Sheffield Hallam University | College of Business, Technology and Engineering

MSc Dissertation Report

DEPRESSION DETECTION USING LANGUAGE MODELS

A dissertation submitted in partial fulfilment of the requirements of Sheffield Hallam University for the degree of Master of Science in MSc Big Data Analytics

Student Name	DEVI PRIYA BIJOSH MOHAN
Student ID	31056531
Supervisor	DR BAYODE OGUNLEYE
Date of Submission	09-01-2023

ABSTRACT

Depression is the leading cause of disability worldwide and a major contributing factor towards suicide. The WHO estimated that 280 million individuals worldwide had depression in 2021, and that number is increasing day after day. Early identification is necessary for providing necessary precautions which could stop the illness from getting worse. Prior studies have demonstrated that language use is impacted by depression and many depressed individuals use social networking sites or the internet in general to gather information or discuss their struggles. People are increasingly prone to use social media platforms for online communications instead of face-to-face interactions. This paper investigated the use of machine learning models for the early detection of depression from posts on social media. In this supervised learning project, 2 labelled Twitter datasets are used to train the machine learning models, while a third manually annotated validation dataset is used for performance evaluation. This study compared traditional models versus language models to determine which model performs the best at detecting depression. The language models employed are BERT, BiLSTM, XLNet, and ELMO as compared to the traditional models SVM, DT, LR, and KNN. Text is converted into vectors using the TF-IDF and Word2Vec methods. The study found that the Language models outperformed the Traditional model at detecting depression, with XLNet being the model that performed the best with an accuracy of 90%. Statistical performance detecting parameters like Accuracy, Precision, Recall, and F1score are used to evaluate performance. The proposed EDDS framework with XLNet language model outperformed them all and proved to be quite effective at detecting depression in the very beginning stages. With the use of this framework, a medical expert will be able to identify those who require assistance and prevent the sickness from getting worse because of their ignorance.

ACKNOWLEGEMENT

I would like to thank my supervisor Bayode Ogunleye, who guided, supported, and encouraged throughout my research. This study could not have been completed without his guidance and support. Secondly, my special thanks go to my family who gave me an immense support and motivation. Finally, I would like to show my gratitude towards the participants who took part in the validation survey by providing there valuable and suggestions.

Table of Contents

CHAPTER 1

1.0 Introduction

Depression is one of the most prevalent mental illnesses which adversely affects people's lives (Chiong et al., 2021). It can have a significant impact on many facets of life, including academic achievement, work productivity, family relations, and the capacity to engage in community activities (WHO, 2021). Previous studies discovered that depression and physical wellbeing, particularly cardiovascular disease and tuberculosis, had substantial correlations (Stephen & Prabu, 2019). All age group and socioeconomic status is affected by depression (WHO, 2021). According to the WHO in 2021, 3.8% of the population, including 5% of adults suffer from depression (WHO, 2021). This means approximately 280 million people worldwide experience depression, which is a leading cause of disability and a major contributor to the overall burden of diseases (Chiong et al., 2021).

Depression can be stated as a chain of mental conditions characterised by persistently sad mood or lack of interest in life (Burdisso et al., 2019). The individual may eventually suffer and perform poorly at work, school, and within the family. As a result, the person experiences severe trouble functioning in their personal, family, social, academic, occupational, and/or other key domains during the period of depression (WHO, 2021). It has an impact on the patient as well as the patient's family. Early depression diagnosis might provide individuals who are affected better chances to get the right care and recover from the condition. However, due to a dearth of understanding about mental health and since there is no pain associated with mental health problems, many individuals with depression do not identify it (WHO, 2022). Everyone is susceptible to depression and the friends and family members might not be able to understand it earlier. Although some know a little about depression, they are often reluctant to seek professional help because of a sense of shame. Depression takes hold gradually, without a person realising that depressive thoughts and feelings are increasingly dominating their perspective - and their life (Angskun et al., 2022). It will result in a significant health deterioration and even the possibility of suicide if it is left untreated due to self-denial or unawareness (Shetty et al., 2020). As per WHO, every year, around 700,000 people die by suicide and for people aged 15 to 29, suicide is the fourth most common cause of death (WHO, 2021).

According to a scientific brief released by the WHO, COVID-19 pandemic triggers a 25% increase in prevalence of anxiety and depression worldwide (WHO, 2022). The countries have already been warned to consider the event as a wake-up call to take better precautions towards mental health and do a better job of supporting the mental health of their citizens. People have been forced to stay inside and engage in fewer social interactions due to COVID-19, which has made the depression problem worse (Stephen & Prabu, 2019). The pandemic has resulted in unprecedented levels of stress due to social isolation. This was related to restrictions on people's capacity to work, ask loved ones for help, and participate in their communities. Stress factors that can cause anxiety and depression include loneliness, fear of illness, suffering and death for oneself and loved ones, grieving following a loss, and financial concerns (WHO, 2022). During the pandemic, the prevalence of depression in the general population was reported to be 33% (Angskun et al., 2022). It is evident that the number of cases of depression in the US has tripled. While in Thailand, the cases increased massively and became the reason behind 60% of suicides (Angskun et al., 2022). It is crucial to diagnose the condition as soon as possible to take appropriate treatment before it reaches a suicidal stage. The below figure shows the statistics of the depression cases worldwide for the past years.

Several methods are proposed to help people who cannot receive adequate diagnosis and treatment. Depression can be diagnosed by medical methods such as medical history, physical exams, lab tests or psychological evaluation that answers questions about the thoughts, emotions, and behaviour. Depending on the severity of the symptoms, depression may be treated with, for example, behavioural activation, cognitive behavioural therapy, interpersonal psychotherapy, selective serotonin reuptake inhibitors, and tricyclic antidepressants (Cha et al., 2022). Though the methods mentioned before are effective for depression, early diagnosis, and the gap between who can and cannot receive the treatment is still relevant. Traditional methods rely on a doctor's subjective evaluation following a consultation with the patient and analysis of the relevant questionnaires constitutes nearly all the basis for the diagnosis of depression (Xu et al., 2019). This method is inadequate in the case of patients who are unaware about their mental stage. The issue is proven by the fact that only 4.6% of the world population suffer from depression, 43.3% of them do not take their symptoms seriously and do not care to be treated professionally (Trotzek et al., 2018). If the symptoms could be identified early, either by therapy session or medications, the affected people can come back to normal life (AlSagri & Ykhlef, 2020).

To aid the early identification of depression, researchers have focused on developing supplementary approaches. Many scholars used social media as one of the potential solutions in exploring and diagnosis of depression symptoms (Biradar & Totad, 2018). For instance, Kim et al. (2020) conducted a study and found out that social media datasets are useful in detecting social media users' emotional statements and potential mental illness.

1.2 Why social media?

Social networking platforms have grown significantly over the past 10 years and are now utilised extensively in practically every facet of modern life. Social networking sites such as Twitter, Facebook, Instagram, WhatsApp, etc., have significantly altered how individuals communicate with one another (Islam, Md Rafiqul et al., 2018). Social networking is a type of communication which does not require eye contact or facial expression. Thus, people became very comfortable expressing their ideas, sentiments, goals, and other accomplishments in social media such as posts, messages, and comments rather than speaking or sharing the emotions to others (Burdisso et al., 2019). The techniques used to detect depression using social media data have been found to be efficient and affordable when compared to conventional methods. People are accustomed to sharing their innermost thoughts and sentiments on social media. The enormous corpus offers a wealth of writing addressing feelings that may indicate depression, such as melancholy, fatigue, and collapse (Cha et al., 2022). Thus, social media can be a fruitful tool for gathering data about the state of the public's mental health.

1.3 Justification of Twitter Data Selection

From the numerous social networking sites, Twitter seems to be a more appropriate data source for this research. Twitter is a micro blogging site which allows users to post concise, 280-character length short messages to either the public or to a specific group of followers. According to statistics compiled by e-marketer, there were an average of 330 million active monthly users on Twitter in the third quarter of 2017, and that number is growing quickly. The below Figure 1.1 shows the statistics of Twitter's daily active users from Statista website (Statista, 2022).

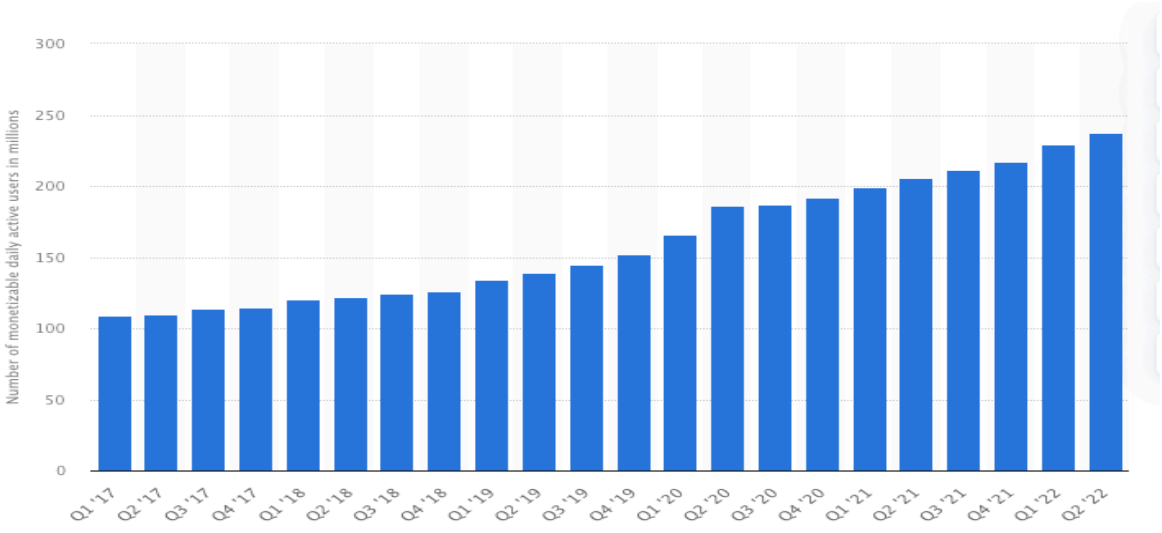


Figure 1.1 Twitter User. Statistics

People uses Twitter as a media to share their thoughts, feelings, and daily activities (Biradar & Totad, 2018). Twitter is commonly used as a source of personal data in academia, and the same reasons apply for this dissertation too. The primary reason is that there are over 500 million postings made each day on Twitter (Twitter, 2018), the majority of which are visible to the public (through Twitter API). Twitter serves as a medium for tracking, analysing, and comprehending global and local trends. Each tweet contains a condensed block of information that provides a brief insight about users' personal life and emotions because of the social network's emphasis on personal themes and the volume of public data it collects (Rajaraman et al., 2020). Twitter is ideal for academic research because posts are always public until the user actively changes the settings for their profile (Biradar & Totad, 2018) and thereby used for this study.

