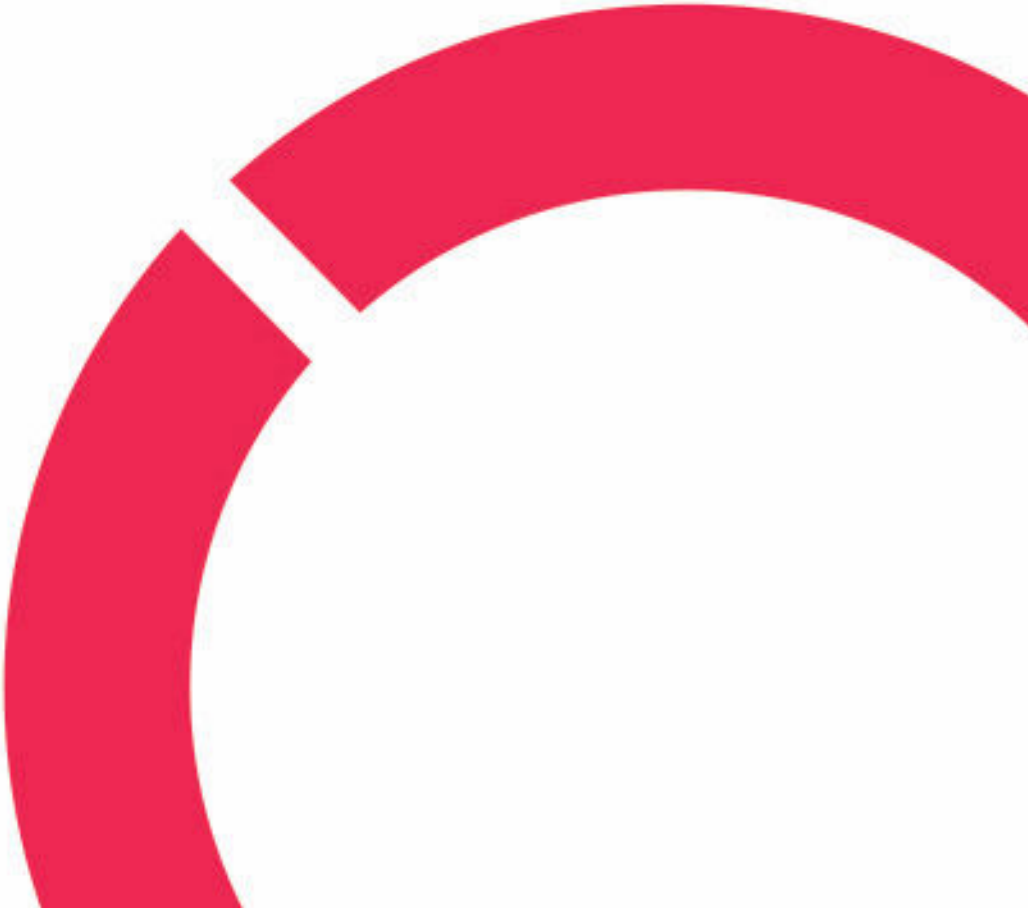
Md Sadbin Neymat

EXPLORING THE CORRELATION BETWEEN SOCIAL MEDIA LANGUAGE PATTERNS AND DEPRESSION

A Sentiment Analysis Approach

Thesis
CENTRIA UNIVERSITY OF APPLIED SCIENCES
Master of Engineering, Cloud-based Software Engineering
May 2024



ABSTRACT

Centria University of Applied Sciences	Date May 2024	Author Md Sadbin Neymat
Degree programme Master of Engineering, Cloud-based Software Engineering		
Name of thesis EXPLORING THE CORRELATION BETWEEN SOCIAL MEDIA LANGUAGE PATTERNS AND DEPRESSION. A Sentiment Analysis Approach.		
Centria supervisor Aliasghar Khavasi		**Pages** 63
Instructor representing commissioning institution or company -		

The advancement in social media technology has provided modern communication platforms through which the users can express their thoughts and emotions towards a particular event. The increase in mental health issues are mainly based over the depression that may arise because of a particular event causing impact in change in the sentiment of the user over the social media.

This thesis focuses on the identification of the correlation between the language patterns and depression that is reflected by the social media users for reporting the underlying sentiment. For this purpose of identifying the depression content root causes, the use of prediction schemes has been reported in the work.

The implementation of logistic regression and Lexicon based sentiment analysis computation has been provided in the work. It uncovers the scheme through which the extraction of the emotion content can be extracted from the review texts. An accuracy of around 86% has been obtained by using logistic regression and accuracy value of 63.38% is reported from the lexicon-based scheme. In other words, this shows that machine learning can be used to find the real reason behind an event.To this reason, the evaluation has been considered by including the precision and recall measures as well. From identifying the negative sentiment root cause, Covid and Cardinal disease has been identified as the base which are mainly causing depression in the tweets of collected data.

Key words
Behaviour Analysis, Depression, Machine Learning, Sentiment Analysis.

CONCEPT DEFINITIONS

CSV

Comma Separated Value, is a file type used in datasets to make it more accessible for data analysis.

LSTM

RNN comes in different forms, and Long Short-Term Memory is one of them.

SVM

Support Vector Machines (SVM), Mostly used for sorting and predicting in machine learning.

NLP

Another name for machine learning is natural language processing (NLP). NLP lets computers understand, change, and process human words.

API

API is Application Programming Interface which enables the connection or communication between two or more computer programs.

NLTK (Natural Language Toolkit)

This is used for NLP which is a popular API for Python.

ML

The system which automates a machine through training the machine with various algorithms is known as Machine learning (ML).

Social Media

Online platforms enabling users to create and share content, fostering digital interactions and communities.

Depression

A mental illness that causes people to feel sad, hopeless, and uninterested in doing normal things for a long time.

Sentiment Analysis

A natural language processing method is used to look at text data and pull out subjective information, like feelings and views.

Machine Learning

A type of artificial intelligence that lets computers learn from data and make guesses without being explicitly programmed to do so.

Logistic Regression

Statistical technique used for binary classification tasks, predicting the probability of a given outcome based on input features.

Lexicon

A list of words or sentences that are used for mood analysis and have scores assigned to them.

Emotion Detection

Method for finding and grouping feelings shown in text or speech data.

Feature Extraction

The process of choosing and representing important features in raw data so that machine learning programmes can use it.

Mental Health

State of well-being encompassing psychological, emotional, and social aspects of an individual's life.

Digital Communication

Exchange of information through digital channels, including social media platforms, emails, and instant messaging.

Early Detection

Identification of potential indicators or risk factors associated with a condition or disorder before its full manifestation, facilitating timely intervention and treatment.

ABSTRACT
CONTENTS

FIGURES

1 INTRODUCTION

This chapter has been formulated for the purpose of presenting the basic concepts that have been adopted during the development of the thesis work. The incorporation of these concepts is important from the context that it provides the overview of the features that has been included in the implementation process. Creating a social media-based mood analysis is seen as an important job for getting a sense of how people are feeling. Generally, the brands focus over the concept of target audience analysis which includes a certain type of the audience that are primary relevant to their business. They offer products to those specific subsets of the users that have highest chances to increase their growth. Hence, for the business related to the medical industry, there are important reasons why users are sad that need to be found. The sections given in this chapter mainly present the core concepts of the topic along with the problem statement and the objectives that are formulated for this work.

1.1. Introduction to Social Media and Review Analysis

The practice of utilizing social media data as a foundation for constructing a business strategy is increasingly prevalent. (Alslaity & Orji, 2022). The core reason behind this activity is the identification of the root causes and the relevance of a certain emotion depiction event along with their sentiment type. These platforms are also becoming the first supply line through which the business can target the user for their marketing. To this reason, the focus over the review analysis from the social media has been considered for this work.

1.1.1. Background of Sentiment Scoring

The fundamental principle underlying sentiment analysis of reviews is to determine the polarity of user input feelings. One of the major reasons behind the analysis of these reviews can be understood from the perspective that these reviews enable the user to understand the consumer mindset and the latest trends which are reported. Generally, the change of mindset is based over different factors and there are generally some root cause driven events which cause a certain type of change in the mindset of the user (Asif, Ishtiaq, Ahmed, Aljuaid & Shah, 2020). For this reason, the user depicts a particular behaviour on the social media for reflecting their views towards the event. Now, a particular review based reflection can be in both positive or negative type. The identification of root causes behind each type is

important to reflect for the business which are the driven factors that are changing the mindset of the user from positive to the negative. Hence, the adoption of the technique for the reporting of these polarity computation approaches are mainly reported as the baseline through which the user can get the relevant feedback analysis approach to report certain problem. The businesses can use these reviews as the baseline over which they can build the possible solutions in terms of handling the problems of their consumer (Bharti, Varadhaganapathy, Gupta, Shukla, Bouye, Hingaa, & Mahmoud, 2022). Moreover, they can also incorporate the possible ways of the offering and marketing of their products to the specific type of the audience by identifying their emotion status.

1.1.2. Use of Automated Techniques for Sentiment Analysis

Automated solutions for sentiment calculation are considered the foundation upon which users may determine the fundamental reason for their company concept. Generally, the business models are based over certain factors and each of these factors have certain impact over the mindset of the consumer (Birjali et al., 2021). Hence, it is important to enhance the mindset of the user by providing relevant factors based solution through which their understanding of the business can be improved. Also, these solution can help the user by getting them out from the depression state and make the overall environment happy by accepting the offers of the business. It also reflects the success of the business with high margin growth as shown in figure 1 below:

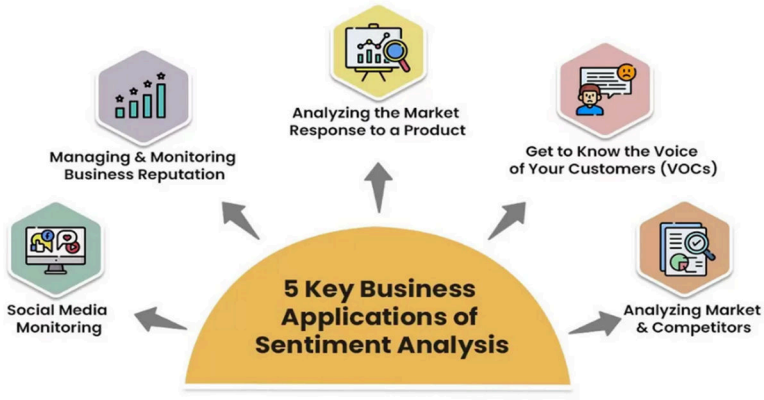


FIGURE 1. Applications of Sentiment Analysis (Adapted from Cacheda, 2019, 2)

The use of sentiment analysis can be observed in different paradigms of the business. These paradigms are generally used by the strategy building department as the core reason behind the defining of the

success behind a certain the business. The medical industry has been reported as one of the major businesses in which the mental condition of the consumer is considered as an important factor based on which the user can relate with the product (Cacheda, Fernandez, Novoa, & Carneiro, 2019). To gain a better understanding of the underlying causes of a specific event in society, it is important to identify the factors contributing to it. This understanding can then be used to make informed decisions and effectively promote products by targeting the consumer mindset. Likewise, these can help in presenting the needful factors through which it can be noted that what features are good for the promotion.

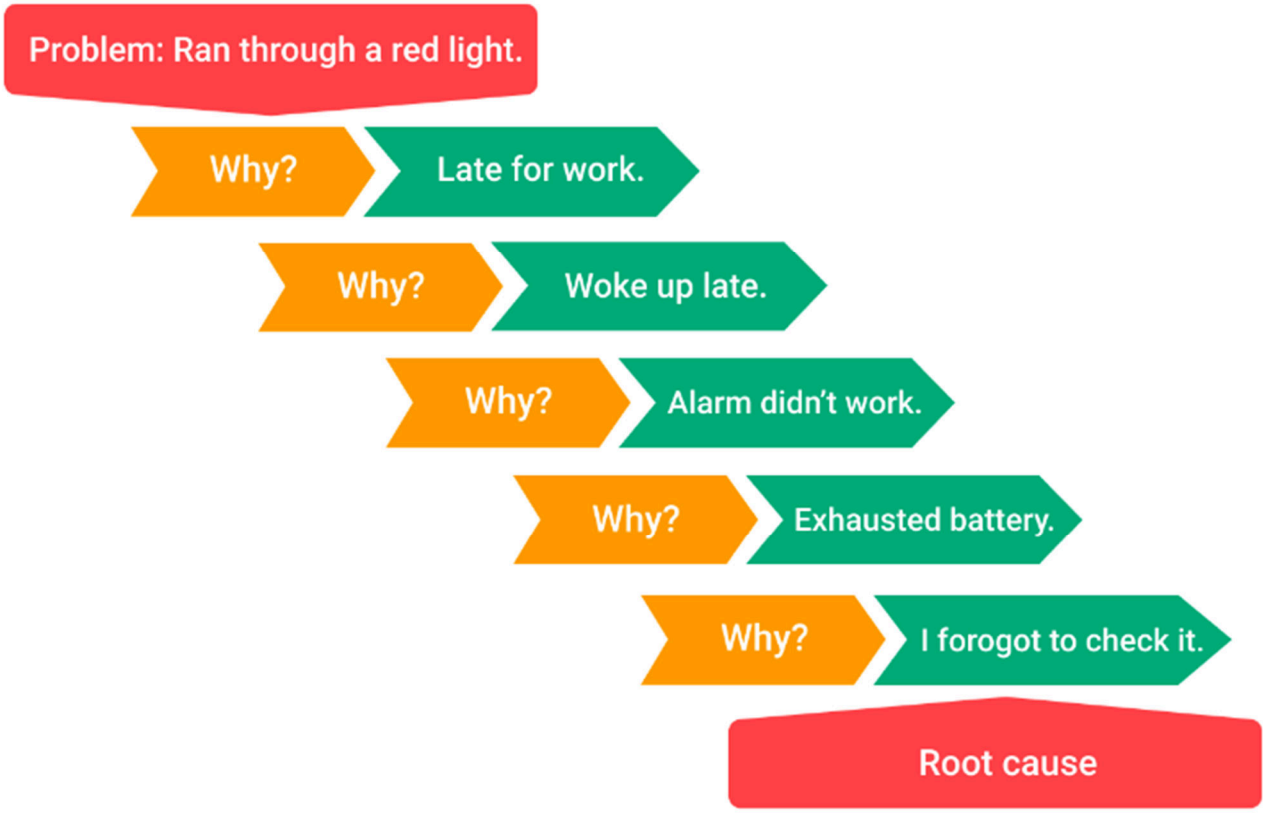


FIGURE 2. Root Cause Analysis using Sentiment Computation (Adapted from Chiong, 2021, 3)

For the success of the business, it can be noted that the movement with the mindset of the consumer is important. These provides an overview of the understanding related to the consumer understanding of the business as shown in (Figure 2.) For a better understanding, it can be noted that the business generally focuses over the providing of the factors through which the identification of the features can be provided for the user by which it reflects the factors related to the relevant processing factors in the system. Through this analysis it can be noted that the defined features of the user can help in shaping

the strategy based on the root cause behind a certain event. The points behind the defining of the factors can be used as the feature set by which the user can get relevant factors to report the success of their business.