1.4. Why sentimental Analysis

Sentiment analysis became an efficient method to understand people's sentiments in multiple situations. It is a natural language processing (NLP) technique used to understand the sentiment associated with the user's post. Social media data, which includes text data, emoticons, emojis, and other visual elements, would be used for this purpose. By combining linguistics and computing, natural language processing enables the creation of intelligent systems that can comprehend, interpret, and extract meaning from spoken language and written text (Pachouly et al., 2021). Twitter sentiment analysis has become a very popular and efficient method to detect depression in the current era. People often express their feelings through twitter in the

form of tweets. Likewise, when a person is starting to experience depression, it will get reflected in the posts shared by him/her. Sentiment analysis is used to uncover the feelings concealed in these posts of the social media users. (Joshi & Kanoongo, 2022; Rajaraman et al., 2020). These posts can be examined into polarity (positive or negative) to decide the user's mental state to identify the symptoms of depression as well as suicidal thoughts (Joshi & Kanoongo, 2022). Thus, sentiment analysis with Machine Learning (ML) models can effectively be used to detect the depression in its early stage (Biradar & Totad, 2018).

1.5 Problem statement

The problem addressed in this study is the need for an early depression detection framework which can be used as an aid to the health sector. An efficient early depression detection machine learning framework can predict the depressive symptoms in its early stage and can be used to provide necessary support and steps to treat the illness. This can lead the people to get back to normal life. As per the past studies, researchers proved that sentimental analysis on depression detection became a cost effective and efficient method to detect depression.

Depression has become a serious problem in society, and it is only getting worse. 43.3% of people who are having depressive symptoms do not take their symptoms seriously and do not care to be treated professionally (Trotzek et al., 2018). The unawareness or ignorance of people can lead to acute depression which may end up with suicide. It is evident that depression has become one of the major reasons behind the suicides happening in worldwide. This became a life-threatening problem globally.

The ineffectiveness of the conventional professional consultation techniques to detect depression from the people who are unaware/ashamed leads to finding alternative methods to identify early depression. Social media sentiment analysis with machine learning models became one of the prevalent approaches to detect depression. However, there have been a few studies done to detect early-stage depression utilizing pre-trained Language models. This research allows to find the suitability of pre-trained language models by doing a comparative study to understand how effective the language models than the traditional ones to the detect depression. The proposed framework can be adapted by medical professional to easily identify the people with depression. This framework classifies the tweets into depressed or non-depressed by analysing the sentiments of the tweets using the state-of-art language models and

thereby overcome the research gap in the depression detection using the language models. Additionally, our study identifies the cause of the tweet misclassification, which was not covered by the earlier research.

1.6 Research Aim

The project's goal is to develop a framework with sentiment analysis concept to detect early-stage depression.

1.6.1 Research question

How can sentiment analysis concept help detect early-stage depression?

1.6.2 Research Objectives

1. To conduct a literature review to get a good background of the study and to understand the methodologies appropriate for this problem.
2. To conduct a comparative study to detect depression using machine learning.
3. To ascertain the best performing model suitable to detect depression.

1.6.3 Deliverable

A framework for the health sector to use as an aid to detect depression in its early stages by analysing the tweets of a user.

1.7 Project Benefits

The project benefits of the dissertation are as follows.

- The intention behind this dissertation is to benefit the society with a ML model framework which can detect the early-stage depression symptoms using social media data.
- This study provides a strong background on depression detection using pre-trained language model which will help to bring more future studies in this area.
- It can help the health sector to identify early-stage depression for taking necessary steps.

- It is a very time efficient and cost-effective method since it's only using people's social media communication to predict their mental state.

1.8 Project structure

Chapter 1: the motivation and the background for this study is presented. The problem statement, aim and objectives were stated and then the research question to be answered was specified. In brief, this chapter provides details on the problem, its complications, and the suggested solution. This chapter also provides details about the succeeding chapters.

Chapter 2: serves as the foundation on which this study is built. It examines, compares, and analyses the limitations, gaps, and shortcomings of the previous literatures that attempted to detect depression. To determine the areas for improvement and cutting-edge solutions, it examines the range of data sources, various NLP methodologies, and machine learning models and their performances.

Chapter 3: discusses the chosen research design as well as the justification for the methodological preference, reasoning approach, and research philosophy that are used to solve the problem and accomplish the objectives. The details of datasets, text pre-processing, exploratory data analysis, machine learning models, and the framework are presented in the subsections of this chapter.

Chapter 4: presents the comparative study results of this project. The performance evaluation of all the models conducted using approach evaluation matrix such as accuracy, precision, recall, and F1 score are captured and compared. Thus, this chapter answers the research question and outlines the main conclusions.

Chapter 5: demonstrates how the EDDS framework was evaluated. The chapter presents how this study findings might be useful, relevant, and reliable. The comments provided will helps to evaluate the reliability and relevance of the research findings.

Chapter 6: summarises the study, discusses the challenges and limitations of the research. Thus provide the recommendation for the future study.

Chapter 2: Literature Review

2.0 Introduction

Depression is a mental disorder that affects a person's mood. People who are experiencing a depressive episode feel unhappy, irritated, empty, and lose interest in their current activities (WHO,2021). It will begin by prolonging a depressing feeling for the entire day. These depressed moods impede concentrate, cause excessive guilt feelings, low self-worth, lack of future hope, suicidal thoughts, disrupt sleep, modify body weight, and make you feel extremely tired or low energy (WHO, 2021). Additionally, some people face physical symptoms like pain, exhaustion, and weakness. Social networking sites have become the ideal companion in people's lives; thus, most individuals communicate these emotional changes there via posts, comments, audios, or videos. This inspires a lot of research, which is evidently very effective, to identify the emotional condition of people's comments or postings in social media to evaluate the symptoms of depression (Angskun et al., 2022). Generally speaking, different words may convey different emotions. For example, the terms wonderful and tasty convey happiness, gloomy and weep convey sadness, yell and boiling convey rage, and so on (Shetty et al., 2020). This can be efficiently used to predict depression with sentimental analysis techniques. This study's goal is to provide a framework for early depression detection using sentiment analysis, and as a result, this chapter provides a review of the literature on the various machine learning models that will be employed in the following sections.

2.1 Machine Learning Models

Chiong et al. (2021) did a comparison study using several single and ensemble models such as Logistic Regression(LR), Decision Tree, Random Forest, and Adaptive Boosting to diagnose depression by analysing social media posts. To train and test the machine learning models, they initially used two publicly available, labelled Twitter datasets. They then used three additional, non-Twitter depression-class-only datasets to evaluate the algorithms' performance. The first Twitter dataset, which has 6493 depressed and 5384 non-depressed observations, has 11877 total observations, while the second has 10314 total observations, 2314 of which are depressed and 8000 of which are not. The other datasets have 62,50000 and 100 data points, respectively, that only include posts on depression. During pre-processing, components from the NLP toolkit and Peter Norvig’s code for spelling correction were applied. The 10-fold CV method was used

in all experiments, and the findings show that the suggested strategy can successfully identify depression in social media posts even when the training datasets lack specific keywords. They discovered that LR showed better performance of Accuracy 70.52%, Precision 90.47, Recall59.84 and F1 75.26.

Joshi & Kanoongo (2022) have also conducted a comparative study on Twitter sentiment analysis. The dataset contained 43,000 tweets which divided training and test data in a 70:30 ratio. In addition to emoji extraction, stop-word removal, spelling correction and lemmatization, they employed lengthy and efficient pre-processing techniques. Further, the model is trained using a bag of words models to identify the frequency of the term on the text for the predictive model. The 30% of test data is pre-processed to train the model by adding the tweet positive or negative column. The performance of the model is then evaluated using a confusion matrix and they found that the Multinomial Naive Bayes works better than Support Vector Machine (SVM) on their analysis. Although the study appears to be extremely effective, it does not provide the model performance time, which is essential for an early-stage detection.

Saha et al., (2021) conducted a study to understand how well social media data can be utilised to detect depression by utilising linguistic features. The depression dataset was manually created by collecting textual data from Facebook and Twitter communities which was further confirmed by a psychologist. However, the details of data collection and the number of data was not discussed precisely. The work is segmented in to 2 states such as sentiment detection and training the machine learning models. The textblob package of Python is used to identify sentiment analysis and then that data is applied to traditional machine learning models. The models implied were Naive Bayes, Decision Trees, Random Forest(RF), Logistic Regression, Support Vector Machine, Adaboost, Sequential Minimal Optimization, Bagging , Stacking and Multilayer Perceptron. They argue that RF was the one who outperformed with other models with an accuracy of 60.56%, precision 0.585, Recall 0.605 and F1 score 0.547. However, this result does not provide a satisfying performance in detecting depression.

Bi et al., 2021 proposed a User Generated Content (UGC) based method to detect depressed users on Sina Microblog, which is very popular in China, using 3 classifiers such as SVM, RF and LR. The researchers constructed a dataset from Sina Microblog containing depressed and non-depressed users by collecting profile information of the user such as nickname, gender, number of followers and personal signature. The dataset contains 753 depressed and 10163 non-depressed users. The features used in the feature extraction stage were social engagement,

user profile, emotion, and depression-lexicon. These features were vectorized using TF-IDF and then employed classifiers to detect the depression. They discovered that SVM outperformed the LR and RF classification models with a 71% F1-score, 86% Precision, and 60% Recall. Despite their claim of the SVM's high accuracy, there is no supporting evidence in the literature.

Rajaraman et al. (2020) created a deep learning model to detect depression from tweets. For the investigation, they used a deep learning classifier called RNN. The study also uses a variety of Python libraries, including NumPy, Matplot, Scikit Learn, Natural Language Toolkit (NLTK), WordCloud, and Kera. The pre-processed Twitter data is utilised to train and evaluate the algorithm by classifying messages as normal or depressive. During their analysis, the Long Short-Term Memory networks (LSTM), Recurrent Neural Network (RNN), a deep learning classifier, is discovered to be the most accurate classifier after a comparison analysis of several techniques including TF-IDF, Naive-bayes, LSTM, Logistic Regression, and Linear Support Vectors. The LSTM RNN acquired 99% accuracy, precision, and F Measure. Although the study provides a solid understanding of the research, it was unable to provide precise and thorough information regarding the data gathered.

AlSagri & Ykhlef (2020) suggested a machine learning method for identifying depression in Twitter users by considering both their network activity and tweets. Data from 111 user accounts and more than 300,000 tweets were used in this investigation. SVM, Naive-Bayes, and Decision Tree were the classification methods used to find the severity of depression. The best outcomes are produced by the Support Vector Machine (SVM)-linear, which has an F-measure of 0.79 and an accuracy of 82.5. User behaviour and tweets were used to predict the identification of depression. Features are extracted after pre-processing and vectorized using TF-IDF. They argue that the feature extractions boost the accuracy and F1 score.

Islam, Md et al. (2018) conducted a comparative study to detect depression using Facebook comments with 4 machine models such as Decision Tree, k-Nearest Neighbour(KNN), Support Vector Machine, and ensemble. The feature extraction for the study was carried out using the LIWC library. Two experts manually labelled the raw dataset of Facebook comment data into the categories of YES and NO based on the numerical values in the context of psycholinguistic factors. In all, 7145 comments were received, of which 58% were positive indicators of depression and 42% were negative indicators of non-depression. This study even focused on

identifying the most influential time which makes depression symptoms worse and evident that day is the time people feel so lonely. They discovered that Decision Tree surpasses the KNN, SVM, and ensemble models. with the highest precision (59%), recall (98%), and F-measure results (73%). The authors concur that additional research in that area is possible because the study lacked the methods to extract passphrases from a wider variety of emotional traits.

Islam, Md Rafiqul et al. (2018) used a variety of k-Nearest Neighbour (KNN) classifiers, including Fine, Medium, Coarse, Cosine, Cubic, and Weighted KNN to detect depression. The feature extraction was done with the help of the LIWC library. The dataset includes 7145 comments from Facebook, of which 58% indicated depression and 42% indicated non-depression. They discovered that Coarse KNN outperforms all other models with Precision 59%, Recall 0.77%, and F-Measure 67, and that outcomes for all KNN techniques range from 60 to 70% depending on the level of various metrics.

Tadesse et al. (2019) conducted an experiment to increase the performance of the depression detection through a proper feature selection and their multiple combinations. The data was collected from the Reddit social media platform which contains 1293 depressive-indicative posts and 548 standard posts. They employed five text classification techniques, including Logistic Regression, Random Forest, Adaptive Boosting, Support Vector Machine(SVM), and Multilayer Perceptron classifier models, after conducting feature extraction using LIWC, LDA, and TF-IDF. Based on three single feature sets and their various feature combinations, they compare the performance results. They claimed that the suggested approach might considerably raise performance accuracy and SVM classifier was selected as the best single model for detecting depression with 80% accuracy and F1 scores. The Multilayer Perceptron classifier achieves the best performance for depression detection, reaching 91% accuracy and 0.93 F1 scores, and they have identified a lexicon of words more prevalent among the depressed accounts.

Li et al (2020) constructed a depression-domain Chinese lexicon to identify social media users who could be depressed. They examined two data sets from users on Weibo microblogs, a well-known social networking site in China, that contained 58,265 depressed and 52,787 non-depressed users, respectively. Additionally, they used a dataset from Twitter to test the lexicon by choosing 7,455 depressed persons and 10,118 non-depressed users from their 1-year tweets. They automatically produced a lexicon for the depression domain using Word2Vec, a semantic

association graph, and the label propagation method. 111,052 microblog entries from 1868 users 1868 of whom had depression or not were used to create the lexicon. They assessed the performance of detection with and without the lexicon using six characteristics and five classification techniques, including Naïve Bayes (NB), Decision trees, Logistic Regression (LR), Random forests, and Support Vector Machines. The lexicon was able to increase the overall accuracy of the depression detection models. With the aid of the Lexicon, the Logistic Regression beats other models and achieves 77% accuracy, precision, F1 score, and recall. The accuracy, however, drastically decreased when the data set was unbalanced, raising questions about the model's usefulness in a real-world scenario where the data had less than 10% of depressed posts.

Cacheda et al. (2019) proposed a dual Random Forest model for the early detection of depression utilising writing features from postings in the Reddit social network, such as textual spreading, time gap, and time span. Reddit posts totalling 531,394 were used in this study, 49,557 of which were depressive and 481,837 of which were not. Each post consists of a group of tuples of the pattern (id, writing), where id denotes the individual's social network unique id and writing denotes the specific post they posted. They developed a single random forest (RF) classifier with two threshold functions as well as a dual RF, which consists of two independent RF classifiers. The method employs a time-aware methodology that penalises late detections and promotes early detections. The findings demonstrate that a dual model performs noticeably better than the singleton model, however the literature lacks the evidence of the values.

Pirina & Çöltekin (2018) examined the efficacy of training data in detecting depression. They have used 8 different datasets from numerous Reddit forums known as 'subreddits' for their experiment. 400 observations each were used for training and testing purpose. They used multiple classification methods including logistic regression, SVM and recurrent neural networks, in several different settings and stated that the SVM with a combination of word n-grams performed the best. In addition to that, they used 5-fold cross validation and identified a significant drop of F1 measure from 91.40% to 58.28% when evaluating with normal test set. The authors agree that the lower F1 score might be due to the small amount of training data as well as the non-harmonized data sources.

Biradar & Totad (2018) developed a hybrid model with SentiStrength technique and Back Propagation Neural Network (BPNN) to detect depression. They used the Twitter dataset,

which consisted of 61,400 tweets in total. The datasets are gathered by extracting data using the terms "guilt," "sadness," "anxiety," "mental health," suicidal, "tired," and "random," which represent the major 7 symptoms of depression. About 50,000 of the 60,400 tweets were chosen for training, while the remaining 10,000 tweets were used for testing. The sentimental analysis is done using SentiStrength and using the sentiment value they have labelled the observations as depressed or non-depressed. Then the data is used for training the BPNN model. The author claims that the model achieved an accuracy score ranged from 77% to 81% but they failed to provide the values of accuracy, precision and recall analysing the model's performance.

Shah et al., (2020) have proposed a hybrid model using deep learning techniques for early depression detection. The data was chosen from Reddit, which contains 531,453 posts of 892 different users and the users were divided into test (401) and train data (486). The feature selection and classification components were part of the model architecture. They used a range of features, including Glove Embed Features, Word2Vec embed Features and Metadata Features in the feature selection. Additionally, they used 31 helpful LIWC components. Embedded features were given to Bidirectional Long Short-Term Memory (BiLSTM) to enhance the model's forecasting capabilities. The models used hidden layers to address overfitting and binary classification and the person was classified as depressive or non-depressive if the model consistently generated the depressive forecast for the subsequent n posts. They evaluated the models with Latency-weighted F1, Early Risk Detection Error and F1 Score to measure the standard and speed of the model. They found out that Word2VecEmbed+Meta features performed well with a highest F1 Score of 81%, with precision of 78% and recall of 86% at Risk Window 23. However, the model takes too much time to identify participants as depressed, which precludes an early identification.

Cha et al. (2022) proposed a lexicon- based deep learning model in 3 different languages such as Korean, English, or Japanese to early diagnose the high-risk group of university students. They gathered postings from users whose primary languages were Korean, English, or Japanese, totalling 31, 565, and 363 thousand posts, respectively. The authors contend that each keyword in the lexicon dictionaries has been verified by experts because the lexicons were developed in English and translated into other languages. They classified each post as depressed or not using part-of-speech (POS) tagging and the depression lexicon collection. The depressive posts were categorised using CNN, BiLSTM, and BERT, the three baseline classification models. The BERT model can achieve the best performance of 99% accuracy,

precision and F1 score. Additionally, this model was successful in identifying depression in South Korean students using a new dataset from the social network namely ‘Every Time’, and the BERT model achieved an accuracy rating of 77%.