1.2. Root Cause based Correlation Computation

The concept behind the correlation analysis can be understand in a way that each factor depends over a certain other factors that are considered to be driven methodology for causing a certain event (Chandra, 2022). It enables the defining of the highlights by which it can be noted that the user can take particular decision in understanding of the reason that are driving the impact related to the user. The industry of different businesses provides the cross-selling and up-selling schemes by enabling the user through which it can be understand that the user can help in reflecting the approaches by which the generation of the factors can be defined to report the most intelligent strategies that can be adopted by the user to depict intelligent decisions (Chiong, 2021). This decision taking approach is mainly depends over the root cause analysis and identification of the correlation to reflect the possibilities of the outcomes as shown in (Figure 3.) below:

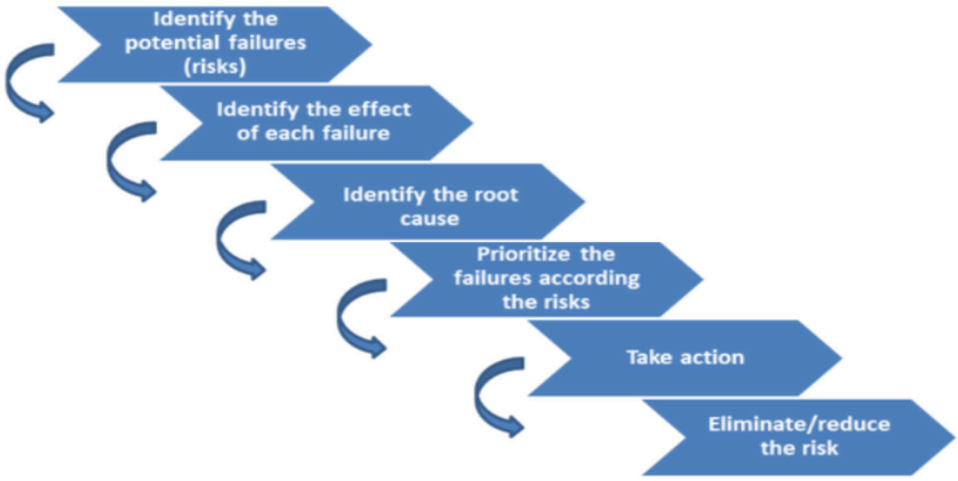


FIGURE 3. Root Cause identification with Correlation (Adapted from Chiong, 2021, 4)

1.2.1. Event based Sentiment Change Impact

The pandemic of the Covid-19 was not just a threat for physical health but it also affected the mental health of many people around the world. Due to the quarantine imposed during the pandemic, most

people were confined within their homes and this resulted in the rise of loneliness and depression as humans are social creatures and require interaction. This quarantine increased a state of emotional anxiety, stress, and depression among the people around the globe, which they expressed on social media posts. Social media platforms have significantly influenced the manifestation of users' melancholy condition. Analysing these expressions might offer valuable information to avert undesirable circumstances. (Kim, Park & D.M, 2022). This analysis is very important because it helps in gaining deeper insights into the depressive states of the individual and facilitates advance preventive measures. Additionally, the COVID-19 virus has caused unprecedented anxiety and uncertainty, sparking discussions about mental health online. By analyzing social media data, this study aims to uncover the ways in which people report depressive symptoms, suicidal thoughts, and negative thoughts in the digital space during transmission. By identifying early signs and risk factors of depression online, policymakers, mental health professionals, and social media can create targeted interventions, increase mental health awareness, and encourage community support. Overall, this study helps us understand the relation of mental depression and social media in emergencies, providing a better understanding of how to address people's health needs in the digital environment (Li , 2014).

To investigate causes related to suicide, negative speech during COVID-19 pandemic, self-harm expressions such as "hate myself" keywords, social media discussions regarding mental health during COVID-19 lockdowns, as well as social media texts containing depressive texts, and data on suicide during COVID-19 lockdowns, we will be analyzing multiple datasets containing tweets and retweets on these topics. The objectives for this work includes the followings:

With the help of sentiment analysis, this study will find correlation among language patterns found on social media as well as depression by using a comprehensive, in-depth analysis of depressive social media posts. Evaluating the texts and patterns between the two terms in social media posts, there seems to be a complex correlation between linguistic features and depressive symptoms, which needs further explanation. Furthermore, the research also tries to gain a better assumption of how different linguistic expressions, sentiments, emotion, and tones have the ability to reflect on social media posts of depressed individuals by analysing their words and expressions in the social media text patterns as shown in (Figure 4.) below:

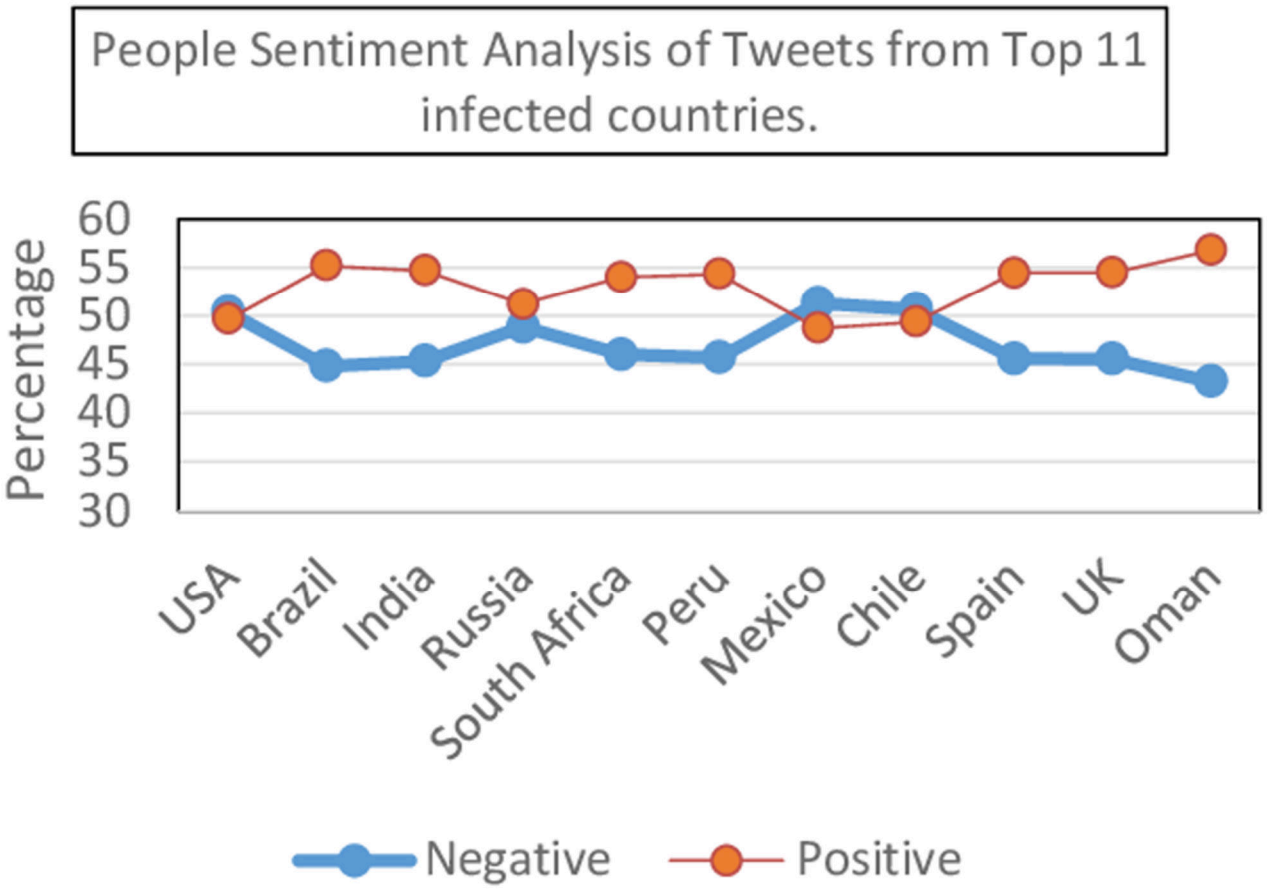


FIGURE 4. Sentiment Reported during Covid (Adapted from Li, 2014, 3)

Different concepts have been reported for the defining of the process through which the reporting of the sentiment analysis based over the root cause identification can be performed. These approaches have been listed below for the better defining of the process by which the use of correlation can be considered in the work.

Keyword Analysis involves identifying specific words and the phrases which are mainly considered to be strongly linked with the computed positive or negative sentiments.

Topic Modelling is a statistical approach used to find abstract topics and themes in a collection of texts.

Sentiment Correlation Analysis mainly identifies the relationships between different sentiment categories and the dimensions within social media data.

Temporal Analysis: It examines how sentiment has been changed over time based on the identifying of temporal trends and patterns.

Social Network Analysis depicts the observation of the relationships and interactions between individuals and the entities within a social network.

This workflow is depicted in (Figure 5.) below:

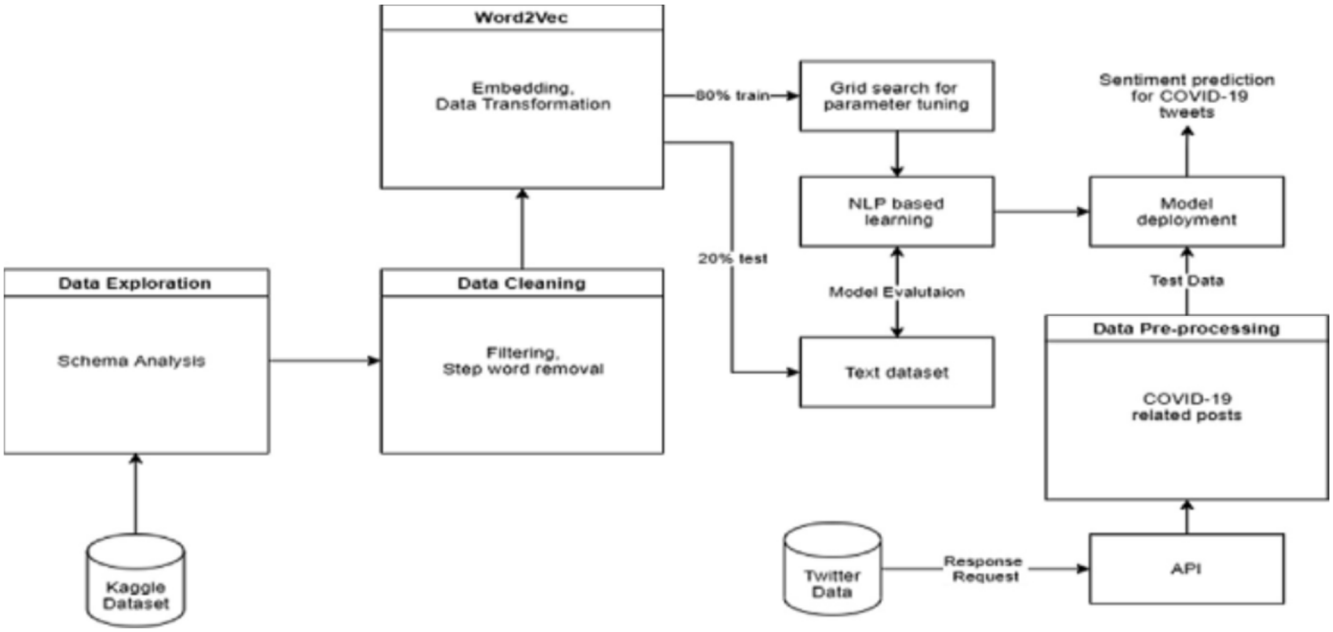


FIGURE 5. Process for the Sentiment Computation (Adapted from Chiong, 2021, 6)

1.2.2. Correlation based Language Pattern Detection

Correlation-based social media language pattern detection is recognised as an advanced method for analysing relationships and reporting linguistic features and trends in social media patterns. The text describes the correlation analysis in the factor reporting scheme, which helps identify the influence and sentiment expression in social media platforms. This analysis may be used to report the various mindsets of consumers. To this reason, the steps has been listed below to depict the process through which the patterns can be identified for the purpose of reporting the correlation:

Data Collection step is considered as the first process which requires the collection of the data based on which the forthcoming steps can be performed. The use of social media APIs can be defined for this purpose to get this data.

Preprocessing of the data extracted from the social media is mainly in raw format and requires text analysis by which the clean format of the data can be reported.

Correlation Analysis based on the clean data, the use of feature extraction process can be performed over the pre-processed textual data. It provides the inclusion of the sentiment scores and the relevant measures through which the user can get the topic modelling and identify the link of each event with certain factors.

Pattern Detection step presents the identification of the pattern in the text defined by the user. It includes the reporting of the change and trends that are reported from correlation analysis
Insights Development step presents the insights of the language can be presented through which the user sentiment score can be considered for the usage in the system to depict public opinion over a certain type of the event. The work flow is shown in (Figure 6.) below:

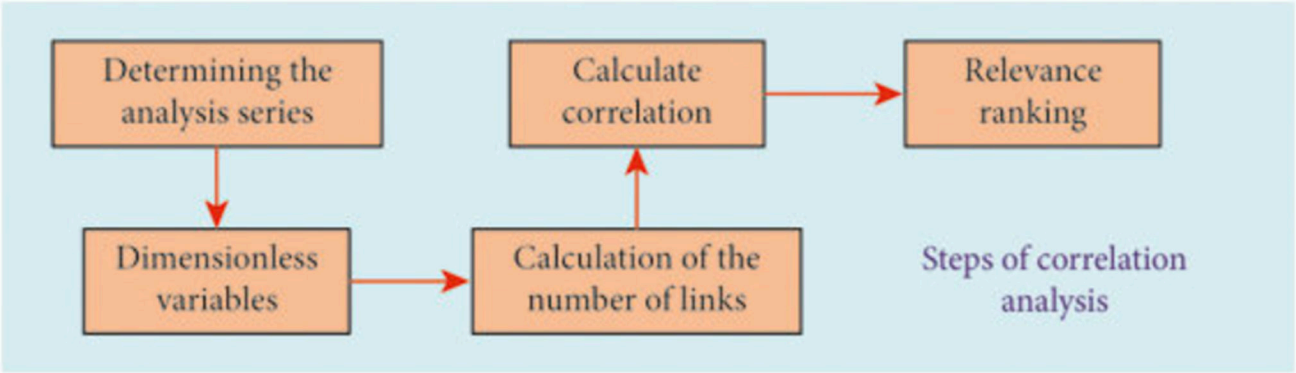


FIGURE 6. Correlation Analysis Process (Adapted from Li, 2014, 5)

1.3. Social Media Views Analysis

In order to have a better grasp of how these evaluations might be analyzed, it is essential to comprehend the process of combining user perspectives from social media. The purpose of these evaluations is to give users with a baseline for providing variables that should be included in the job. It is because of them that users are able to freely express their thoughts and feelings on various occurrences. The purpose of this section is to offer several methods for reporting the process by which user evaluations can be taken into consideration for analysis. The purpose of these reviews is to determine the user's feelings and attachments to events based on their subjective experiences. This contributes to a better knowledge of the processes involved in incorporating user perspectives into work.

1.3.1. Process for Reviews Analysis

The process of review analysis is mainly based on defining different techniques for extracting insights from customer feedback and opinions across different platforms. To understand the different types of

review analysis, the following points have been defined to outline the possible processes for analyzing these reviews.:

Sentiment analysis: This scheme is considered as the most frequently reported approach for the reviews analysis. It involves the method related to the gaining of the sentiment polarity that has been presented by the users within textual data. The reviews are categorised into three main groups: good, negative, or neutral. Sentiment analysis employs natural language processing (NLP) algorithms to discern the emotional tone of users on social media. By reporting this sentiment score, the businesses can gain a valuable insight for the purpose of defining the patterns that are required to be defined for the user.

Aspect-based analysis scheme is considered as a type of approach through which the defining of the process for sentiment computation can be reported. It provides an overview of the process by which the sentiments of the user can be considered based on the focus related to the specific aspects and the features that are provided by the user reviews. This scheme helps in defining the reviews overview by which the identification of the individual attributes can be defined through which the products for the quality, pricing, customer service, and performance can be considered from the user perspective.

Comparative analysis provides the process related to the using of comparing reviews and the sentiments that has been reported over the time towards a brand. This approach helps the businesses to presents the trends, strengths, weaknesses, and competitive advantages that are mainly reported from the market related to their business. Comparative analysis provides the businesses the focus for making informed decisions and report optimized marketing strategies based on which the refinement can be taken place for their product offerings.

Opinion mining approach is sometimes referred to as entity-based sentiment analysis, which primarily aims to extract the opinions stated by users in reviews pertaining to specific entities of interest. This scheme helps in involves identification of the opinion-bearing words and the phrases that are associated with corresponding entities mentioned in the reviews by the users. Opinion mining provides interesting insights into customer sentiment that can help the user in enabling organizations related to the marketing strategies for the sake of enhancing brand reputation that may lead towards the drive customer loyalty.

1.3.2. Impact of Root Cause Identification

The importance of the root cause identification can be understand in a way that the brands and the market are correlated with each other. They mainly focus over the defining of the approaches that can be used as the baseline for the defining of the features that are important for the users. This can be consider in a way that the reporting of the sentiment analysis scheme can help in identifying the factors that are behind the change in the sentiment polarity of the user. The positive and negative features are use as the baseline to present the factors that can be dependent over the feature set to report the highlighting of the system (Nandwani et al., 2021). Market value refers to the fundamental value upon which price fluctuations may be measured. It allows users to analyse and report the impact of certain events on the market (Neelavathi et al., 2021). Similarly, it involves determining the set of features that may be incorporated into the system to analyse user review reports and provide a solution based on the model. In order to achieve this, the user can utilise the system to illustrate the factor scheme that defines the generation of features in the system, therefore representing the remedy for the highlighted problem trend. (Pan, J., Liu, B. & Kreps, 2018).

FIGURE 7. User Emotion Correlation Analysis (Adapted from Ruz, 2020, 6)

Different factors have also been reported from the link of sentiment analysis with the correlation computation (Ruz et al., 2020) as shown in (Figure 7.) above. This may be interpreted as the calculated correlation reflecting the emotional perspective of the customer, which can be categorised as either positive or negative sentiment. The examination of emotional perspectives may also be applied to the development of a context-aware reporting scheme, which ensures the accuracy of the process by emphasising important aspects in the system.