2.2 Summary

In conclusion, the depression can be detected using various lexicon- based as well as machine learning model approaches. Most of the studies (Islam, Md et al., 2018; Chiong et al.,2021; Joshi & Kanoongo, 2022; Li et al.,2020) employed traditional models such as SVM, NB, LR, RF etc where Cha et al. (2022) used deep learning methods such as CNN, BiLSTM etc. Even though the leading state-of-the-art language models, like BERT, XLNet, ELMo, etc., have shown to be effective for language processing, there has been limited research on the depression detection. To determine the sentiment behind each post and the relationship to the participants' depressive states, some research used various feature extraction techniques. Glove, Word2Vec and TF-IDF, n-grams, Bags-of-words , Term Frequency Inverse Document Frequency (TF-IDF) were the main feature extraction techniques used (Shah et al., 2020). Additionally, as text pre-processing is a part of the study, pre-processing techniques have become crucial in the identification of depression using social media data. Number removal, stop-word removal, link removal, emoji removal, lower case conversion, special character removal, tag removal, username removal were the common pre-processing techniques used. Some authors such as Chiong et al. (2021) used lemmatization or stemming, spelling correction, elongated word removal and contraction removal for their studies. For pre-processing, libraries from the NLP toolkit were used, and some studies even included Peter Norvig's code for spelling correction. Although Chiong et al. (2021) claimed that their study didn't check for the phrases "diagnosis" or "depression" in the tweets, the datasets used were labelled based on the keyword "depression". In addition to that, the study by Chiong et al. (2021) shows that the scores were only high when they use the validation dataset generated from the training data itself. This makes suspicion of its performance for a real word experience. Moreover, their comparative study only focused on traditional approaches. Furthermore, no studies examine the reason of the incorrect classification of the tweets. This influence to focus on the reason behind miss classification for future studies. On the top of it, the less studies in pre-trained language models on depression detection leads to open a window to this study.

CHAPTER 3: METHODOLOGY

3.0 Research Methodology

This chapter presents the methods adopted in this study and the justification for chosen approaches will also be discussed. The research problem and the philosophical position have mostly been used to define the methodological decision. Moreover, it discusses why the quantitative and deductive research strategy is found to be appropriate for this study in more depth.

3.1 Research Design

Research designs refer to the processes for gathering, analysing, interpreting, and reporting data in research studies (Creswell & Miller, 2000). It specifies the methods used to collect and analyse data and how all of this is going to answer the research question (Queiros et al., 2017). There are two main types of research design namely, quantitative, and qualitative research design. However, it is worth stating that there is a mix of both methods named as mixed method research design which is gaining popularity recently. The quantitative research design is primarily based on applying mathematical and statistical analysis techniques to examine theories and hypotheses (De Villiers, 2005). It is grounded on the positivism philosophical tradition. Positivism is typically the researcher-created ideology in quantitative studies that focus on discovering truth and presenting empirical evidence. This method entails employing statistical procedures to generate a valid generalizable result (Queiros et al., 2017). The process begins with theory and progresses to data, which is then narrowed down to generalizable results. The study design is well-liked in the research community due to its statistically reliable outcome or generalizability. However, has been chastised for failing to investigate humans and their behaviour in depth (Amaturo & Punziano, 2017). While the qualitative research design entails comprehending the significance of social phenomena (Queiros et al.2017). It is based on the philosophical tradition associated with post-positivism. Data from observations, interviews, focus groups, audio and video recording are used for this type of research. This research design focuses on data from a small number of group and new theories/hypothesis are then developed by grouping, classifying, and analysing that data (De Villiers, 2005). Even though this method is effective to understand the feelings behind the person correctly, the chance of bias is very high due to the lack of generalization (Amaturo & Punziano, 2017).

The deductive approach starts with an established theory and testing it with the collected data (Saunders, 2015). It can be considered as a ‘top-down’ approach which means it's theory and hypothesis are dependent. Whilst the inductive approach is considered as a data driven approach which is grounded up with the collected data (Saunders, 2015). In contrast to deductive approach, the inductive approach can be considered as a ‘bottom-up’ approach which emphasises on the observation to conclude. Also, it does not need a pre-trained hypothesis or framework (Woo et al. 2017). The increased data availability influences the evolution of natural language processing research methods. This demonstrates the significance of mixed research methods as an investigation toolkit when considering their strengths and weaknesses (Hesse-Biber & Johnson, 2013). This study has a solid foundational theory based on literature reviews; however, we are not using a fully deductive methodology because we are aiming to create a better, more modern framework. To summarise, this study employs a quantitative and deductive approach, as well as statistical approaches, to predict the depression sentiments underlying people's tweets. This research will be enhanced by understanding the sentiment behind people's posts to recognise their mental state using language learning techniques and provide a proper classification. The specific timeline of the project is to adapt the time horizon into cross-sectional and to complete within the timeline.

In the previous chapter, literature review findings showed that there are no studies that utilised the advantage of latest language models such as Bert, XLNet and ELMo. The literature review shows the lack of comparative studies between traditional machine learning models and the transformer language models to detect depression even though people compared different traditional models. On top of it, the studies do not have a comprehensive discussion about the misclassification. Additionally, most of the studies used validation set from the same training dataset. Based on this, this chapter focuses on the development of depression detection model and will therefore compare different well performing models to ascertain the best performing model in this field of study. In addition to that, the chapter describes how the validation dataset is generated and annotated to make the depression detection.

Moreover, this chapter will provide an overview about the machine learning models and statistical analysis technologies employed in this domain to fill the identified research gap and accomplish the research goal. The chosen methods' rationale is to be subject to their performance in text classification tasks, particularly in natural language processing as reviewed

in previous chapter. On the top of it, this chapter gives an insight to the processes for data gathering, data wrangling and the data exploration using quantitative approaches for fulfilling the research aim. Finally, this chapter defines the ethical implications behind the study. The research question, objective, and gap that will guide this chapter are shown in Table 3.0.

Research Question	How can Twitter sentiment analysis help detect early-stage depression?
Research Objective	To conduct a comparative study to detect depression using Twitter data.
Research Gap	1) The lack of comparative study on depression detection using state-of-the-art methods. 2) There is not enough discussion about the reason behind misclassification of depression.

Table 3.0: Research guide for this chapter

3.2 Research Strategy and Method

The research strategy is considered as the action plan because it refers to how the research will be conducted in practical terms, based on the research question, aim and objectives. While the research design focuses on the procedures for collecting, evaluating, interpreting, and reporting data in research investigations (Creswell & Miller, 2000). It outlines the methods used to collect and analyse data as well as how all of this will help to answer the research question(Queiros et al.2017). Social media platforms have become a popular place for people to openly express their emotions and feelings, making them a valuable resource for learning about their mental health. Depression symptoms can be identified more easily and quickly by using classification models. The study delves into the specifics of the quantitative and qualitative research designs, as well as the deductive versus inductive approach. Based on the data collected and methods applied, the research design can be classified as quantitative research.

3.3 Research Ethics

Ethics in research refers to the moral principles that guide research. This dissertation is being done under the supervision of the university, which is subjected to high ethical standards. On

the top of that, I have make sure the study satisfied beneficence, non-malfeasance, integrity, Informed Consent, Anonymity and Impartiality. This research is mainly focused on benefiting the society by detecting early-stage detection problems to prevent the damage in society due to it worsen stage by people's ignorance. It is beneficial to the healthcare sector to easily identify depression using social media posts apart from the old traditional way such as questionnaires or consultations. This study does not do any harm to vulnerable people such as kids or any specific groups because it is purely based on the analysis of already collected social media data. Integrity is assured by the appropriately qualified and experienced supervision of the project. To fulfil the requirement of the university, I have filled the UREC2 Ethical form, (low risk) human participation since it involves the human data which is used to analyse and create the framework. The Ethics form is attached to the appendix-B.

3.4 Data Collection

Data gathering is one of the most important phases of a research. Data is divided into two categories, referred to as primary data and secondary data, depending on the methods used to collect it. Primary data are those that the researcher collects for the first-time using interviews, surveys, questionnaires, experiments etc while secondary data is gathered from open data sources, articles, journals, web pages, books, and blogs. This study involves the data from twitter which have been already collected and annotated in numerous existing studies and examined social media depression. So, the secondary source of is considered suitable for this research, considering the time delay and cost associated with data extraction. In this study, we have utilised the same datasets used by Chiong et al. (2021) in their comparative study to train our models. The 1st dataset is from Shen et al. (2017) and the second dataset is an open dataset from Kaggle developed by the author Eye (2020). Both datasets were collected from Twitter and were already labelled. To test the model's performance in a real-world entity, a validation dataset was created by randomly selecting 250 depressed and 250 non-depressed datasets from Eye's (2020) dataset. This validation dataset was manually annotated, verified by an expert, and used to test the models generated. The main columns of these three datasets are tweet and label, where tweet represents each tweet of the user and label represents depression status annotations. The label column contains 1 and 0 which indicates depressed and non-depressed respectively.

Shen et al. (2017) added the limitation that a record would only be labelled as "Depression" if its anchor tweets strictly followed the pattern "(I'm/I was/I am/I've been) diagnosed with

depression." The record would be labelled as "non-Depression" if the user had never posted any tweets containing the character string "depress". The sizes of the depression (first set) and non-depression (second set) classes are nearly equal, totalling 6493 and 5384, respectively. The statistic of the dataset is added in the Figure 3.1.