1.3.3. Correlation of User State with Sentiments

In this section, the reporting of the process related to the defining of the user mind-set with the sentiment has been presented. It includes the defining of the feature set that is reported as the baseline for the highlighting of the factors for the inclusion in the analysis process (Tillman., March, Lavender, Braund, & Mesagno, 2023). To this reason, the process that has been reported as the core approach that is generally presented for the defining of the steps for the system are reported in the approach to depict the correlation based occurrence of the factors by which the generating of the solution can be ensured by looking over the sentiment analysis values reported from the user texts over social media (Yadav & Vishwakarma, 2020) as shown in (Figure 8.) below:

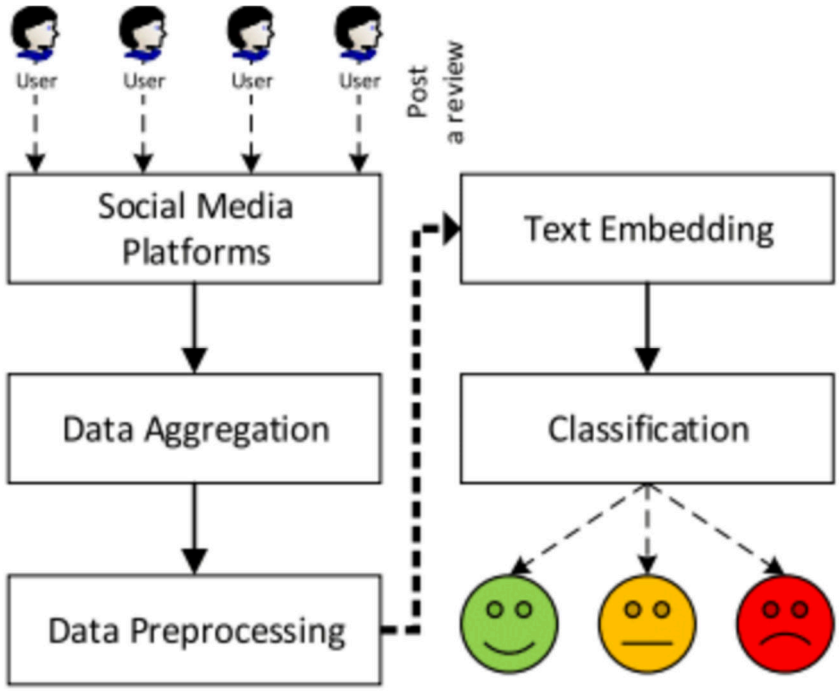


FIGURE 8. Emotion Analysis using Sentiments (Adapted from Li, 2014, 7)

1.4. Problem Statement

In this section, the formulation of the problem statement for this work has been presented. Social media platforms have been recognised as the primary means of disseminating specific reviews to the online community. The use of language patterns that are reflected by the user have direct relation with their mental health. It can be understood that the relevance of the user feedback with the provided reviews can reflect the mental condition of a person. The use of online feedback may be described as the means by which the relationship between the language used on social media and the user's health status

can be identified (Wankhade et al., 2022). It reflects the condition of the person on different parameters that includes the depression state or the happy state for the person. The major challenge in reporting the correlation between the mental state and the language pattern can be identified in a way that it requires the use of natural language processing techniques that can help in reporting a score-based pattern for reflecting positive or negative association.

An analysis of social media language and its correlation can assist in identifying patterns that may indicate a potential relationship between the language used and the user's mental condition, considering the intricate nature of the language.It reflects that extensive research efforts has to be performed in this area for the sake of identifying the patterns that can be used as the baseline for the developing of the features reflected for the defining of the conclusion. This conclusion driven approach is mainly used as the factor by which the analysis can be performed over the social media language process by which the mental condition of the user can be reported. The primary issue that has been defined as the major problem related to the extraction of the relevant linguistic features is the incorporation of the social media data. The data has to be retrieved from the social media API in order to report the elements that may be constructed using the issue statement for the project. To this reason, the use of different factors can be reported in the work that can help in defining the features by which the validation and the development of the process can be reported for defining the computational models for the sentiment analysis and the detection of the depression state. This can be noted in a way that the tradition approaches that has been performed over the defining of the solution lacks the accuracy through which the user feedback analysis approach can be improved. Therefore, it is crucial to document the methods by which the development of the correlation model may be achieved in order to demonstrate the connection between user feedback and the identification of system characteristics. Thus, the generating of the results has reported the need for the incorporation of the advance approaches that can be use as the baseline to compare with the traditional scheme to develop the effective strategy driven solution for the assessing of the sentiments by the users.

1.5. Research Objectives and Goals

This section presents the research objectives and goals set to define the study path. It includes the incorporation of the possible factors through which the generating of the aims can be defined for the work to ensure correct execution of the thesis. Using the research questions, the research will gain a better insights of social media language patterns and depression in order to create opportunities for the

development of more helpful strategies for detecting depression on social media platforms, intervening with them, and offering them support when they are experiencing it.

What is the correlation between the linguistic patterns seen on social media platforms and the intensity of depressive symptoms?

What are the limitations of using social media data for studying depression, and how do factors such as biases, and data sparsity affect reliability of findings from sentiment analysis studies in this study?

What interventions or support mechanisms can be implemented based on insights gained from analysing social media language patterns for depression, and how effective are these interventions in improving mental health outcomes?

What are the differences in language usage between individuals with diagnosed depression and those-without influence on the expression of depressive symptoms through language patterns?

How to identify the most crucial linguistic features for accurate classification of changes in sentiment, linguistic pattern, or language style over time?

How can researchers ensure privacy, confidentiality, and informed consent when analysing user-generated content for sentiment analysis?

How depression-related language patterns on social media are used for culturally sensitive sentiment analysis algorithms?

The study objectives have been formulated based on the research questions provided above:
RO1: To report the problem statement and the limitation in the previously reported works
RO2: To design and develop the scheme that can help in resolving the problems related to the factors related to the identification of the correlation analysis for the work
RO3: To provide the evaluation of the scheme by defining the feature set that can be taken into account for the defining of the feature set to depict correctness of the models

1.6. Research Aims

The task aims develop the valuable insights into how early detection, intervention, and support can be effectively deployed for individuals experiencing mental health challenges in digital environments

while providing valuable insights into effective strategies for early identification, intervention, and support. The primary purpose of including this study objective is to demonstrate the method by which the user may have a fundamental grasp of the approach that has to be described in the work in order to provide relevant variables. To this reason, the defining of the features in the system has been reported through defining the solution of correlation analysis with sentiment computation process.

1.7. Thesis Organization

This part presents the organisation of the thesis work to enhance user readability. It includes the defining of the information summary that has been presented in the upcoming chapters of this thesis work.

Chapter 2 provides the insights related to the problem statement and the limitations that have been identified from the past studies. It can help in defining the possible tools and techniques which are incorporated by the traditional research works

Chapter 3 outlines the research approach that has been used for the development of the proposed strategy. This methodology aims over the highlighting of the sequence through which the user can get the insights of the work

Chapter 4 chapter presents a thorough examination of the experimental setup and the specific features that were incorporated into the experiment. Additionally, it encompasses the process of determining the outcomes that have been produced by the trials.

Chapter 5 chapter contains the verified conclusion of the thesis study. It guarantees the achievement of the research goals and potential future paths.

2 LITERATURE REVIEW

This chapter provides an overview of the existing literature relevant to the thesis study for the reason of defining the traditionally adopted techniques for the work. Generally, the adoption of the correlation based techniques for the identification of the root cause has been defined as the base over which the user can present the features for the highlighting of the system. The user can get the insights related to the factors such as the previously adopted technique and the features that can be defined through the work processing approach in the system. The inclusion of the forthcoming section has reported the research works over which the generating of the problem statement factors has been defined.

2.1 Research Background

In today's times, the use of social platforms has attracted attentions of a lot of researchers, as the usage pattern and the behaviours of the user related to the activities from these platforms has become increasingly concerning. Here, from the use of social media, several behavioural patterns can be analysed which yields negative impact in the society, thus giving rise to much research. Several negative social activities including suicide from depression are prominent behavioural pattern among users. This gives rise to many questions in order to prevent such negative aspects from the society, which requires analysing the sentiment of the user through their expression in terms of words and language used in social media. Amidst the Covid-19 shutdown and epidemic, social media played a significant role in providing easy access to news and information available on the internet, which proved essential in our everyday operations. As a result of the lonesome circumstances and lockdown at the time, many people were depressed, and it was evident through social media platforms that they were expressing their sadness. Analysing these, one can easily find patterns of depressive words, and understand the sentiment of the user and take preventive measures appropriately. A review of 30 research studies were done on the subject of analysing behaviour or behaviour identification utilising the social posts from till 2017 (Wankhade et al., 2022). The study was mainly done by analysing the other studies and their method of analysis, where it was seen most of the study was done on Twitter posts. As part of their analysis, they also used a random user list to analyse the behaviour of the user, which was then pooled in the final analysis, which revealed that different social media sites displayed statistically important differences in multiple parameters with respect to their respective social media sites. The study related to the presented approaches has been defined over the years as shown in (Figure 9.):

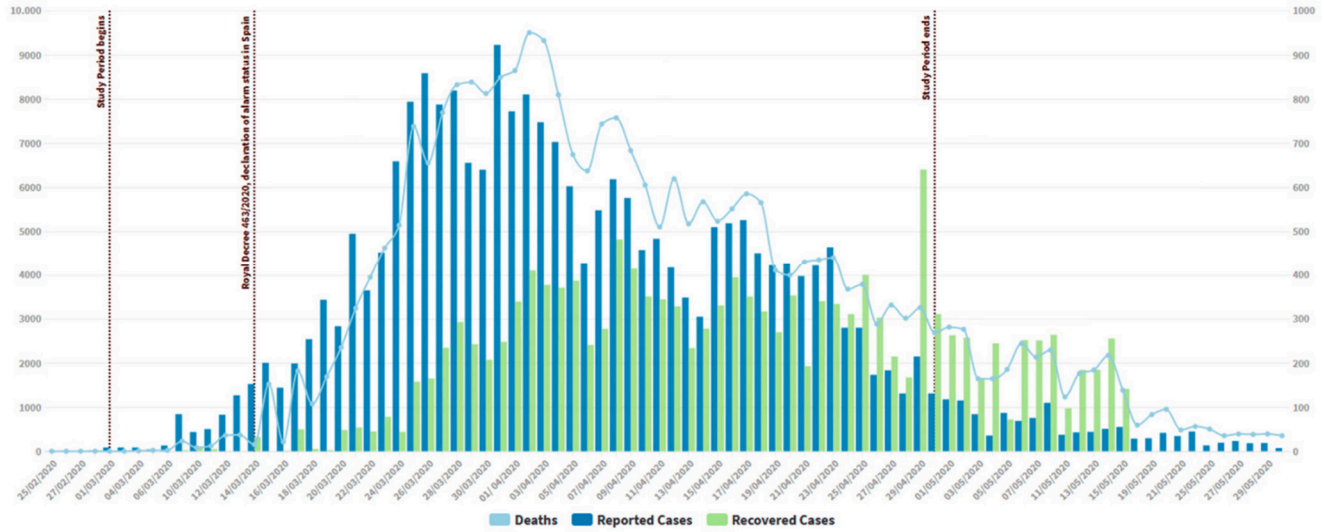


FIGURE 9. Research work conducted on the Covid (Adapted from Chiong, 2021, 5)

Sentiment analysis is the process of assessing and analysing individuals' perspectives, opinions, and perceptions of various behaviours, goods, issues, and services from several user sources. The posts by users becomes very important for concerned authorities for collecting data and information making decisions based on the opinions of users. During the process of sentiment analysis a researcher can face many challenges that has quite an impact on the capability to understand sentiment properly and find the proper polarity of the sentiment to implement in the analysis process. A sentiment analysis contains the identification and extraction of different kinds of information from texts, generally using NLP techniques and mining for text techniques. In this system there will be evaluation, comparison and investigation of different approaches inspired from various sources and data for a deeper interpretation of the strengths and weaknesses.

According to Nandwani and Verma (2021), the major objective of sentiment detection is to analyse the depth of the text containing opinions that lets one to separate and categorise them as various emotions based on how they were expressed. Part from finding the patterns of emotions in texts, the sentiment analysis with the help of emotion identification understand the patterns in text through recognising the mental state of the user with accuracy. When people express polarity or emotions, or when they interact with interactive systems, it has been found helpful to use techniques such as detecting the emotion and sentiment to comprehend the emotions that are displayed. Alslaity and Orji (2022, 1-26) in their study says that the machines that detects sentiments will give accurate natural detections and emotion identification that will assists in building systems that are human-centred and give an adapting change in behaviour and interventions based on users.

Sentiment analysis can be looked at as a subset of emotion detection, in that it detects the unique emotions than simply tagging them as negative, positive, and neutral statements. However, it is difficult to detect emotion in text because there are no cues to assist in the process., unlike in speech, which are cues such as the tone of the voice, the facial expression, the pitch, etc. In the study by Bharti et al. (2022, 1), it is mentioned that NLP techniques was proposed to identify emotions in texts: lexicon-based approaches, keyword-based approaches, and machine-learning approaches have all been proposed for this purpose: approach based on keyword, approaches based on lexicon, and ML approaches. Recent studies have proposed different approaches that can detect the emotional content of texts, such as keyword-based, lexical affinity, learning-based, and hybrid approaches that can detect the emotional content within texts based on their keywords.

Machine learning and deep learning models have been widely utilised in studies aimed at detecting emotions, with good results. According to the research study by Li and Xu, 2014, 1742-1749, the researchers are facing some challenges that should be addressed. For example, there is no way to extract semantic information from non-standard language, the feature extraction process is inefficient and time-consuming, it is hard to identify different emotions from non-standard language, and the datasets are imbalanced. For this purpose, they have used the LSTM model, Naive Bayes as well as SVM classifiers to find emotions using LSTM and Naive Bayes models.

Based on the study by Asif et al. (2020, 48), The researchers use sentiment analysis on social textual data to ascertain the frequency of depressed sentiments and classify the included textual perspectives into several groups based on the degree of extremism. First, domain experts design and evaluate a multilingual vocabulary that corresponds to the intensity weights. The lexicon achieves an accuracy of 88% on validation, as determined by domain experts. During the subsequent stages, various classification methods were utilised for the task of classification, including multinomial Naive Bayes and linear support vector classifiers. The Linear Support Vector Classifier, trained on a multilingual dataset, achieved an accuracy of 82% in predicting the categorization of the item.

Depression is a phrase commonly used to refer to a mood condition that can lead to serious mental breakdowns, self-harm, and suicide, among other consequences. According to research by Pan, Liu and Kreps (2018), More than 300 million people worldwide are believed to experience depression, a mood illness that impacts their mood, perceptions, thoughts, and everyday activities like sleep, appe-

tite, and work. Individuals afflicted with depression experience a comprehensive effect on several aspects of their lives, such as their emotional state, perspectives, thought processes, and the routine activities they partake in regularly. Depression mainly refers to describe the existing disorder in mood change of the user that results in mental imbalance, harming oneself and committing suicide. Millions of people worldwide suffer from depression, affecting their mood, thoughts, and daily activities.. The negative impact of depressive states on people's lives and daily activities can lead to misinterpretation of these states, resulting in incorrect treatment identification in several cases while evaluating this situation. There can be a misunderstanding about depressive states among people that might result in wrong treatment in most cases as well as generalisation in other cases when it is time for analysing the situation.

According to research (Kim et al., 2022, 8), there has been an increased curiosity in automated depression analysis methods in present years, due to rise of social platforms like Twitter, etc. The incidence of mental illness is lower than the incidence of physical diseases, because patients suffering from mental illnesses have no idea what their illnesses are and do not receive treatment because they lack motivation. Additionally, they do not realise
that mental health services can improve their problems to some extent. When mental illness is detected and treated at an early stage, it is often possible to get a full recovery, however, if treatment is delayed, the consequences can be even worse. This is why awareness and identifying depression are integral for treating mental illness. Further, by recognizing and identifying the disease, as well as knowing its precise name, there is a higher chance of early detection of the disease as well as a greater probability of successful treatment as shown in (Figure 10.)

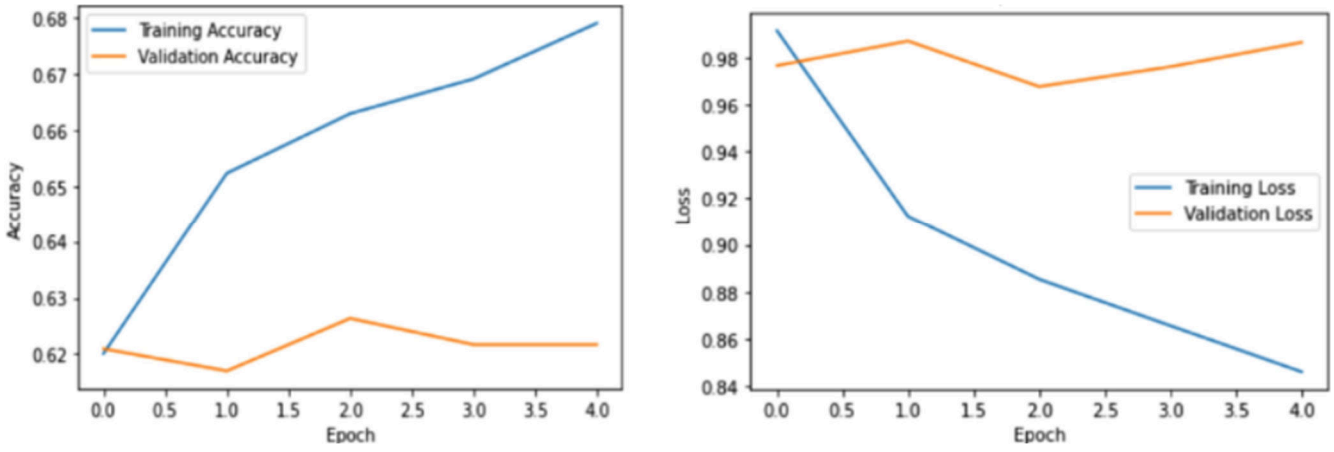


FIGURE 10. ML based Accuracies Comparison (Adapted from Chiong, 2021, 138)

A research carried out by Cacheda et al. (2019, 21), uses the data available on Reddit to examine different methods that can provide early detection of major depressive disorder MDD utilising artificial intelligence, based on the results of which two methods were used. In both algorithms, the subject's depression condition can be predicted using textual, semantic and writing features (WFs) In addition to machine learning algorithms, all of which are derived from machine learning algorithms. The first approach employs a solitary machine learning algorithm to forecast instances of depression, whereas the second model use two machine learning algorithms to anticipate non-depression occurrences. There are two algorithms developed, one for predicting depression cases and one for predicting non-depression cases, in which the first algorithm is trained to predict depression cases. In the subsequent study, the study examined the effectiveness of each model in terms of early detection and late detection, as well as how time-aware the approach was, and the results of that evaluation uncovered that a dual model can improve detection up to 10%. Furthermore, the methods used to conduct the study were implemented using freely available software tools, which facilitates the reproducibility of the research conducted.

Here, in their study by Chiong et al., (2021, 137), by applying machine learning classifiers that are used to solve prediction problems, the study solved prediction problems. These methods were selected based on their excellent prediction performance and their ability to predict a variety of situations. The functions of the models, depends on the data used to train the system. The function of the generalised method has been analysed against different and globally available sets of data containing social-posts and text from various sources, including social platforms and the function of the normal approach against daily posts. To construct the detection model, ML algorithms are used, along with textual-based features, which are used because they extract input features from the text itself; because of this, they are more independent than feature extraction processes based on the system; they are therefore free from the machine than different feature extraction system relying on the system. The aim of this study is to determine the activity of the models in the study utilizing the datasets from multiple social posts in an attempt to identify that the given method that relies on portions of the exact datasets for training and experimenting the machine learning algorithms, as well as those for exact keywords, will work.

Specifically, the study (Chiong et al., 2021, 56) applies text preprocessing techniques, like tokenization, eliminating stop word, elongation word correction, and part of speech (POS) lemmatization, in order to facilitate the detection of negation words, correct elongation words, and remove illegible

words. It was determined that n-gram words (from unigrams to trigrams) as well as the count vectorization method would be effective in extracting the input features based on the bag-of-words (BOW) features extraction technique. By taking a normal approach for detection of sentiment like depression which is effective, when social platform users do not have any knowledge of their depression or denies it, the study assists in the research findings of machine learning, NLP, and mental health problems. This in combination with the aspect of mental health issues has resulted in a significant contribution to research field of machine learning, NLP and even mental health as shown in (Figure 11.)

FIGURE 11. Root Cause Finding Technique (Adapted from Neelavathi, 2021, 138)

According to the study by Neelavathi et al. (2021, 134-139), research centres state that 72% of people at this time use different types of social platforms, and over 300 million people experience depression on regular basis; only a small percentage of them receive sufficient treatment for depression and despondency on a regular basis. Nearly 800,000 people routinely lose their lives as a result of suicide based on the World Health Organization, thinking about that suicide is a major reason of death among teenagers. It is recognized as discouragement is the leading cause of depressed people throughout the world. It is well known that Social Media offers a very strong opportunity to aid early depressions, especially among young people. One week, the time it now takes for users to send a billion Tweets. 50 million. The average number of Tweets people sent per day, one year ago. 140 million. The average number of Tweets people sent per day, 456. Tweets per second (TPS) while over 200 billion tweets are being shared annually on social media (Twitter Official Blog., 2011). This rich source of information will enable the design of an efficient model, which will provide a report of the person's symptoms of depression based on the data and information provided. A social media analysis model, which

can examine the Tweets expressed by people that they self-assess to have negative characteristics, can be devised by analysing the linguistic markers in the Tweets as shown in (Figure 12.)

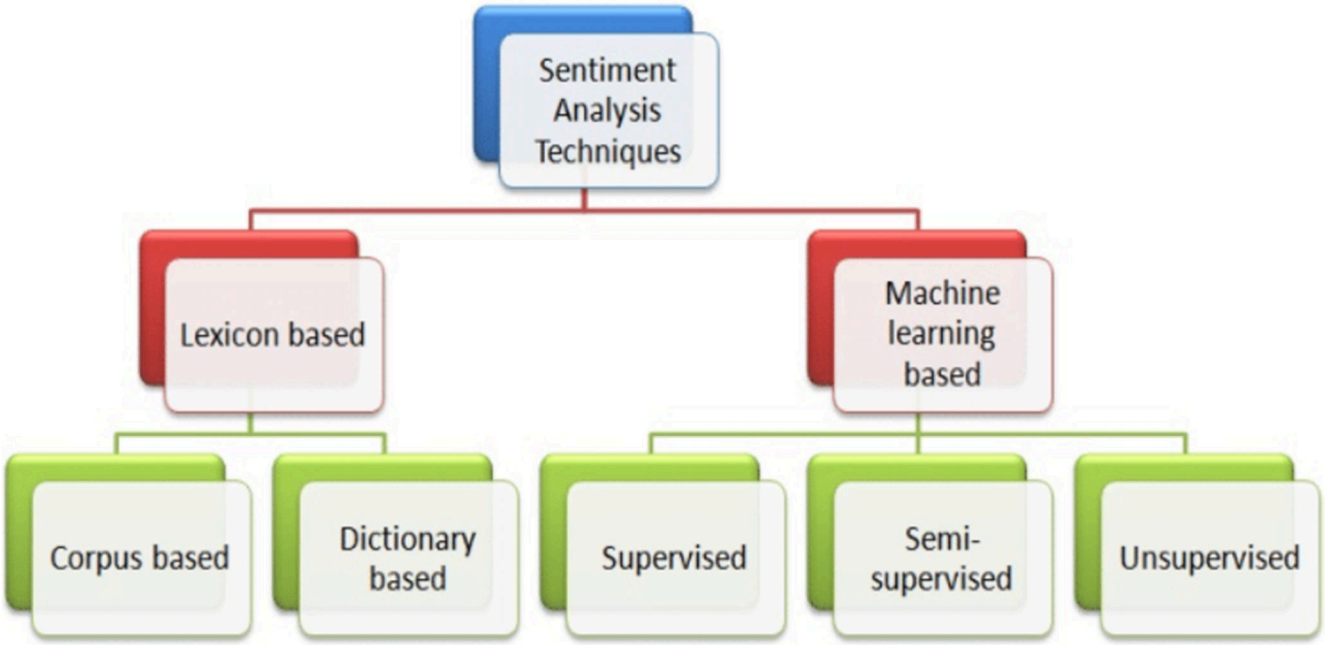


FIGURE 12. Existing Approaches for Sentiment Analysis (Adapted from Li, 2014, 8)

Among all the social media platforms that were explored, Facebook emerged as the most popular social media platform out of all the others. While Facebook is a widely used social media platform, it also needs to be explored in the contexts of sentiment detection, NLP, computational text identification, analysing texts, and applications, examining the sentiments and emotions of users. The study used the methods of Multinomial Naïve-Bayes estimation, SVM (Support Vector Machine) estimation, K-Nearest Neighbors Classifier estimation, Decision Tree estimation, & Random Forest estimation for the purpose of determining the best classifier for an unstructured text from social media posts, in order to determine the most appropriate classifier for this text. For example, there has been a study conducted for evaluating functions of a Random Forest model for categorising Facebook comments and unstructured textual data, as well as in predicting the performance of a Random Forest on a classification task. In case of maximum accuracy, the random forest can analyse text bodies, like comments made on Facebook posts, with efficient performance. The summary of the articles has been given in (Table 1.) below:

TABLE 1: Existing Studies Summary

Author	Contribution	Advantages of Work	Limitations	Conclusion
Nandwani and Verma (2021)	Analyzing sentiment depth in textual data	Provides insights into sentiment analysis and emotion identification based on textual data. Offers a methodical approach for understanding user sentiments and categorizing them based on expressions.	Limited focus on specific social media platforms or datasets. Lack of consideration for contextual factors influencing sentiment expression.	Understanding sentiment depth is crucial for extracting meaningful insights from textual data, aiding decision-making processes and interventions. Implementing effective sentiment analysis techniques can enhance user experience and inform targeted interventions.
Bharti et al. (2022)	Proposal of NLP techniques for emotion detection	Introduces various NLP approaches for emotion detection in textual data, including lexicon-based, keyword-based, and machine-learning approaches. Provides a comprehensive overview of techniques for identifying emotional content in texts.	Challenges in emotion detection from text, such as lack of cues present in speech. Difficulty in extracting semantic information from non-standard language. Imbalanced datasets and inefficiencies in feature extraction processes.	Emotion detection in textual data is a nuanced process requiring sophisticated techniques. By exploring various NLP methods, researchers can develop more accurate models for understanding emotional content, leading to improved sentiment analysis and user engagement.
Li and Xu (2014)	Application of machine learning for sentiment analysis	Supports machine learning models for sentiment analysis, reporting successful textual emotion detection. Highlights the challenges faced in feature extraction and dataset imbalance.	Difficulty in extracting semantic information from non-standard language. Inefficiencies in feature extraction processes and dataset imbalances pose challenges.	Machine learning models offer promising avenues for sentiment analysis, providing accurate detection of emotions from textual data. Overcoming challenges in feature extraction and dataset balance is crucial for enhancing model performance and facilitating more effective sentiment analysis.
Asif et al. (2020)	Sentiment analysis of social textual data	Conducts sentiment analysis of social textual data to categorize sentiments related to depression. Utilizes a multilingual lexicon and	Reliance on classification algorithms may limit the flexibility of sentiment analysis.	Social textual sentiment analysis may discover and categorise depression-related sentiments. Machine learning algorithms and multilingual

		classification algorithms for sentiment classification. Demonstrates the effectiveness of linear support vector classifiers in predicting sentiment categories.	Challenges in classifying diverse sentiments accurately.	lexicons can improve sentence classification, enabling early mental health detection and intervention.
Cacheda et al. (2019)	Detecting serious depression early using AI	Looks at Reddit and AI strategies for early depression identification. Propose text-based machine learning algorithms for depression prediction. Demonstrates improvements in detection accuracy with dual-model approaches.	Reliance on Reddit data may limit the generalizability of findings. Challenges in accurately predicting depression cases based on textual features.	Early detection of depression using AI holds promise for improving mental health outcomes. Leveraging textual features and machine learning algorithms can enhance the accuracy of depression prediction, facilitating timely interventions and support for individuals at risk.

2.2 Research Limitations

Different studies have analysed the relation between the social posts language patterns and the depressive mental states through conducting sentiment analysis. However, there are many research gaps in the existing studies and methods that exist in the nature and depth of the study while analysing the data. Due to the existing research gaps, it is very crucial that the datasets are mostly accurate and have an extended scope for research, otherwise, it will result in the inadequate evaluation of consider nature for the depression. It also demonstrates over the related issues that are expressed in such posts. Here, most of the datasets that are used in the study of the connection between the words and texts used in social platforms mainly focus on some of the unique and specific keywords related to mental health on the posts with negative references of depressive symptoms. Most of these datasets are mainly based on the negative wordings and texts related to the mental health, as an example, when a datasets contains some social posts that has references to self-harm or involves suicidal tendencies, this will help in detecting the texts and the users who are depressed.

Furthermore, datasets related to hate speech during the COVID-19 pandemic, or those that contain keywords associated with self-hatred, can also provide valuable insights for understanding the relationship between user's negative views of themselves and their mental health in order to use these data to interfere in and prevent any unwanted situation. In addition, Post-COVID depressive social media

texts and datasets provide insight into how the level of mental health has deteriorated in social media platforms, so they are relevant to understanding how social media users are feeling psychologically as a result of the pandemic and lockdown. In the same way, there are datasets that are related to the Covid 19 lockdowns also provides some good insights to the researchers to find out how lockdown and the pandemic overall impacted the lives of the users who has exhibited symptoms social media postings about depression (Tillman et al., 2023, 698).

Despite the fact that the datasets provide valuable insights into different aspects of depression as well as the related issues, they also present several limitations that actively contribute to the research gap identified as a result of this literature review. There is a main limitation in this very approach which is the possibility of bias in the selection of data. This is because the datasets maybe presented in a way which only captures the posts directly referencing depression and not other insignificant yet serious posts. Here capturing the data of those who are mostly active users may often result in missing out those who are not that much active or vocal about their depressive states and rather expresses their conditions indirectly that does not directly reference to depressive states. This can often result in over-looking and misunderstanding which might result in not diagnosing the users who needs immediate help. Again, the excessive dependency of the dataset which contains the references for depressive states and the simple relations with the keywords and will result in avoiding the very simple and easier expression of depressive posts from different demographics, cultures and places based on their words, texts and patterns. Now, for example, users from the marginalised groups or cultural communities have different language expression or texts that are used to explain their experiences. In previous studies, this could have been overlooked resulting in research gaps.

In other cases, not including the datasets with various kinds of topics and wordings that are not directly related to the depressive states of users, may result in the limited findings that leads to the less comprehensive study of the data and the complexities used in the posts language regarding depression and mental health of social media users. Therefore, using data with wide range of information, texts and insights can be beneficial in filling the research gaps of the study as shown in (Figure 13.)

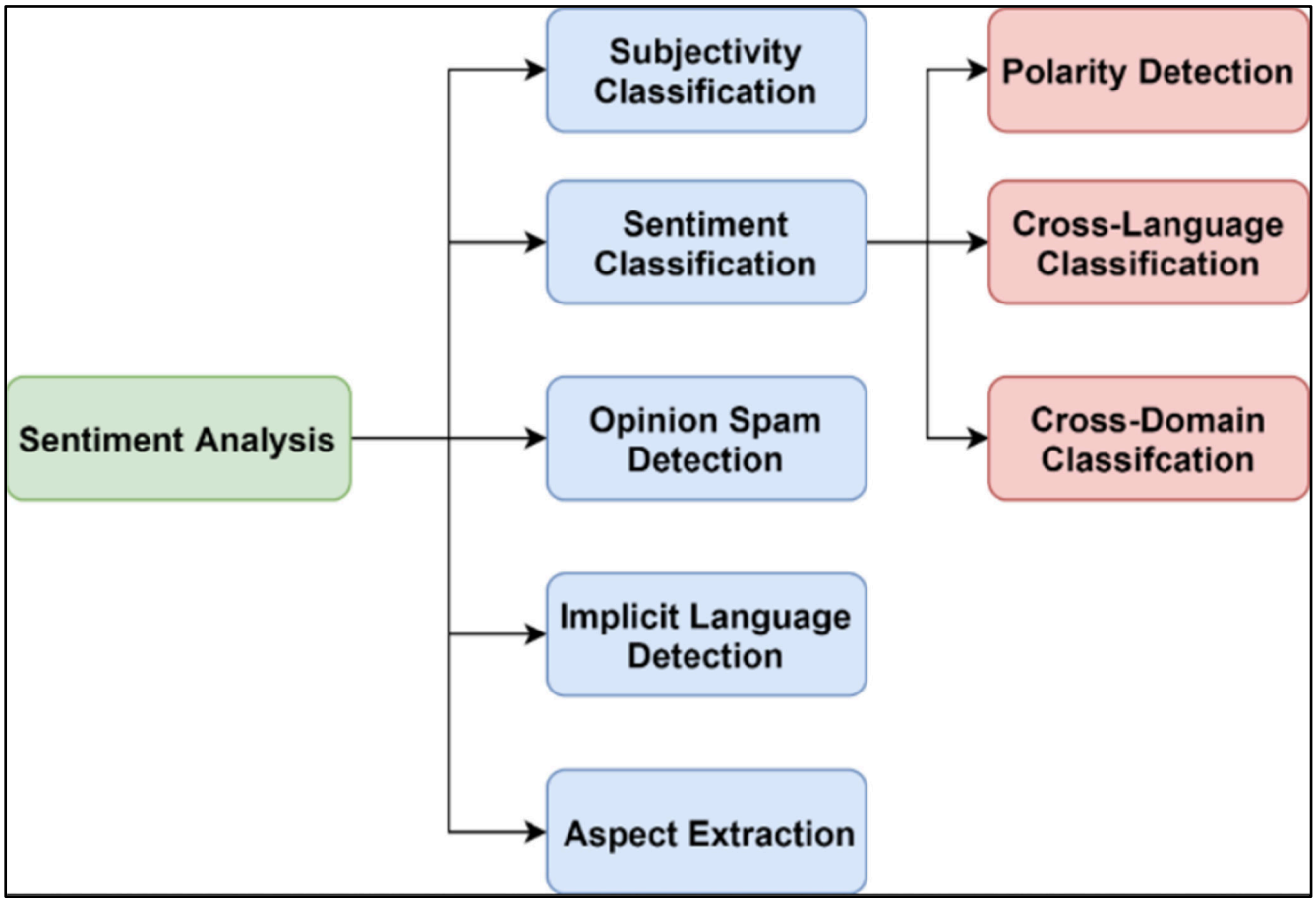


FIGURE 13. Trends in the development of the Sentiment Analysis (Adapted from Chiong, 2021, 136)

One possibility of getting a more comprehensive understanding of the social platform posts exists where researchers will be able to relate the language and word patterns by incorporating large datasets and finding internal similarities in the consistency of words and analyse them overall. Additionally, the efforts to reduce selection bias and ensure the proper evaluation of the datasets are very important for improving the validity of the findings in this study.