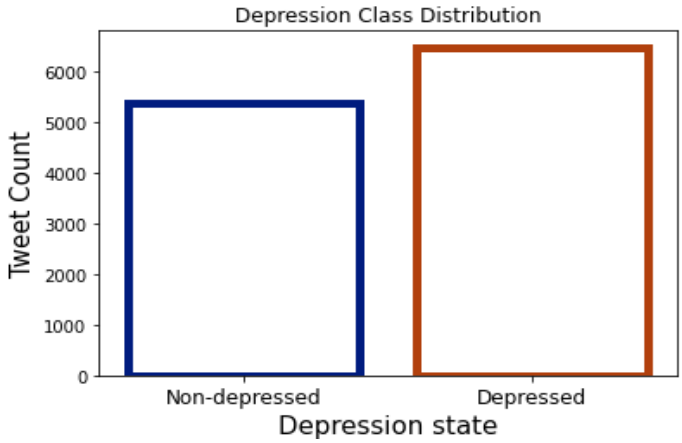


Figure 3.1: Shen et al.'s dataset Distribution

The Eye (2020) dataset, on the other hand, only looks for the word 'depression' in tweets. If a tweet contains the word "depression," it is labelled as depressive otherwise non-depressive. This dataset is highly skewed, with depression class records accounting for only 22% of all records. That is, out of a total of 10314 records, 2314 are depressed and 8000 are non-depressed. The class distribution added in the Figure3.2.

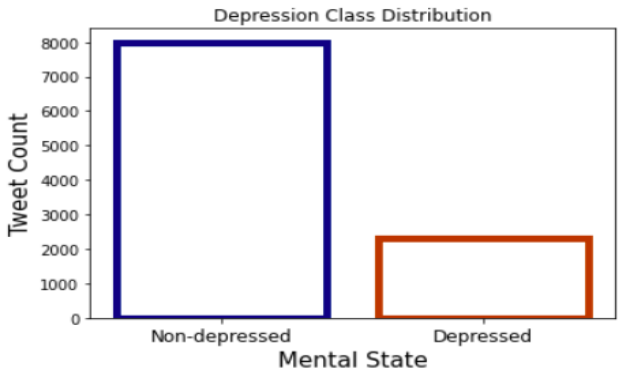


Figure 3.2: Eye's dataset Distribution

The validation dataset was prepared by taking 250 depressed and 250 non-depressed data from the Eye's (2020) dataset which have been annotated manually by two human annotators. The validations were verified with a second annotator to confirm the annotations made correctly.

After the annotation, the validation set contains 150 depressed and 350 non-depressed tweets. The class distribution of the validation dataset is added in the Figure 3.3.

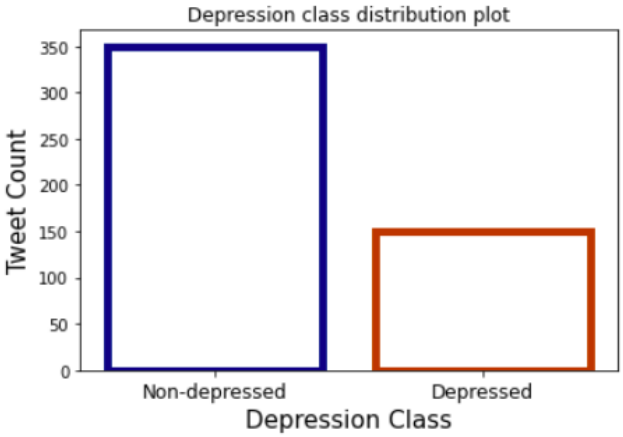


Figure 3.3 Validation dataset class distribution

The detailed descriptions of all the datasets used in the study are shown in below Table 3.1.

Dataset	**Total Records**	**Depression Records**		**Non-depression records**	
		Total	**%**	**Total**	**%**
Shen et al. (2017)	11847	6493	54.67	5384	45.33
Eye (2020)	10314	2314	22.44	8000	77.56
Validation Dataset	500	150	30	350	70

Table 3.1 Depression dataset description3.5 Tools and Technologies

3.5.1 Depression Detection Framework

To fill the research gap identified in chapter 2 and provide an efficient depression detection system, this comparative study uses multiple language models as well as traditional machine learning models. The various tasks of this framework are Dataset collection, Validation set creation by manual annotation, Text pre-processing, Vectorization using TF-IDF and Word2vec to identify the sentiment, model training and validation of the models. The programming language chosen for this project is Python because of its platform independence, less complexity, better readability, and huge community support. Moreover, python has a

significant role in data analytics, machine learning, data science, data engineering, and even artificial intelligence due to its library support, especially NLTK, Scikit learn etc (Lasser et al. 2021; Robinson, 2017). This research makes use of the Python libraries namely NumPy, Matplot, Scikit Learn, NLTK, WordCloud, and Kera. The support of tensor flow and Py-torch makes it easier to use the pre-trained models deployed in TensorFlow and Py-torch. The architecture of the proposed Early Stage Depression Detection System (EDDS) is shown in Figure 3.4 below.

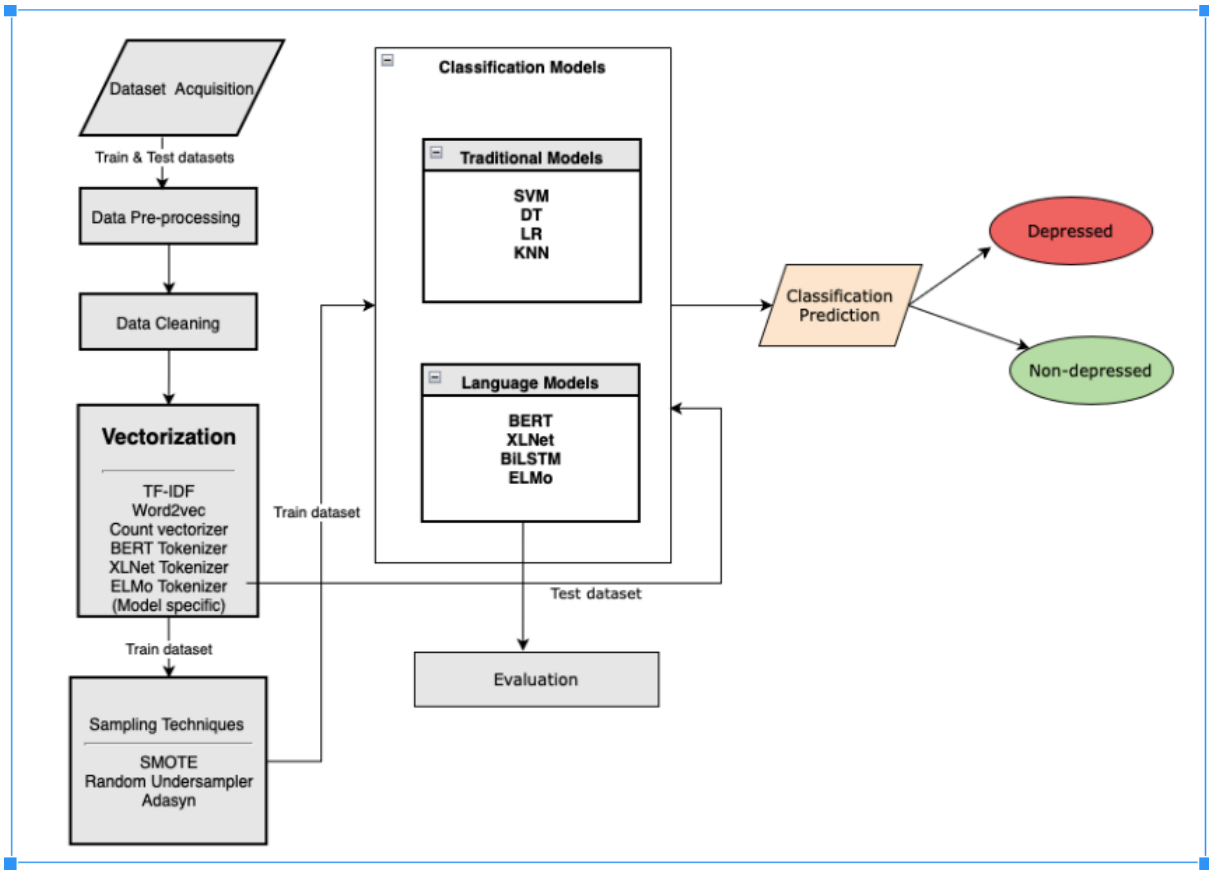


Figure 3.4. EDDS Proposed Framework

3.5.2 Text pre-processing Techniques

The designated framework for detecting depression is to make use of social media text data by using various methods to pre-process the data before feeding it to the models. In this study we have used the proposed techniques used by Budhi et al., (2021) and Chiong et al. (2021) in their research.

- Convert to lowercase: To ensure that words are processed appropriately in text classification, especially in natural language processing, words must be changed to lowercase. Therefore, all texts have been changed to lowercase.

- Removal of punctuation, numbers, and URLs: Numbers, punctuation, and URL links can add unwanted noise to text data. Consequently, eliminated to simplify text processing.

- Removal of Stop-words: Stop-words do not add anything to the sentence's meaning (Dey et al. 2020). Therefore, removing words like "a," "the," "is," and "are" will lessen the unwanted processing they each cause. However, since the goal of this study is to detect depression by analysing the sentiment underlying the texts, not and can stop-words be kept.

- Removal of special characters, tags etc: Social media messages are packed with unnecessary phrases, tags, usernames, special characters, emojis, and other text. These words or characters are eliminated (Fithriasari et al. 2020).

- Negative word correction: Although negative words can take many different forms, all of them aim to incorporate negativity into the sentence. So, we have changed them to their fundamental "not" negative form.

- Lemmatization: Lemmatization is an algorithmic process of finding the lemma/base form of a word depending on their meaning. This is an essential step to reduce the word diversity and easy recognition.

Table 3.2 below shows the example of raw vs cleaned data.