3 RESEARCH METHODOLOGY

This chapter presents the debate on the research approach of the thesis work. The chapter elucidates the principles utilised in the thesis study to explicate the experimental procedure. This research methodology has been considered as the root cause by which the proposed scheme has been presented. It also includes the defining of the architectural diagram based on which the proposed scheme design has been presented. The forthcoming section has been defined in a way that it elaborates the focus of the work by explaining the approaches and the predefined schemes that has been reported by the thesis. In this way, the generated proposed scheme has been considered for the work through which the reporting of the thesis journey has been considered.

3.1 Research Methodology

Regarding the establishment of the research approach for the study, it is important to consider the scheme in a way that the proposed methodology has been built by considering the reports that are previously defined in the research efforts. These research schemes have been taken as the base over which the new proposed methodology has been designed. For this reason, the working process has been started by defining the fundamental scheme of the work through which the user feedback has been taken into account for the reporting of the process to depict the information details. The processing scheme is defined in a manner that takes into account the created factors for the development of the system's performance over time. For this reason, the adoption of the design science research methodology has been considered for this work. This research method mainly elaborates the working phenomena through which the user can define the approach to depict the resultant of the work. In this way, the research scheme that has been used for the work has been started by defining the research problem that has been extracted from the past studies. This research problem has been defined as the limitation of the previous works and based on which the generating of the possible future direction has been presented. It also includes the elaboration of the methodology through the presenting scheme by which the user can get the insights over the data. The obtained data can be further analysed to generate variables linked to the information processing of the data. The pristine data may be utilised to demonstrate the correlation between the factors associated with the emphasis on the scheme, hence illustrating the model's performance. The computation of work accuracy is determined by the presentation of aspects that demonstrate how experimental details are reported. This has been given in (Figure 14.) below:

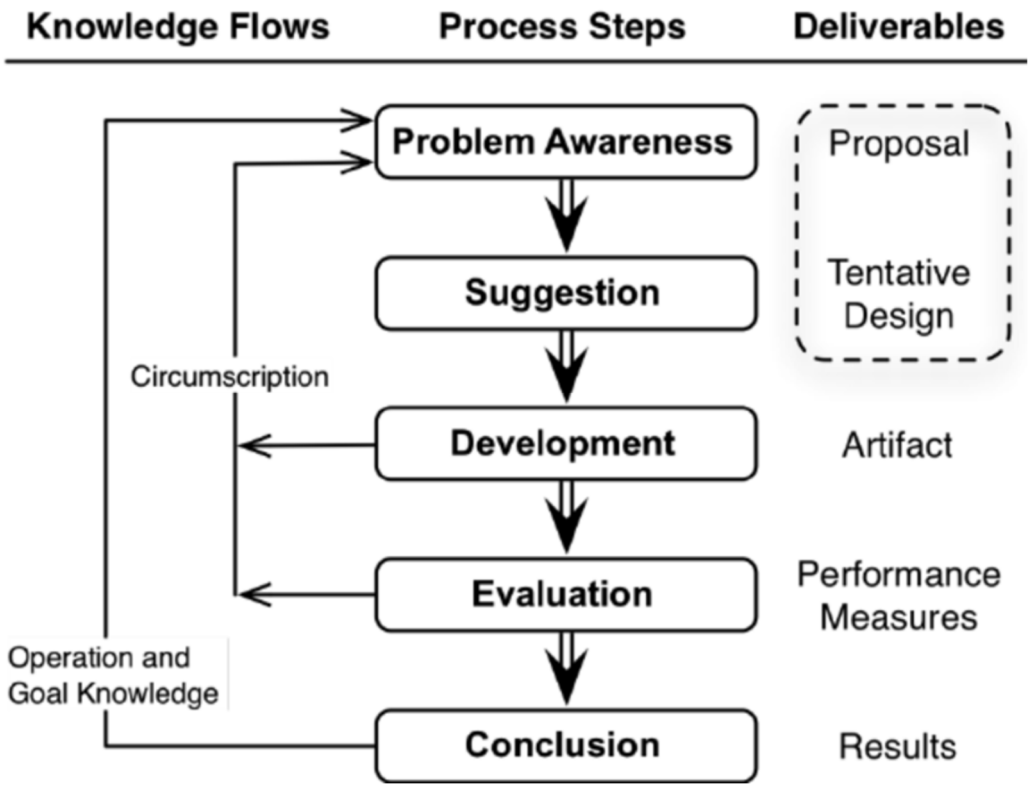


FIGURE 14. Design Science Research Methodology

It can be observed from the above given diagram that the information flow has been defined in a way that the generating of the research problem in combination with the development of the research design and implementation can be presented for the thesis. This study endeavour has been further delineated in order to present a comprehensive assessment of the work. Therefore, the assessment outcomes have been utilised as the benchmark for the inclusion in the artefacts.

3.1.1. Problem Defining

The research methodology starts with the defining of the problem statement have been considered as the baseline over which the formulation of the details has been presented. It includes the defining of the issues that can be arised in the work related to the elaboration of the factors through which the user can highlight the scheme to depict the features of the work. The problem statement has been defined from the previous work analysis by identifying the limitations and the gaps that has been reported by the authors. These research gaps have been considered in the work by which the formulation of the

new research scheme can be presented. Furthermore, it illustrates the characteristics of the system by demonstrating the aspects that allow the user to emphasise the operational structure of the system.

3.1.2. Research Scheme

The research scheme defined for the thesis has been elaborated by showing the factors by which the user can get the attributes related to the model development. In terms of sentiment analysis computation, The analysis of the characteristics may be described as the process of elucidating the factors in a manner that allows the model's reporting to be defined for the user, whereby the explanation of the user vectors can be reported to establish the methodology of the system.

3.2 Research Design

The research design has been given in the form of an architectural design of the work, as follows in (Figure 15.) below:

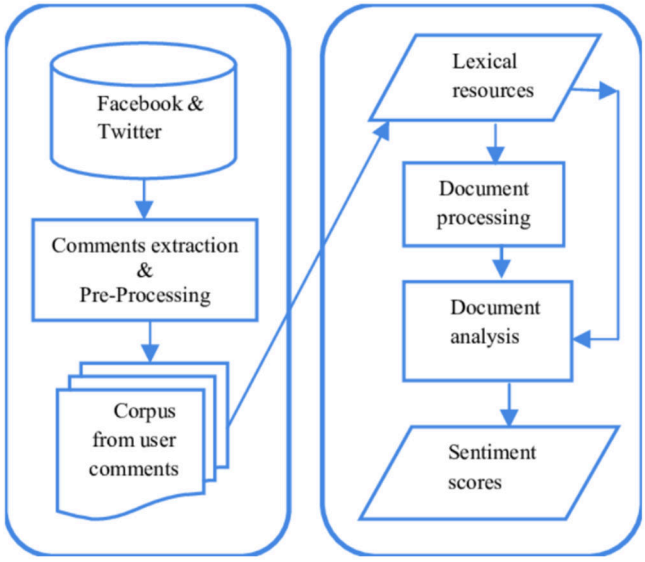


FIGURE 15. Research Architecture Design

The research strategy of the paper is characterised by the systematic approach used to gather and analyse data using the Twitter API. The data has been gathered by identifying the elements that the user has reported, specifically in relation to developing the operational framework for reporting comments and extracting tweets. The tweets have been processed to provide the resulting data insights for the user. This may be understood as representing the entirety of the data, which can then be analysed to

illustrate the outcome of the characteristics. The processed data has been created to display the machine learning-based analysis, providing a report on the document processing strategy in the system. This can be noted that the generated results have been defined for the highlighting of the sentiment score that can be used as the base over which the feedback can be elaborated. It can also present the processing scheme in which the resultants can be based over the defining of the correlation model. The finalized module of the work will be focused over the highlighting of the features by which the usage of the scheme can be presented to depict the working paradigm of the root cause in the system.

3.3 Data Collection

The process of data collection has been presented in a way that the Twitter API has been used as the baseline through which the hashtags and the details of the data has been gather from different tweets to depict the resultant of the work. This scheme has been elaborated as follow:

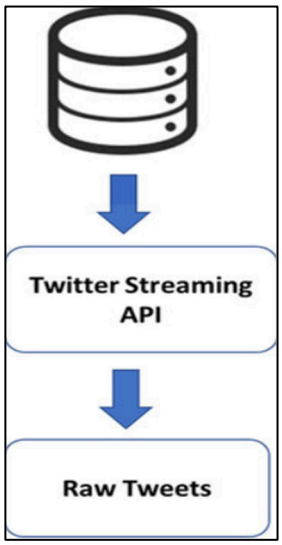


FIGURE 16. X API Extraction (Adapted from Chiong, 2021, 135)

The extraction of the data from X has been presented in a way that the developer account has been considered for the reporting of the use case by which the extraction process has been defined to report the system implementation The visualization of this scheme has been shown in (Figure 16.)

4 EXPERIMENTAL DETAILS AND RESULTS

The chapter has been offered to demonstrate the experimental details, including the experimental setup and the outcomes acquired from the study effort. It should be noted that the development process of the model implementation has been defined in the work for the purpose of elaborating the technique by which the data can be processed for the user feedback. The process of sharing thoughts over the social media has been identified as the most frequently used technique for expressing the mindset of the user towards a certain event. This process defining scheme has been reported for the elaboration of the factors by which the user can get the highlights related to the information processing of the model. In this way, the generated factors that has been used for the highlighting of the information processing has been defined with different steps to elaborate the performance of the system.

4.1 Experimental Details

The process for the defining of the experimental details has been presented in several steps that includes the elaboration of the features by which the user can report for the work. This can be noted that the defining of the work has been reported for the showing of the experimental work along with the factors related to the sentiment analysis of the data. The analysis of the tweets has been delineated in relation to demonstrating the application of the model. The next parts will demonstrate the experimental phase of the work.

4.1.1. Experimental Setup

The procedure has commenced by establishing the experimental configuration for the task. The use of model development is characterized by its ability to demonstrate the characteristics that enable users to extract the emotions of tweets from processed data. The information provided for the depiction of the work encompasses many technologies, as outlined below:

- Jupyter Notebook

- Python
- String
- Tokenziation
- Counter
- Collections
- WorldCloud
- Spacy
- SentimentAnalyzer

The model creation process encompasses many technological stacks that consider the factors involved in defining the method for information development. The starting of the model development has been defined in different steps that includes the data processing along with the data analysis for the reporting of the information that is meaningful for the illustration of the work. It can be noted that the highlighting of the libraries for the inclusion of the information defining scheme has been reported for the system. The inclusion of basic libraries over the python has been considered through the use of Jupyter Notebook as the IDE. Based on the use of the string and tokenization libraries, the process of the data can be defined for the work. It uses the collection and the counter libraries mainly for the sake of generating the information processed for the tokens that has been incorporated in the work to report the adoption of the details for the system. It has been noted that the user can get the processing scheme-based model in terms of showing the processing through matplotlib and SNS to get the desired resulting of graphs.

4.1.2. Data Extraction Process

For this research, I have chosen Twitter to get the data. To get the data, the access to Twitter API was required. The Twitter API is a service that gives direct access to the Twitter's data which allows researchers and coders to get tweets from different user profiles, and also related meta-data if required. If someone wants to get access to the Twitter's API, it can consider as 'consumer_key', 'consumer_secret', 'access_token', 'access_token_secret'. Different end points in Twitter are available to the researchers and programmers to avail different types of data such as consumer tweets, tweets trends, and various search results.

There are some third-party software's and platforms available who offer access to the Twitter data through using API or license agreements. These third-party providers sometimes offer great capabilities and access to the past data which researchers can't get by using Twitter API directly. For example: Nvivo, Datashift, Etc.

For the research, I have produced my 'consumer_key', 'consumer_secret', 'access_token', 'access_token_secret' using Twitter
To get specific data from Twitter we can choose three different options: They are: 'Search Queries', 'Keywords and Hashtags', and 'User Profiles'.
As a researcher, if someone want to specify particular thing, they can by using keywords, phrases or even Booleans to retrieve their research relevant tweets. Also, they can specify some keywords or hashtags which are relevant to their research topics. For example #hate, #depression, #suicide etc. Also, users can mention Twitter profile names to download tweets from a specific Twitter profile. For data collection, I was required to install different packages such as, tweepy, nltk, and textblob. For installing them, I used two types of codes. I wrote codes to import my required packages.

4.1.3. Setup for Twitter API

The use of Twitter consider the change in credentials everytime. I will have to regenerate the API credentials next time I try to mine Twitter data. The dataset is then set up for preprocessing. First, authentication with Twitter is done. Then, data cleaning operations are conducted, including the removal of URLs, mentions, hashtags, and punctuations. Additionally, all text is converted to lowercase to ensure consistency. Finally, tweets are fetched based on specified location and keywords. These preprocessing steps ensure that the dataset is refined and standardized for accurate analysis and generating insights. By removing irrelevant information, the dataset becomes more manageable and meaningful for sentiment analysis.

The main() function sets parameters for fetching tweets related to the query- depression, within a specified location. It calls fetch_tweets_by_location_and_keywords() to obtain tweets, then through each tweet, cleaning them using clean_tweet() function. Cleaned tweets are taken for inspection. The if name == "__main__": block ensures main() is performed when the script is immediately performed. This script demonstrates a simple workflow to fetch and clean tweets based on location and keywords,

which could be further extended for sentiment analysis or trend monitoring tasks. Logging provides visibility into the cleaning process, helping in troubleshooting and analysis. After that, I came up with two datafiles. One is the main data file, another one is the cleaned data file. We will further do analysis with the cleaned data as shown in (Figure 17.)

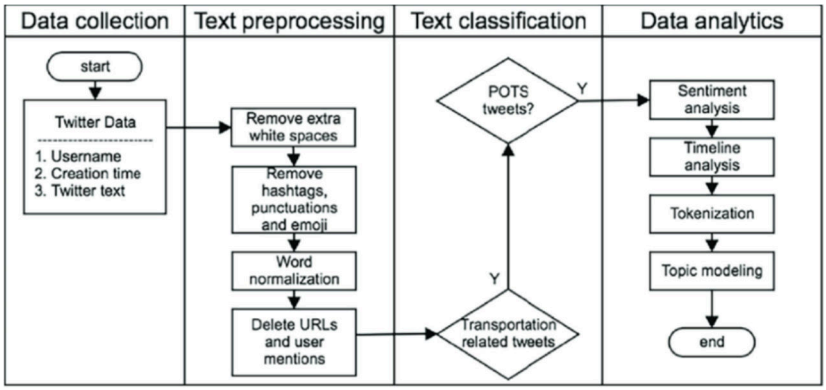


FIGURE 17. Setup for the Twitter API Data Collection

4.1.4. Data Preprocessing

The processing of the data has been specified in this step for the sake of reporting different analysis results through which the user can get the factors to report the findings related to the information highlighting of the work. The processed data presented as shown in (Figure 18.):

	tweet
0	"just as important in weighing the prolonged i...
1	wordle is cool but i much prefer to play this ...
2	i know this sort of data model is hard to cove...
3	(also, as far as long term consequences; not t...
4	is there a concept such as social impotence? I...
...	...
16867	look at the mental illness of the covid cult. ...
16868	yeah. ain't gonna happen. nyc is fucked. this ...
16869	let me tell you what ~ chronic insomnia, heada...
16870	the leftist riots and jan 6 were both caused b...
16871	2/ mental illness is on the table but is under...

16872 rows × 1 columns

FIGURE 18. Clean Dataset Sample

The procedure has been outlined in terms of reporting the characteristics that allow the user to obtain information on the data processing. The model development has been built on presenting information that may provide data to analyse the elements and report them using the model development scheme as shown in (Figure 19.)

	tweet	tokens
0	"just as important in weighing the prolonged i...	[``, just, as, important, in, weighing, the, p...
1	wordle is cool but i much prefer to play this ...	[wordle, is, cool, but, i, much, prefer, to, p...
2	i know this sort of data model is hard to cove...	[i, know, this, sort, of, data, model, is, har...
3	(also, as far as long term consequences; not t...	[(, also, ,, as, far, as, long, term, conseque...
4	is there a concept such as social impotence? I...	[is, there, a, concept, such, as, social, impo...
...	...	...
16867	look at the mental illness of the covid cult. ...	[look, at, the, mental, illness, of, the, covi...
16868	yeah. ain't gonna happen. nyc is fucked. this ...	[yeah, ., ai, n't, gon, na, happen, ., nyc, is...
16869	let me tell you what ~ chronic insomnia, heada...	[let, me, tell, you, what, ~, chronic, insomni...
16870	the leftist riots and jan 6 were both caused b...	[the, leftist, riots, and, jan, 6, were, both,...
16871	2/ mental illness is on the table but is under...	[2/, mental, illness, is, on, the, table, but,...

16872 rows × 2 columns

FIGURE 19. Tokenized Data

The data presented in the figure above indicates that tokenization has been implemented to report the elements that allow users to access specifics pertaining to the processed data. The generated tokens have been considered in terms of showing the tokens through which the user can get the resultants of most used words from the depicted results.

4.1.5. Development Process

The technique has been developed to demonstrate the qualities that allow the user to understand and analyse the sentiment of the work. This must be ensured that the defined factors with the stemming and tokenize words has been reported as shown in (Figure 20.):

	tweet	tokens	stemmed_tokens
0	"just as important in weighing the prolonged i...	[``, important, weighing, prolonged, impact, p...	[``, import, weigh, prolong, impact, pandem, m...
1	wordle is cool but i much prefer to play this ...	[wordle, cool, much, prefer, play, daily, mult...	[wordl, cool, much, prefer, play, daili, multi...
2	i know this sort of data model is hard to cove...	[know, sort, data, model, hard, cover, every, ...	[know, sort, data, model, hard, cover, everi, ...
3	(also, as far as long term consequences; not t...	[also, far, long, term, consequences, talking,...	[also, far, long, term, consequ, talk, ``, lon...
4	is there a concept such as social impotence? I...	[concept, social, impotence, like, really, wan...	[concept, social, impot, like, realli, want, s...
...	...	...	...
16867	look at the mental illness of the covid cult. ...	[look, mental, illness, covid, cult, says, 's,...	[look, mental, ill, covid, cult, say, 's, afra...
16868	yeah. ain't gonna happen. nyc is fucked. this ...	[yeah, ai, n't, gon, na, happen, nyc, fucked, ...	[yeah, ai, n't, gon, na, happen, nyc, fuck, n'...
16869	let me tell you what ~ chronic insomnia, heada...	[let, tell, chronic, insomnia, headaches, cons...	[let, tell, chronic, insomnia, headach, consta...
16870	the leftist riots and jan 6 were both caused b...	[leftist, riots, jan, 6, caused, mental, illne...	[leftist, riot, jan, 6, caus, mental, ill, cov...
16871	2/ mental illness is on the table but is under...	[2/, mental, illness, table, discussions, also...	[2/, mental, ill, tabl, discuss, also, one, ca...

16872 rows × 3 columns

FIGURE 20. Data processing with Stems

Based on the extracted results for the stems and the development scheme, it can be noted that the generated values have been reported for the emotion images detection as shown in (Figure 21.):

	tweet	tokens	stemmed_tokens
0	"just as important in weighing the prolonged i...	[, important, weighing, prolonged, impact, pan...	[``, import, weigh, prolong, impact, pandem, m...
1	wordle is cool but i much prefer to play this ...	[wordle, cool, much, prefer, play, daily, mult...	[wordl, cool, much, prefer, play, daili, multi...
2	i know this sort of data model is hard to cove...	[know, sort, data, model, hard, cover, every, ...	[know, sort, data, model, hard, cover, everi, ...
3	(also, as far as long term consequences; not t...	[also, far, long, term, consequences, talking,...	[also, far, long, term, consequ, talk, ``, lon...
4	is there a concept such as social impotence? I...	[concept, social, impotence, like, really, wan...	[concept, social, impot, like, realli, want, s...
...	...	...	...
16867	look at the mental illness of the covid cult. ...	[look, mental, illness, covid, cult, says, s, ...	[look, mental, ill, covid, cult, say, 's, afra...
16868	yeah. ain't gonna happen. nyc is fucked. this ...	[yeah, ai, nt, gon, na, happen, nyc, fucked, n...	[yeah, ai, n't, gon, na, happen, nyc, fuck, n'...
16869	let me tell you what ~ chronic insomnia, heada...	[let, tell, chronic, insomnia, headaches, cons...	[let, tell, chronic, insomnia, headach, consta...
16870	the leftist riots and jan 6 were both caused b...	[leftist, riots, jan, caused, mental, illness,...	[leftist, riot, jan, 6, caus, mental, ill, cov...
16871	2/ mental illness is on the table but is under...	[mental, illness, table, discussions, also, on...	[2/, mental, ill, tabl, discuss, also, one, ca...

16872 rows × 3 columns

FIGURE 21. Emotions Extraction from Data

The installation and importation of packages laid the groundwork for the subsequent analysis. Through the subprocess module, we executed pip commands to ensure the installation of essential libraries such as tweepy, nltk, and textblob. This automated installation process ensures that all required dependencies are present, minimizing manual intervention and streamlining the setup process.

Following the installation, we imported the installed packages into our Python environment. This step is crucial as it allows us to leverage the functionalities provided by these packages in our analysis. For instance, tweepy provides access to Twitter's API, enabling us to fetch tweets for analysis, while nltk offers a suite of natural language processing tools for text manipulation and analysis as shown in (Figure 22.)

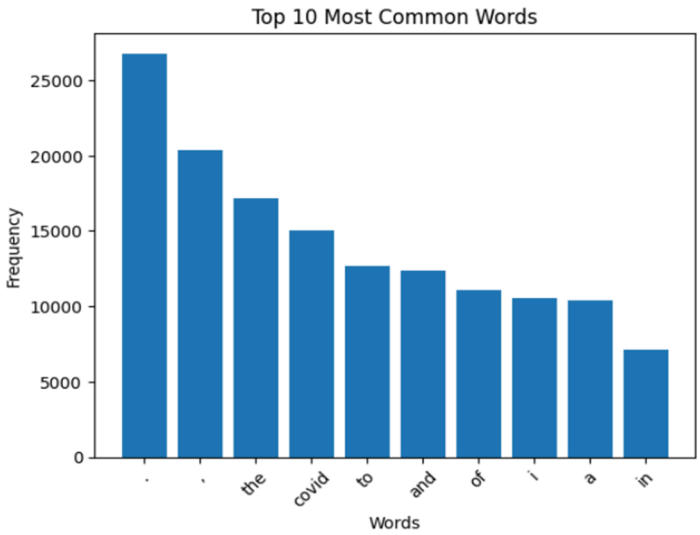


FIGURE 22. Most Common Words Identified from Data

The reported distribution of the sentiment scores has been shown as shown in (Figure 23.):

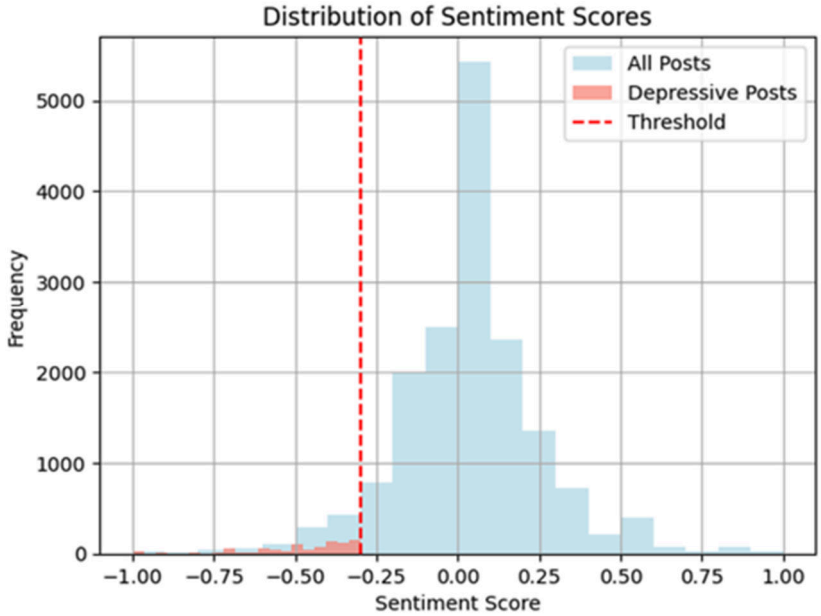


FIGURE 23. Distribution of sentiment score

The processed data that has been reported from the word cloud generation has been shown for the defining of the data as shown in (Figure 24.):

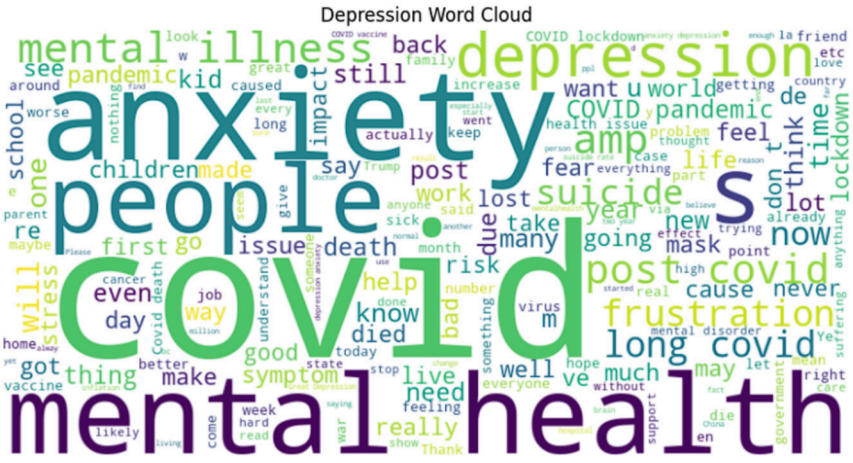


FIGURE 24. Wordcloud generated from the datafile

A graph has been generated for showing "suicide", "death", "dying", "sadness", "cry", "want to die" and "covid-19". Frequency of stressed 'df' words ' one. The code first creates a dictionary containing the frequency of each offending word by checking whether each word appears in the "tweet" field of the DataFrame as shown in (Figure 25.)

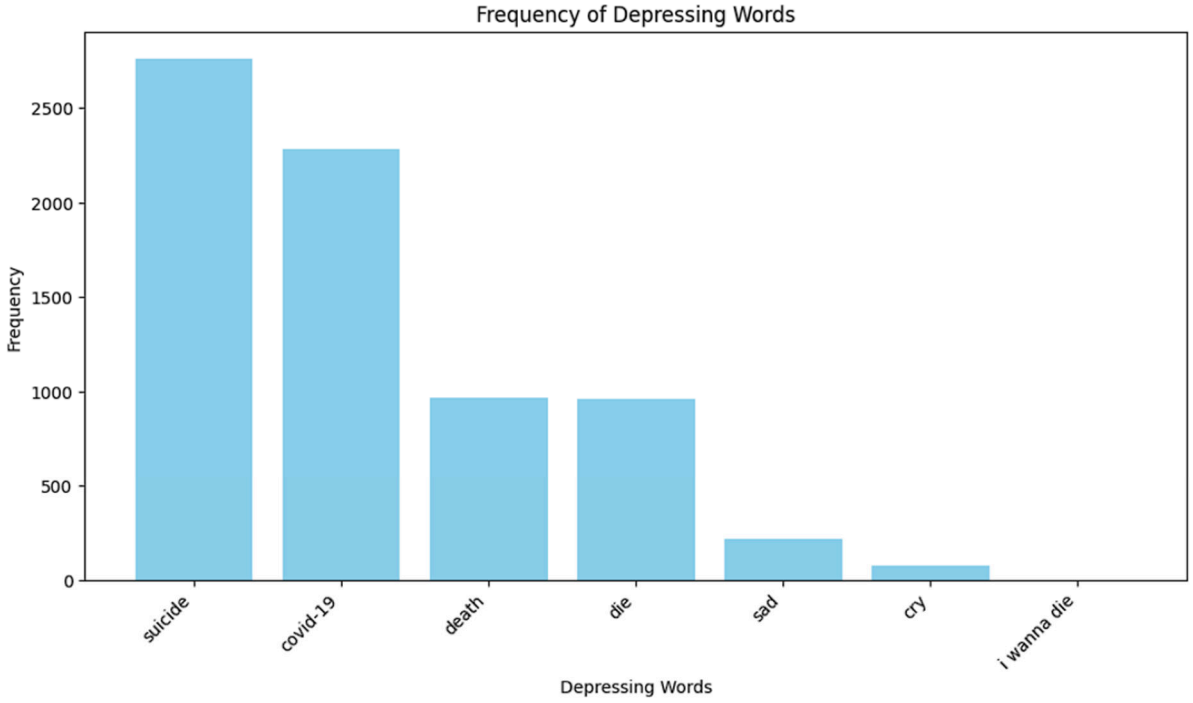


FIGURE 25. Frequency of Depression Words

The reporting of the tweets length has been defined in the work for the identification of the tweets characters that has been extracted from the Twitter API as shown in (Figure 26.):

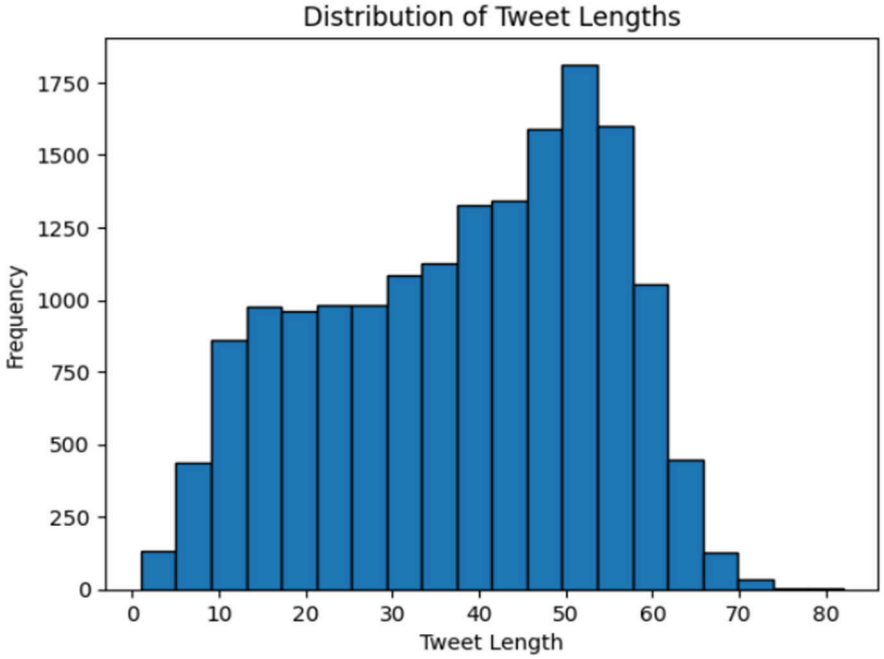


FIGURE 26. Distribution of the Tweets length

The commonly used stems word has been used in terms for the showing of the factors in terms of depicting the stems for the frequency as shown in (Figure 27.):

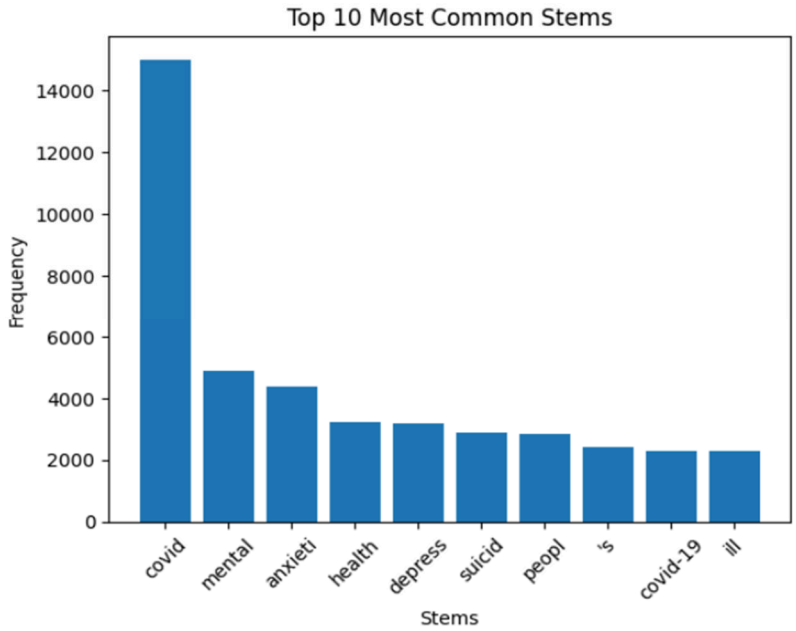


FIGURE 27. Stems reported for the frequency

The defined frequencies for the data have reported in the graph that the most frequently used word is the Covid that has caused the most depression in the data as shown in (Figure 28.)