Raw Tweet	Clean Tweet
@lapcat Need to send 'em to my accountant tomorrow. Oddly, I wasn't even referring to my taxes. Those are supporting evidence, though. "	need send em accountant tomorrow oddly even refer tax support evidence though
"I'm suffering from depression, I'm thankful that you guys are helping me out, really, I love you guys so much even if you don't notice this, I'll always support y'all because its what you deserve<Emoji: Loudly crying face><Emoji: Two hearts><Emoji: Two hearts> #StrayQuiz #StrayKids @Stray_Kids"	suffer depression thankful guy help really love guy much even notice always support deserve loudly cry face two hearts two hearts

Table 3.2: Raw vs Clean Tweet

3.5.3 Exploratory Data Analysis

Exploratory Data Analytics (EDA) is a common statistical approach to understand the data thoroughly. In this study, we have used WordCloud to understand the frequent words used in each class. The word cloud (figure 3.5 and 3.6) shows a visual representation of a text, in which the words appear bigger the more often they are mentioned. From this WordCloud, we can see that “depression”, “anxiety” and “diagnosed” are frequently used for depressive tweets while “thank”, “good”, “day” are used for non-depressive tweets. Figure 3.5 and Figure 3.6 shows the WordCloud of depressed tweets and non-depressed tweets of the two datasets.

Figure 3.5: Shen et al.'s dataset depression (left) and non-depression (right) tweets WordCloud

Figure 3.6: Eye's dataset depression (left) and non-depression (right) tweets WordCloud

The null value handling and outlier removal is also performed as part of the EDA. The duplicate removal and the outlier removal is also performed as part of data cleaning. The outlier removal is performed by removing a tweet less than 2 words size since that cannot be considered as a tweet. Shen et al.'s dataset retained the same shape of (11877,2) while Eye's dataset changed its shape to (10280, 2).

The distributions of the datasets are already added in section 3.4. Figure 3.2 shows that the Eye's dataset is highly imbalanced. Therefore, next section discusses appropriate strategies for dealing the class imbalance issues.

3.5.4 Handling Class-imbalance

The class imbalance is a problem where the classifier has a bias towards the majority class which may cause over-fitting issue (He & Garcia 2009). The majority classes are predicted correctly, whereas the minority classes are often mis-classified. In the case of our Eye's dataset, the non-depressed cases are very high with respect to the depressed classes. So, the depressed tweets have a high tendency to be mis-classified as non-depressed. Since the Shen et al.'s dataset is almost evenly distributed, the class imbalance issue will not be a problem for the dataset.

There are several approaches to handle class imbalance issues such as undersampling, oversampling, hybrid sampling etc. The oversampling creates exact copies of existing minority classes data whereas undersampling methods delete or merge examples in the majority class.

Hybrid approach is a combination of both these methods. Since this study contains sensitive data, oversampling techniques are appropriate. However, the deep class imbalance study is out of the scope. For the language models, this study uses weighted samplers to solve the class imbalance issue whereas oversampling techniques are used in the case of traditional ones. SMOTE, RandomOverSampler and ADASYN are the sampling techniques used in this study.

I. RandomOverSampler

Random oversampling is the process of randomly choosing samples from the minority class, replacing them, and adding them to the training dataset. Although this strategy is appropriate in this situation, many have criticised it since the randomly reproduced samples lead to overfitting.

II. SMOTE

SMOTE (Synthetic Minority oversampling method) was created by Chawla et al. to resolve the RandomOverSampler issue (2002). With synthetic samples created by random interpolation between many minority samples that lay together, the SMOTE technique sought to rebalance the class distribution. It is used to provide a training set that is artificially class-balanced or nearly class-balanced in order to train the classifier.

III. ADASYN

He et al. (2008) suggested adaptive synthetic sampling (ADASYN), which is similar to SMOTE but produces a different number of samples based on an estimate of the local distribution of the class to be oversampled. It produces more synthetic observations for minority class observations that have more majority class observations inside the k-nearest neighbours region.

3.5.5 Feature Representation of Text

Text (input) data is not directly understood by machine learning models. Only numeric representation can be recognised by these models. The discussion of several methods to that goal is part of this stage. This is crucial since it enhances the performance of the classifiers and supports models' efficiency in text classification tasks (Wang et al. 2019). The mathematical representation of unstructured text as a numeric vector in a computing system is called the "vector space model" (Salton et al. 1975). For validation dataset, count vectorizer is used as a

feature representation tool. TF IDF and Word2vec models are two more sophisticated vector models that I have experimented with and are explained in the below section.

I. TF-IDF

The TF-IDF weighing approach assigns a distinct weight to each term in a document based on the frequency of terms in all documents. It will make it easier to understand the word's significance within the corpus. This study uses TF-IDF because of its higher performance, particularly in terms of increasing recall and precision values. The model is the result of multiplying the term frequency (TF) and inverse document frequency (IDF). TF is the ratio of a word's frequency in a sentence to the sentence's overall length. The log of the ratio between the total number of rows and the number of rows containing the word is known as the IDF. This determines how frequent a word is. For example, if the word "dog" appears 10 times in a passage of 100 words and 20 words, it is more important in the second instance. The TF-IDF vectorizer from Sklearn feature extraction library (Pedregosa et al. 2011) was used to implement the TF-IDF model. The default value of n-gram is given as 1 and used as the unigram for this study where the parameters are tuned to maximum features of 5000. TF-IDF parameter chosen implies that the model will transform text of unigram terms into TF-IDF vector model. The count vectorizer from Sklearn feature extraction library (Pedregosa et al. 2011) was used to create vectors for the validation dataset since the TF-IDF is used only to fit the training data.

II. Word Embedding

It has been demonstrated in the previous studies done by Kim, (2014) and Ruder et al. (2016) that word embedding enhances text classification model performance, particularly when used as a deep learning model for sentiment classification. This is because care is given to the semantics, structure, order, and context in which words are employed in a document. Word2Vec is taken into consideration in this study to train and generate word embeddings and for creating ELMo and BiLSTM, we use the Keras Neural Network model as a base model. Word2vec, an unsupervised model used for this study, creates high quality vector representations of words that capture semantic and context. It uses the distribution of word co-occurrences in the context to learn word embeddings. Word2vec from Gensim (Rehurek & Sojka, 2010) was used in this study to train and produce word embeddings. This aids in capturing the semantic relationship between words and context. Tokenized words from the

cleaned Tweets were used in the implementation. The minimum word count was set to 1, and the context window size was set to 3. These vector space models will be used to track variations in classification model performance.

3.5.6 Machine Learning Models

3.5.6.1 Traditional Machine Learning Models

The chapter 2 literature review shows that the depression detection using traditional models are enriched and have shown good performance. Unfortunately, there is no comparative study that validates the performance with a manually annotated dataset to ensure the performance of these models. Logistic Regression, SVM, K Nearest Neighbour and Decision Tree are the models selected based of the performance shown in previous studies.

I. Logistic Regression

Logistic regression is a generalised linear model, which was developed by Nelder and Wedderburn (1972) and then Hastie and Tibshirani (1990) enhanced it later. These models use continuous, and normally distributed dependent variables to overcome the limitation of linear models. In LR, the independent predictor variables can either be interval/ratio or dummy variables, while the dependent variables can either be ordered polytomous or unordered polytomous.

II. SVM

The SVM is a supervised learning model that uses training data to classify new data. It works by dividing different classes with a hyperplane and then attempting to maximise the distance between them. The greater the distance, the smaller the error produced by the classifier. In this study, the SVM is combined with the linear kernel (LSVM), which is commonly used for text classification.

III. Decision Tree

Quinlan (1986) created the DT, which is based on Hunt's (1966) algorithm, and it is a beneficial tool for studying the cause-and-effect chain. It creates a tree-like decision model for

classification, prediction and is commonly used as a base classifier in ensemble models (e.g., BP, RF, and AB).

IV. K Nearest Neighbour

KNN is an effective supervised learning model that makes no data or functional form assumptions. The model is extensively used in sentiment analysis applications. (Dey et al. 2016), and it outperforms other models (Jain & Katkar, 2015). The non-parametric classification technique classifies instances by calculating the nearest neighbour using distance metrics among the training set and then predicting to the test set by choosing the nearest majority.

3.5.6.2 Language Learning Models

The Pre-trained Language Model (PLM) describes a saved network developed by another entity and trained on a huge dataset to address a similar issue. So, this model can be used as a starting point to create another model. Pre-trained models have various benefits, including the ability to eliminate the requirement of huge, labelled dataset for training and their applicability to several use cases. It also improves the efficiency without fine-tuning. Moreover, the expense of using specialised computational architecture, such as Tensor Processing Units (TPU), can be avoided since we are not developing from scratch (Arslan et al., 2021). PLMs gained more importance after BERT delivered best results on 11 NLP tasks (Devlin et al., 2018). Hence, we have used some well-known PLMs in this study including BERT, XLNet and ELMo. These models will be covered in depth in the following sections.

I. Bidirectional Encoder Representations from Transformers (BERT)

Bidirectional Transformer Encoder Representations (BERT) model is a transformer-based machine learning model developed by Google in 2018 for NLP. This is used to aid Google AI Language in the pre-training of deep bidirectional representations for extracting context-sensitive properties from input text (Devlin et al., 2018). This model has been acclaimed by the machine learning community for producing cutting-edge results on a variety of NLP tasks, including sentiment analysis, question answering, and natural language inference. BERT learns contextual relationships between words in a text using the transformer's attention mechanism. Transformer is composed of two mechanisms: an encoder for reading text input and a decoder

for task prediction. BERT is made up of 12 transformer-encoder blocks stacked on top of each other. A multi-headed self-attention layer exists within each of these encoders. The Feed-forward network hidden layer size has been increased from 512 to 768 in comparison to the Transformer to accommodate the increased number of attention heads (Devlin, Chang, Lee, & Toutanova, 2018). Figure 3.7 represents the architecture of BERT. In our study, we have been fine-tuning the pre-trained model with default parameters and then trained using our labelled dataset for text classification. To identify and separate each sentence we will add a token named [CLS]. The hidden state matching to [CLS] token is used as the aggregate sequence representation for our classification tasks. The token, segmentation, position embeddings based on the input embeddings are added in Figure 3.8.

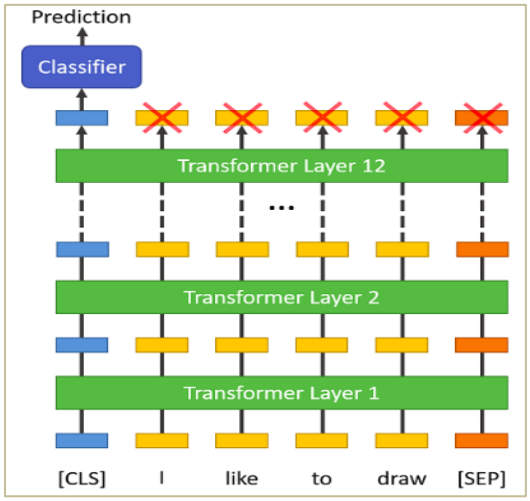


Figure 3.7 Illustration of BERT model architecture

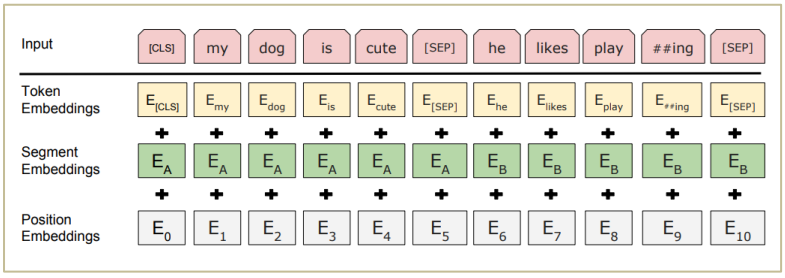


Figure 3.8 BERT input representation (Devlin et al., 2018)

II. XLNet

As previously stated, the main advantage of BERT is its fast convergence time and ability to predict a target token using bi-directional context. However, Yang et al. (2019) pointed out some drawbacks. These drawbacks include noise input and the assumption of independence.

BERT corrupts the input during pre-training with the "[MASK]" token, which does not appear during the fine-tuning process, resulting in pre-training fine-tuning inconsistency. XLNet is pre-trained with a permutation language model (PLM), which allows it to incorporate the auto-regressive nature of language models. Unlike BERT, XLNet can learn from the dependencies between target tokens.

III. Embedding from Language Models (ELMo)

ELMo is a deep contextualization-based text representation. ELMo aids in overcoming the limitations of traditional word embedding methods such as LSA, TF-IDF, and n-grams models (Lim et al., 2020). It recognises both the meaning of words and the context in which they occur, in contrast to TF-IDF and n-grams representations, which only record the frequency and presence of the words and are unsure of the context. The Long Short-Term Memory (LSTM) theorem is used by ELMo (Lim et al., 2020). It employs a bi-directional LSTM that has been trained on a specific task and can create contextual word embedding. The architecture of the ELMo word embedding is depicted in the Figure 3.9 below.

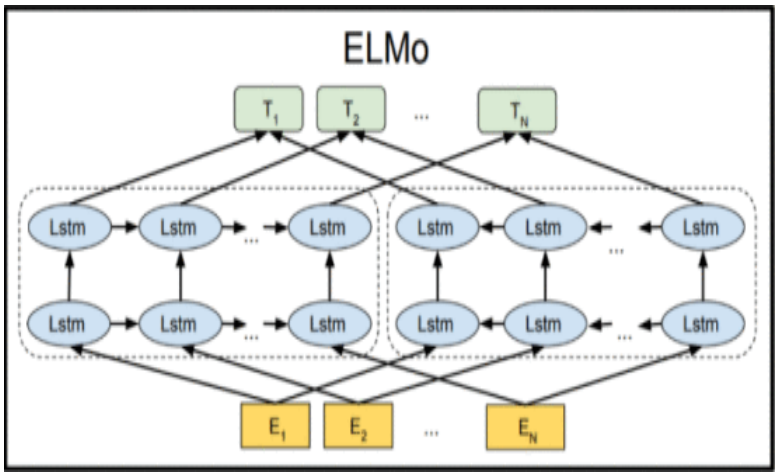


Figure 3.9: Illustration of ELMo architecture

IV. BiLSTM

Long Short-Term Memory (LSTM) networks are a type of RNN model that addresses the problem of vanishing gradients. Based on training, it learns to keep the relevant sentence content and discard the irrelevant ones. Using dynamic gates known as memory cells, this model preserves gradients over time (Xu et al., 2019). A gate can erase, write, and read data

from a memory cell at each input state. Gate values are calculated by combining the current input with the previous state in linear fashion.

The Bidirectional LSTM (BiLSTM) model has two unique states for forward and backward inputs produced by two independent LSTMs. The first LSTM feeds a normal sequence that starts at the start of the phrase, but the second LSTM feeds the input sequence backwards. A bi-directional network's purpose is to gather data from neighbouring inputs. Although it depends on the task, it usually learns faster than a one-directional approach and all inputs are treated equally by the model (Xu et al., 2019). The sentiment polarity of the text is heavily dependent on the words containing sentiment information when performing sentiment analysis. The sentiment reinforcement of the sentiment word vector is realised.

3.5.7 Model Evaluation

This section describes the model evaluation metrics to understand how well they performed in this context and to assist in deciding the best model. In this study, we use the confusion matrix parameters such as Accuracy, precision, recall, and f1-measure for performance evaluation. The proportion of correctly and incorrectly labelled tweets is shown as a 2 by 2 confusion matrix as true positive, true negative, false positive, and false negative. The confusion matrix figure is shown below Figure 3.10.

		Predicted	
		Negative **(N)** -	Positive **(P)** +
Actual	Negative -	True Negative **(TN)**	**False Positive (FP) Type I Error**
	Positive +	**False Negative (FN) Type II Error**	True Positive **(TP)**

Figure 3.10: Confusion matrix illustration

True Positive (TP): Correctly predicted Positive tweets

False Positive (FP): Negative tweets incorrectly predicted as positive (Type I error)

False Negative (FN): Positive tweets incorrectly predicted as negative (Type II error)

True Negative (TN): Correctly predicted Negative tweets.

The following Figure 3.11 shows what the parameters and how it is calculated.

Accuracy is the ratio of the number of samples predicted correctly to the total number of samples.

$$Accuracy = (TP+TN)/(TP+FP+TN+FN)$$

Precision: is the ratio of the number of total positives to the total predicted positive values.

$$Precision= TP/(TP+FP)$$

Recall: is the ratio of true positive to total positive samples.

$$Recall = TP/(TP+FN)$$

F1-score is the harmonic mean of precision and recall.

Figure 3.11 Calculation of confusion matrix parameters

3.6 Chapter Summary

According to the research objective, aim, and question, this chapter explained the research strategies, methods, and methodology used in this study. This study is quantitative research that makes use of Twitter data to detect depressive symptoms by examining the tone of the tweets, and the rationale for choosing a quantitative approach. The information employed comes from indirect observations, which entail textual analysis of material produced directly from narratives like tweets. Human beings do not have any detrimental causes. More than that, the intention behind the study is to provide benefits to society. This study is conducted based on ethical principles of low-risk form.

This chapter also gives a clear outline of the framework development and the tools and techniques used. Data pre-processing, EDA, Feature Text Implementation, Model Development, and Statistical Evaluation are the main stages of this research. Removal of numbers, links, tags, username, contraction, stop word, and emoji are the key data pre-processing procedures. EDA is performed to identify the pattern/trends in the data. The data is then vectorized to make it readable to the machine learning models. For traditional models TF-IDF is used while the language models used its own pre-trained tokenizers and data loaders to convert the text to machine readable format. However, the BiLSTM, ELMo makes use of Keras neural networks, therefore it generated the vector using Word2vec unsupervised models. The data loaders(tensors) can be then used to train the models. The class imbalance problem of the Eye's dataset is also investigated and solved using the sampling techniques namely SMOTE, RandomOverSampler, ADASYN. This will help the models to avoid overfitting issue. The supervised classification models adopted were also discussed in this chapter. Likewise, the statistical valuation parameters such as Accuracy, F1 score, Precision and Recall and its calculation are also described clearly.

CHAPTER 4: RESULTS AND DISCUSSION

4.0 Introduction

This chapter discusses the findings and results from the study conducted by utilising the tools and technologies specified in the previous chapter. As mentioned in the 3.4 section, the study is conducted as 2 cases with 2 different datasets and the validation is done by a manually annotated dataset.

4.1 The validation dataset

The framework is created to identify depression in its early stages. Some of the words like 'depress', 'diagnose' are often misclassified as false positive results. To validate the performance of the models, a manually annotated dataset became very important to ensure performance. The validation dataset is prepared from randomly choosing 250 depressed and 250 non-depressed observations, and further annotated by understanding the sentiment behind it. The annotation was done by 2 human annotators and confirmed with an expert in case of disagreement. After the annotation the depressed entries are reduced to 130 and the non-depressed entries are increased to 370. The class distribution of the validation dataset is added below as Figure 4.1.