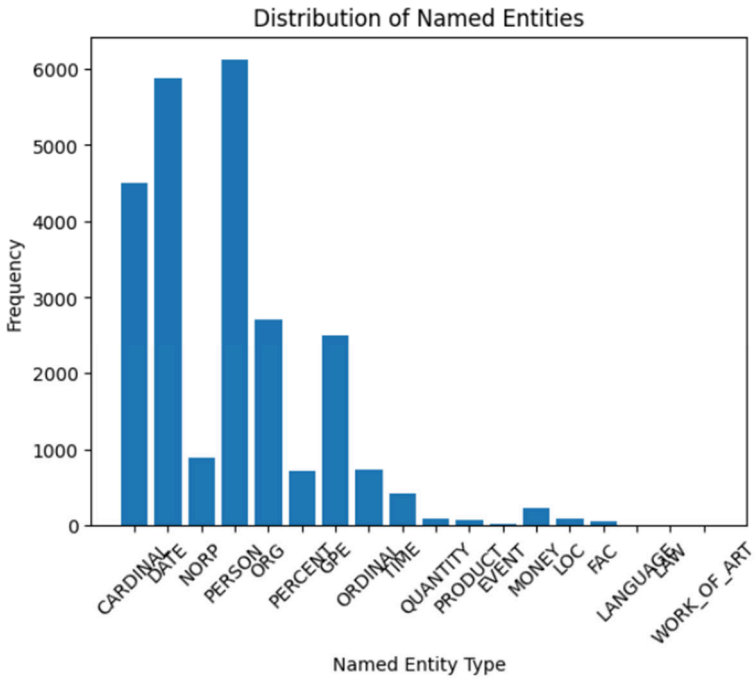


FIGURE 28. Named Entity Type from Data

The resultant has depicted that the frequency for the named entity has resulted in the cardinal in terms of showing the most frequent words in the data as shown in (Figure 29.)

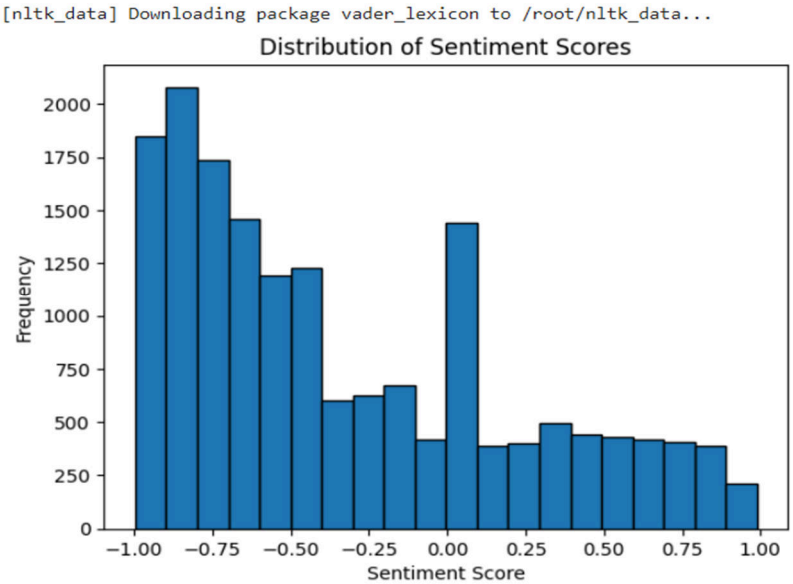


FIGURE 29. Sentiment Score of Distribution

4.1.6. Logistic Regression Model Implementation

The logistic regression models have been developed to identify the elements that contribute to the highlighted aspects. These models yield components that accurately predict the results. The use of prediction scheme has been used for showing the process to generate sentiment analysis. To this reason, the generated values has been considered to depict the libraries as shown in (Figure 30.):

```
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.metrics import accuracy_score, precision_score, recall_score
from transformers import BertTokenizer, BertForSequenceClassification
import torch
from textblob import TextBlob

train_texts, test_texts, train_labels, test_labels = train_test_split(df['tweet'], df['sentiment'], test_size=0.2, random_state=42)
```

```
Some weights of BertForSequenceClassification were not initialized from the model checkpoint at bert-base-uncased and are newly initialized: ['classifier.bias', 'classifier.weight']
You should probably TRAIN this model on a down-stream task to be able to use it for predictions and inference.
```

FIGURE 30. Libraries inclusion for the work

These libraries have been adopted in the work for the purpose of showing the features through which the reporting of the data results has been defined for the model as shown in (Figure 31.)

```
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.feature_extraction.text import CountVectorizer
from sklearn.linear_model import LogisticRegression
from sklearn.metrics import accuracy_score, precision_score, recall_score

df = df[['tweet', 'sentiment']]

X = df['tweet']
y = df['sentiment']

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

threshold = 0

y_train_discrete = (y_train > threshold).astype(int)
y_test_discrete = (y_test > threshold).astype(int)

vectorizer = CountVectorizer()
X_train_vec = vectorizer.fit_transform(X_train)
X_test_vec = vectorizer.transform(X_test)

classifier = LogisticRegression(max_iter=1000)
classifier.fit(X_train_vec, y_train_discrete)
y_pred = classifier.predict(X_test_vec)

accuracy = accuracy_score(y_test_discrete, y_pred)
precision = precision_score(y_test_discrete, y_pred, average='weighted')
recall = recall_score(y_test_discrete, y_pred, average='weighted')
```

FIGURE 31. Model Implementation for Logistic Regression

The task has been implemented by demonstrating the logistic regression classification scheme. This scheme allows the user to obtain specific details about defining information and illustrating the data process. For this purpose, the dataset has undergone data processing, including both training and testing, to extract information from the model. The data has been evaluated using metrics such as accuracy, precision, recall, and the f1-score.

4.1.7. Lexicon-based Sentiment Analysis

The second model used in this study incorporates lexicon-based sentiment analysis. To achieve this objective, the execution of the task has been specified to illustrate the aspects that may be taken into account for the user-defined working model as shown in (Figure 32.):

Lexicon Model Implementation

```
[ ]  threshold = 0
     test_labels_binary = (test_labels > threshold).astype(int)
     def lexicon_sentiment_analysis(text):
         polarity = TextBlob(text).sentiment.polarity
         if polarity > 0:
             return 1
         elif polarity < 0:
             return 0
         else:
             return 0.5

     lexicon_preds = [lexicon_sentiment_analysis(text) for text in test_texts]
     threshold = 0.5
     binary_lexicon_preds = [1 if pred > threshold else 0 for pred in lexicon_preds]

     lexicon_accuracy = accuracy_score(test_labels_binary, binary_lexicon_preds)
     lexicon_precision = precision_score(test_labels_binary, binary_lexicon_preds)
     lexicon_recall = recall_score(test_labels_binary, binary_lexicon_preds)
```

FIGURE 32. Lexicon-Based Model Development

The construction of the lexicon-based model has been described in terms of the processing scheme used for sentiment analysis. This scheme has been built to determine the components that need to be reported for the system as shown in (Figure 33.):

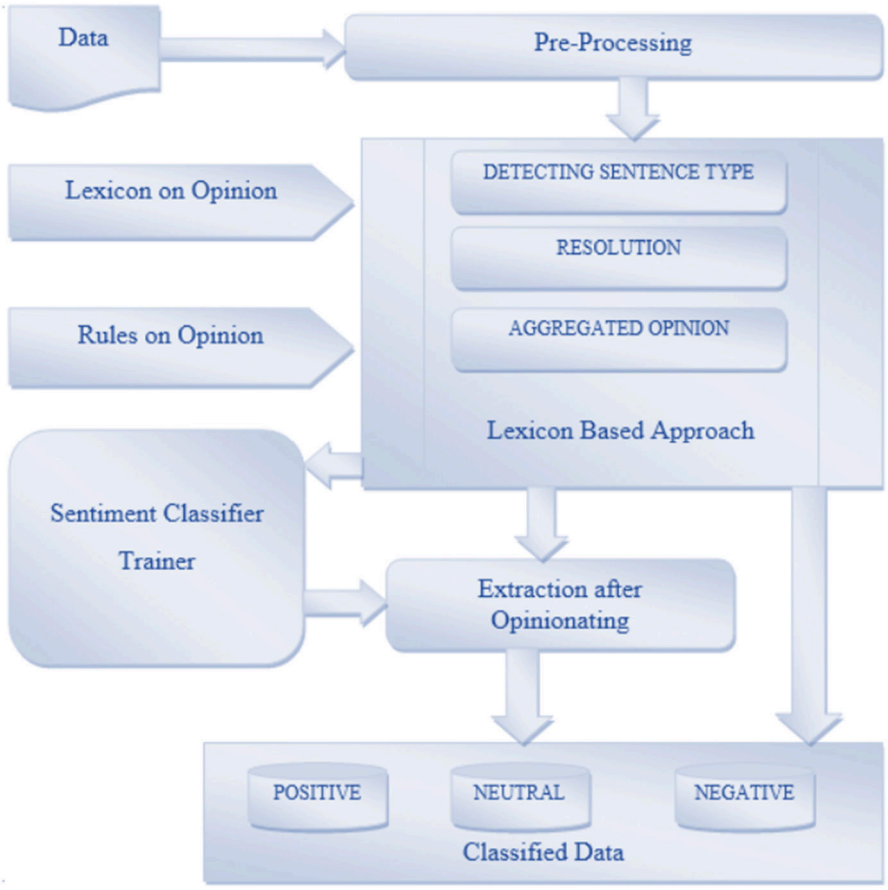


FIGURE 33. Lexicon-based Sentiment Analysis Model

The graphic provided illustrates the defined implementation approach for displaying the Lexicon-based sentiment analysis model.

4.2 Experimental Results

This part presents the experimental procedure by showcasing the specific information about the connection between the data and the work. The assessment of the models is determined by examining accuracy, precision, recall, and f1-score.

4.2.1. Logistic Regression Model

The logistic regression model attained an accuracy rate of 86.43%, a precision rate of 85.78%, and a recall rate of 86.43%. This indicates that the model performed well overall in correctly classifying sentiment. With high precision, it accurately identified positive and negative sentiments, minimizing false positives. The recall score suggests that the model effectively captured the majority of positive and negative sentiment instances in the dataset.

4.2.2. Lexicon-based Model Results

The Lexicon-based model yielded an accuracy of 63.38%, precision of 34.97%, and recall of 71.63%. While the accuracy is lower compared to the ML Model, the recall score indicates that it captured a higher proportion of positive and negative sentiment instances, albeit with lower precision.

The results have been reported in the (Table 2.) as follow:

TABLE 2. Generated Results

Model	**Accuracy**	**Precision**	**Recall**
Logistic Regression	87%	85%	86%
Lexicon-based Model	63%	35%	71%

These results have been shown in the graph as shown in (Figure 34.):

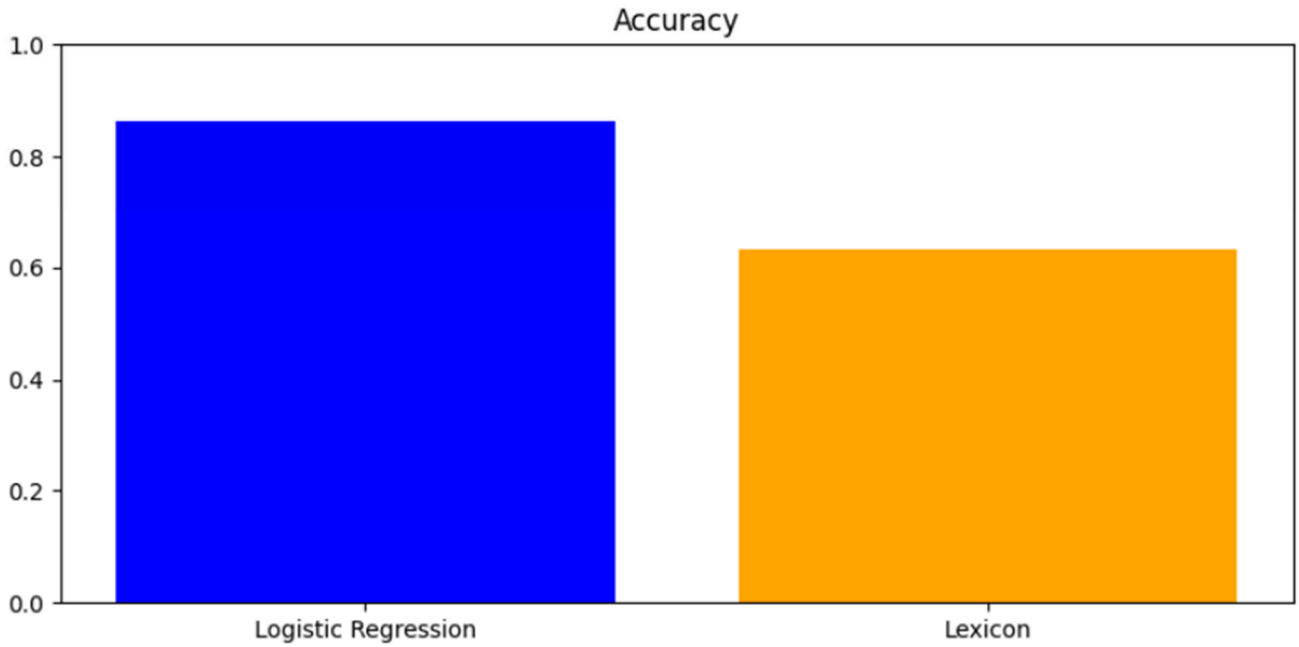


FIGURE 34. Accuracy Comparison of the Models

The precision findings for the created model have been displayed as shown in (Figure 35.):

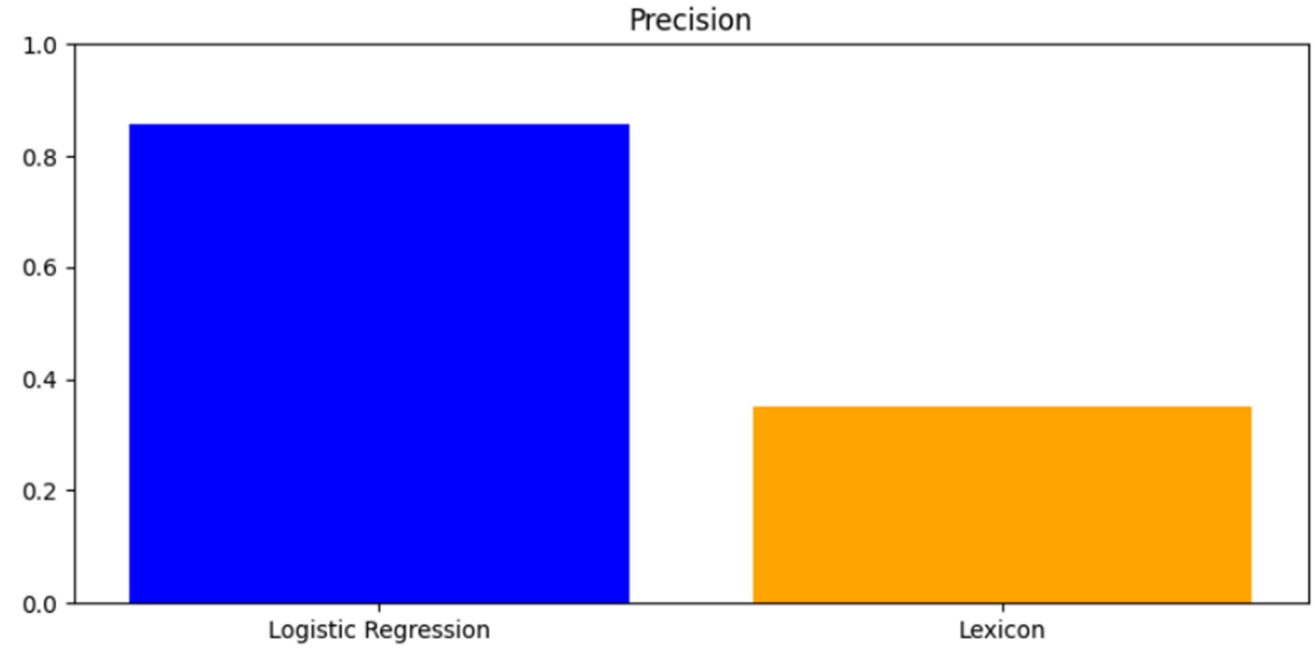


FIGURE 35. Precision based accuracy computation

The recall values have been presented in terms of showing the correctness of the model as shown in (Figure 36.):

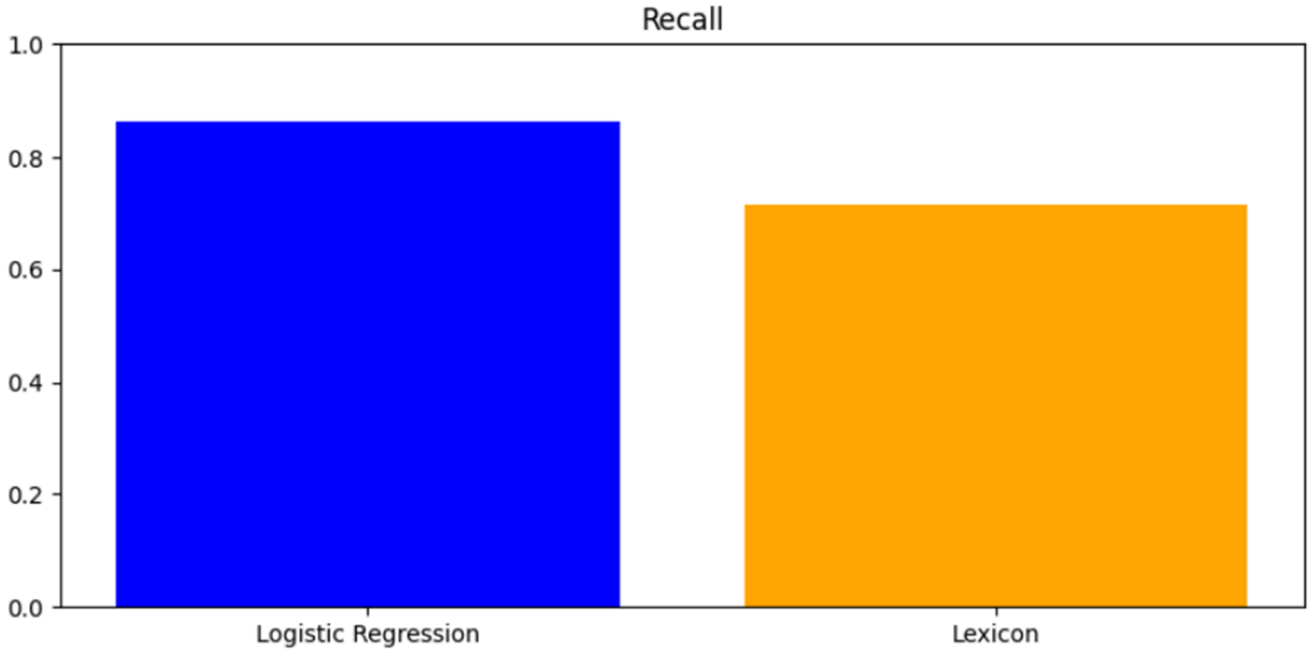


FIGURE 36. Recall based accuracy computation

5 CONCLUSION

Ultimately, this study highlights the significance of studying depressed tweets on social platforms such as Twitter to get a more profound comprehension of the characteristics and types of such messages, as well as the attitude of users during the COVID-19 epidemic through the examination of social post data. The findings of the sentiment analysis offer valuable insights into the consumers' attitudes towards the social media posts, revealing the depressive posts through texts and words used. This study also explains the characteristics and types of depression and non-depression scores derived from the analysis. This sentiment analysis is helpful to address depression issues of the online users and improve the mental health of online communities, through more research, interventions, and support that will provide further research facilities. By using sentiment analysis to detect the depressive posts, the authorities and mental health professionals will be able develop preventive interventions to help the users who are experiencing depressive state of mind. Moreover, by analysing these, positivity can also be promoted for the depressed individuals which will help them in getting mental and emotional support.

It is recommended to apply new strategies to detect and prevent or intervene from the analysis of this depressive posts to prevent any unwanted situations from the depressive posts shared by the online users. An automated system can be developed where upon detecting any kind of depressive posts, a helpline number or support system will be recommended to the users immediately and provide necessary support. Moreover, through implementing positive measure and promotional activity to reduce depression can also play an important role in voicing support for the depressed individuals from analysing their depressive posts online. Including different social media platforms to better understand the depressive message in their posts and explore differences in expression between platforms. Subsequent investigations into these hypotheses might enhance comprehension of depression as expressed on social media and propose methods for cultivating mental well-being in digital contexts. For the purpose of ensuring that all the requirements have been defined in the accurate manner for the getting of objectives are listed as follow:

Objective 1: This objective mainly focusses over the reporting of the features through which the problem statement can be identified from the past studies. These limitations have been reported in a way that the designing of the work has been illustrated for the showing of research gaps to be incorporated in the work.

Objective 2: The development of the model development process has been illustrated as the baseline for which this objective has been designed. It has been noted that the generated results have been illustrated for the model implementation using Lexicon and ML based logistic regression.

Objective 3: The evaluation of the model has been identified as the primary aim for this effort. To ensure accurate model estimate, it is important to include accuracy monitoring and precision/recall reporting in the information processing.

The development of the model has been reported as the baseline through which the generating of the model has been depicted to highlight the performance in a way that logistic regression has gained more accuracy then the lexicon based approach.

5.1 Research Outcome

The results of the sentiment analysis models illustrate the variation in performance levels across different methods. It has been noted that the logistic regression model has presented a much higher accuracy along with the precision and the recall values that indicates its effectiveness in correct way of the classifying for the sentiment. The logistic regression approach is very dependable for accurately determining the relevance of positive and negative sentiment expressions while minimising misclassifications. However, the Lexicon-based model has reported a lower accuracy and precision but a higher recall score that indicates a sufficient performance of the model which implies that it was reported for better result generation for the capturing of the larger proportion of positive and negative sentiment instances that includes the reporting of more false positives. Hence, the summary of the values has been reported in a way that the logistic regression model has reported in high precision and recall values in comparison to the other model that depicts a better performance of the resultant.

5.2 Research Recommendations and Limitations

One limitation of this research is the incorporation with the predefined sentiment lexicons that is mainly reported for the capturing of the context-specific expressions of sentiment in social media text.

Furthermore, the logistic regression-based model has demonstrated superior performance, which is contingent upon the quality and representativeness of the training data. This dependence may result in biases or mistakes in sentiment categorization. In order to overcome these constraints, future research should focus on researching more advanced natural language processing (NLP) approaches. This will enable the creation of deep learning models that can effectively capture complex linguistic patterns and context dependencies.

REFERENCES

Alslaity, A. and Orji, R. 2022. Machine Learning Techniques for Emotion Detection and Sentiment analysis: Current state, challenges, and Future Directions. *Behaviour & Information Technology*, 1(1), pp.1–26. Available at: doi:https://doi.org/10.1080/0144929x.2022.2156387. Referenced 03 May 2024

Asif, M., Ishtiaq, A., Ahmad, H., Aljuaid, H. and Shah, J. 2020. Sentiment Analysis of Extremism in Social Media from Textual Information. *Telematics and Informatics*, pp. 48, p.101345. Available at: doi:https://doi.org/10.1016/j.tele.2020.101345. Referenced 01 May 2024

Bharti, S.K., Varadhaganapathy, S., Gupta, R.K., Shukla, P.K., Bouye, M., Hingaa, S.K. and Mahmoud, A. 2022. Text-Based Emotion Recognition Using Deep Learning Approach. *Computational Intelligence and Neuroscience*, [online] 2022(1), p.2645381. Available at: doi:https://doi.org/10.1155/2022/2645381. Referenced 27 April 2024

Birjali, M., Kasri, M., & Beni-Hssane, A. 2021. A comprehensive survey on sentiment analysis: Approaches, challenges and trends. *Knowledge-Based Systems*, *226*, 107134.

Cacheda, F., Fernandez, D., Novoa, F.J. and Carneiro, V. 2019. Early Detection of Depression: Social Network Analysis and Random Forest Techniques. *Journal of Medical Internet Research*, 21(6), p.e12554. Available at: doi:https://doi.org/10.2196/12554. Referenced 22 April 2024

Kim, N.H., Kim, J.M., Park, D.M., Ji, S.R. and Kim, J.W. 2022. Analysis of Depression in Social Media Texts through the Patient Health Questionnaire-9 and Natural Language Processing. *DIGITAL HEALTH*, 8(1), p.205520762211142. Available at: doi:https://doi.org/10.1177/20552076221114204. Referenced 03 April 2024

Pan, J., Liu, B. and Kreps, G.L. 2018. A Content Analysis of depression-related Discourses on Sina Weibo: attribution, efficacy, and Information Sources. *BMC Public Health*, 18(1). Available at: doi:https://doi.org/10.1186/s12889-018-5701-5. Referenced 21 March 2024

Tillman, G., March, E., Lavender, A.P., Braund, T.A. and Mesagno, C. 2023. Disordered Social Media Use during COVID-19 Predicts Perceived Stress and Depression through Indirect Effects via Fear of COVID-19. *Behavioral Sciences*, [online] 13(9), p.698. Available at: doi:https://doi.org/10.3390/bs13090698. Referenced 10 March 2024

Yadav, A., & Vishwakarma, D. K. 2020. Sentiment analysis using deep learning architectures: a review. *Artificial Intelligence Review*, *53*(6), 4335-4385.

Wankhade, M., Rao, A. C. S., & Kulkarni, C. 2022. A survey on sentiment analysis methods, applications, and challenges. *Artificial Intelligence Review*, *55*(7), 5731-5780.

Data Processing Module

```
def remove_emoji(tokens):
    emoji_pattern = re.compile("["
                           u"\U0001F600-\U0001F64F"  # emoticons
                           u"\U0001F300-\U0001F5FF"  # symbols & pictographs
                           u"\U0001F680-\U0001F6FF"  # transport & map symbols
                           u"\U0001F1E0-\U0001F1FF"  # flags (iOS)
                           u"\U00002702-\U000027B0"
                           u"\U000024C2-\U0001F251"
                           "]+", flags=re.UNICODE)
    return [emoji_pattern.sub(r'', token) for token in tokens]

df['tokens'] = df['tokens'].apply(remove_emoji)
df
```

	tweet	tokens	stemmed_tokens
0	"just as important in weighing the prolonged i...	[, important, weighing, prolonged, impact, pan...	[``, import, weigh, prolong, impact, pandem, m...
1	wordle is cool but i much prefer to play this ...	[wordle, cool, much, prefer, play, daily, mult...	[wordl, cool, much, prefer, play, daili, multi...
2	i know this sort of data model is hard to cove...	[know, sort, data, model, hard, cover, every, ...	[know, sort, data, model, hard, cover, everi, ...
3	(also, as far as long term consequences; not t...	[also, far, long, term, consequences, talking,...	[also, far, long, term, consequ, talk, ``, lon...
4	is there a concept such as social impotence? I...	[concept, social, impotence, like, really, wan...	[concept, social, impot, like, realli, want, s...

Analysis module

```
fig, axs = plt.subplots(3, 1, figsize=(8, 12))

axs[0].bar(models, accuracy, color=['blue', 'orange'])
axs[0].set_title('Accuracy')
axs[0].set_ylim(0, 1)

axs[1].bar(models, precision, color=['blue', 'orange'])
axs[1].set_title('Precision')
axs[1].set_ylim(0, 1)

axs[2].bar(models, recall, color=['blue', 'orange'])
axs[2].set_title('Recall')
axs[2].set_ylim(0, 1)

plt.tight_layout()
plt.show()
```

APPENDIX 3

Sentiment Model

```
def deep_learning_sentiment_analysis(df):
    X = df['tweetText'].values
    y = df['Sentiment_Class'].values

    label_encoder = LabelEncoder()
    y = label_encoder.fit_transform(y)
    X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

    tokenizer = Tokenizer(num_words=5000)
    tokenizer.fit_on_texts(X_train)

    X_train = tokenizer.texts_to_sequences(X_train)
    X_test = tokenizer.texts_to_sequences(X_test)

    maxlen = 100
    X_train = pad_sequences(X_train, padding='post', maxlen=maxlen)
    X_test = pad_sequences(X_test, padding='post', maxlen=maxlen)
    model = Sequential()
    model.add(Embedding(5000, 64, input_length=maxlen))
    model.add(LSTM(64))
    model.add(Dense(3, activation='softmax'))
    model.summary()
    model.compile(optimizer='adam', loss='sparse_categorical_crossentropy', metrics=['accuracy'])

    model.fit(X_train, y_train, epochs=10, batch_size=32, validation_data=(X_test, y_test))
    loss, accuracy = model.evaluate(X_test, y_test)
    overall_sentiment = 'Positive' if accuracy > 0.5 else 'Negative'
```

ABSA Model

Aspect-Based Sentiment Analysis (ABSA)

```
[ ]  def aspect_based_sentiment_analysis(df):
         def aspect_based_sentiment_analysis(text):
             polarity = TextBlob(text).sentiment.polarity
             return 'Positive' if polarity > 0 else 'Negative' if polarity < 0 else 'Neutral'

         df['Aspect_Sentiment'] = df['tweetText'].apply(aspect_based_sentiment_analysis)
         overall_sentiment = df['Aspect_Sentiment'].value_counts().idxmax()
         accuracy = (df['Aspect_Sentiment'] == df['Sentiment_Class']).mean()
         results = pd.DataFrame({'filename': ['filename'], 'overall_sentiment': [overall_sentiment], 'accuracy': [accuracy]})
         return results

     def aspect_based_sentiment_analysis(df):
         def aspect_based_sentiment_analysis(text):
             polarity = TextBlob(text).sentiment.polarity
             return 'Positive' if polarity > 0 else 'Negative' if polarity < 0 else 'Neutral'

         df['Aspect_Sentiment'] = df['tweetText'].apply(aspect_based_sentiment_analysis)
         overall_sentiment = df['Aspect_Sentiment'].value_counts().idxmax()
         accuracy = (df['Aspect_Sentiment'] == df['Sentiment_Class']).mean()
         results = pd.DataFrame({'filename': ['filename'], 'overall_sentiment': [overall_sentiment], 'accuracy': [accuracy]})
         return results

     results_aspect_based1 = aspect_based_sentiment_analysis(Alko_Oy)
     results_aspect_based2 = aspect_based_sentiment_analysis(FrennHelsinki)
     results_aspect_based3 = aspect_based_sentiment_analysis(Kesko)
     results_aspect_based4 = aspect_based_sentiment_analysis(Lidl_Suomi)
```

APPENDIX 5

Fine-grained Model

Fine-grained Sentiment Analysis

```
def fine_grained_sentiment_analysis(df):
    def fine_grained_sentiment_analysis(text):
        sid = SentimentIntensityAnalyzer()
        scores = sid.polarity_scores(text)
        compound_score = scores['compound']
        if compound_score >= 0.05:
            return 'Positive'
        elif compound_score <= -0.05:
            return 'Negative'
        else:
            return 'Neutral'

    df['Fine_Grained_Sentiment'] = df['tweetText'].apply(fine_grained_sentiment_analysis)
    overall_sentiment = df['Fine_Grained_Sentiment'].value_counts().idxmax()
    accuracy = (df['Fine_Grained_Sentiment'] == df['Sentiment_Class']).mean()
    results = pd.DataFrame({'filename': ['filename'], 'overall_sentiment': [overall_sentiment], 'accuracy': [accuracy]})
    return results

results_fine_grained1 = fine_grained_sentiment_analysis(Alko_Oy)
results_fine_grained2 = fine_grained_sentiment_analysis(FrennHelsinki)
results_fine_grained3 = fine_grained_sentiment_analysis(Kesko)
results_fine_grained4 = fine_grained_sentiment_analysis(Lidl_Suomi)
```