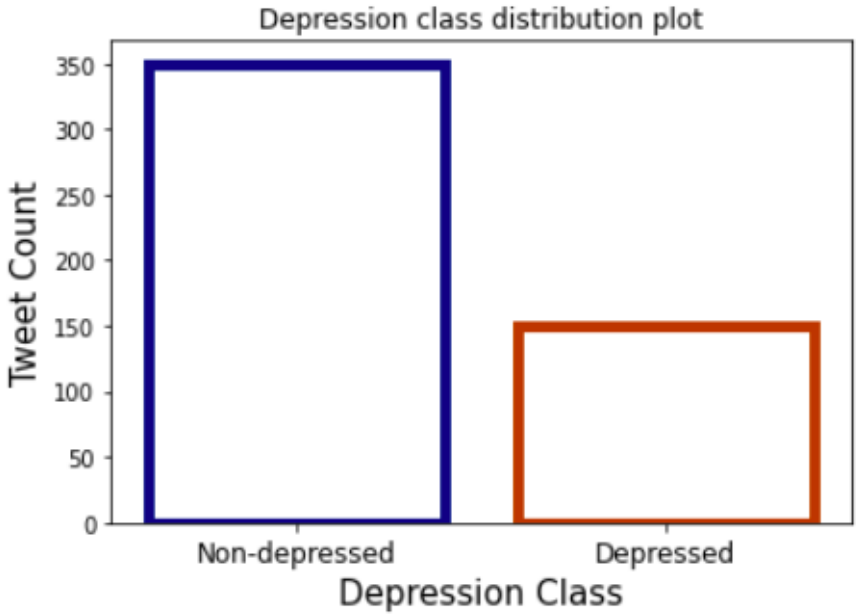


Figure 4.1 Validation Dataset class distribution

4.2 Result

This section details the performance of the models which have been trained with 2 datasets using Scikit learn confusion matrix parameters. The details of the confusion metrics parameters are added in section 3.5.7. Due to the suspicions regarding the overfitting problem while using the test set from the same data, the validation set was developed and used for both validations. In this study, the Depressed cases are considered as positive, and the non-depressed cases are considered as negative classes. As mentioned in the section 3.5.6, the models employed are SVM, Decision Tree, Logistic Regression, KNN as traditional learning models and BERT, XLNet, ELMo and BiLSTM as language models. The result on each case is presented below.

4.2.1 Case 1: Experiment with Shen et al.'s dataset

The following table shows the results of the models with the dataset from Shen et al (2017).

Model	Sampling Technique	Accuracy	Class	Precision	Recall	F1 score	Support	Confusion matrix
BERT		**76%**	0	0.9	0.74	0.81	350	[[258 92] [30 120]]
			1	**0.57**	**0.8**	**0.66**	150	
XLNet		74%	0	0.81	0.82	0.82	350	[[287 63] [67 83]]
			1	0.57	0.55	0.56	150	
BiLSTM		73%	0	0.79	0.84	0.81	350	[[294 56] [78 72]]
			1	0.56	0.48	0.52	150	
ELMo		68%	0	0.75	0.81	0.78	350	[[283 67] [92 58]]
			1	0.46	0.39	0.42	150	
SVM		74%	0	0.78	0.87	0.82	350	[[304 46] [84 66]]
			1	0.59	0.44	0.5	150	
SVM	SMOTE	74%	0	0.78	0.87	0.82	350	[[303 47] [84 66]]
			1	0.58	0.44	0.5	150	
SVM	OVERSAMPLER	74%	0	0.78	0.87	0.82	350	[[304 46] [84 66]]
			1	0.59	0.44	0.5	150	
SVM	ADASYN	74%	0	0.78	0.87	0.82	350	[[303 47]

			1	0.58	0.44	0.5	150	[84 66]]
LR		74%	0	0.78	0.87	0.83	350	[[306 44] [54 66]]
			1	0.6	0.44	0.51	150	
LR	SMOTE	74%	0	0.77	0.89	0.83	350	[[313 37] [95 55]]
			1	0.6	0.37	0.45	150	
LR	OVERSAMPLER	73%	0	0.76	0.89	0.82	350	[[313 37] [97 53]]
			1	0.59	0.35	0.44	150	
LR	ADASYN	73%	0	0.77	0.89	0.82	350	[[311 39] [94 56]]
			1	0.59	0.37	0.46	150	
DT		76%	0	0.84	0.81	0.82	350	[[204 66] [56 94]]
			1	0.59	0.63	0.61	150	
DT	SMOTE	75%	0	0.83	0.8	0.82	350	[[281 69] [56 94]]
			1	0.58	0.63	0.6	150	
DT	OVERSAMPLER	75%	0	0.83	0.81	0.82	350	[[281 69] [56 94]]
			1	0.58	0.63	0.6	150	
DT	ADASYN	75%	0	0.83	0.81	0.82	350	[[282 68] [56 94]]
			1	0.58	0.63	0.6	150	
KNN		59%	0	0.59	0.59	0.67	350	[[207 143] [64 86]]
			1	0.57	0.57	0.45	150	
KNN	SMOTE	67%	0	0.72	0.86	0.78	350	[[301 49] [118 32]]
			1	0.4	0.21	0.28	150	
KNN	OVERSAMPLER	60%	0	0.75	0.64	0.69	350	[[225 125] [76 74]]
			1	0.37	0.49	0.42	150	
KNN	ADASYN	67%	0	0.72	0.87	0.79	350	[[305 45] [119 31]]
			1	0.41	0.21	0.27	150	

Table 4.1 Result of Shen et al.'s dataset

The Table 4.1 above presents the statistical results of all the machine learning models which are trained using Shen et al.'s dataset. This table makes it quite evident that almost all models performed better at predicting negative cases than positive ones. Also, it shows that **BERT** outperformed all the models with **76%** accuracy. However, as the goal of this study is to identify depressed individuals, our primary focus is on accurate positive case prediction. When traditional models and language models are compared, it is evident that the language models outperformed the traditional methods in terms of accurately predicting True positive values. With a **Precision score of 57%, Recall score of 0.8%, and F1 score of 66%**, **BERT** surpassed all models. The model is the best one from this research because it only misclassified 33 positive cases. Furthermore, it showed an overall accuracy score of 76%.

When comparing the traditional models together Decision Tree performed better than SVM, LR and KNN models. It got a Precision score of 58%, Recall score of 63%, and F1 score of 60%. Surprisingly, the Decision Tree has not shown any difference when using samplers like SMOTE, RandomOverSampler, and ADASYN. KNN without sampling methods became the least performing model with an accuracy of 59%. The comparison of the evaluation matrix is added in Figure 4.2.

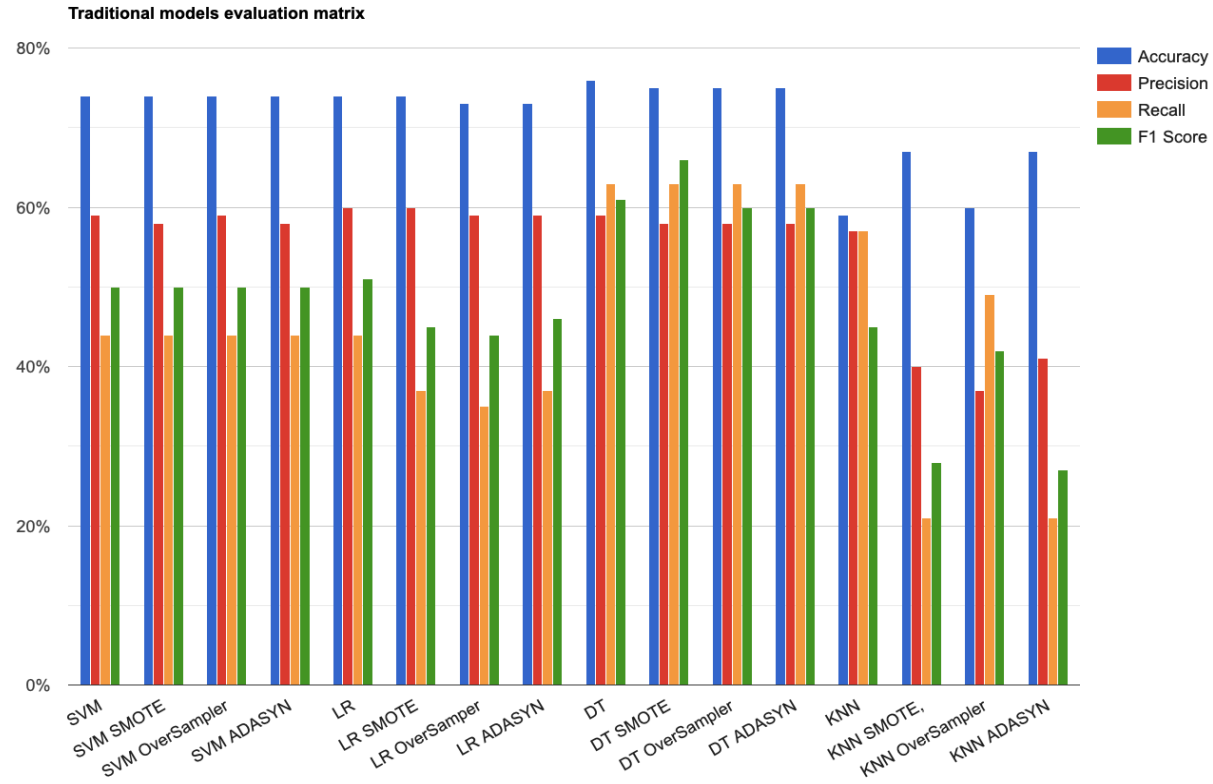


Figure 4.2 Evaluation matrix of traditional models with Shen et al's dataset

The comparison of language models reveals that BERT predicted both negative and positive classes more accurately than XLNet, ELMo, and BiLSTM, while ELMo was the model that did the worst, with an accuracy of 68%. More than 50% of depressed tweets were incorrectly labelled as non-depressed. The comparison between the language models for the positive class is shown in the Figure 4.3.

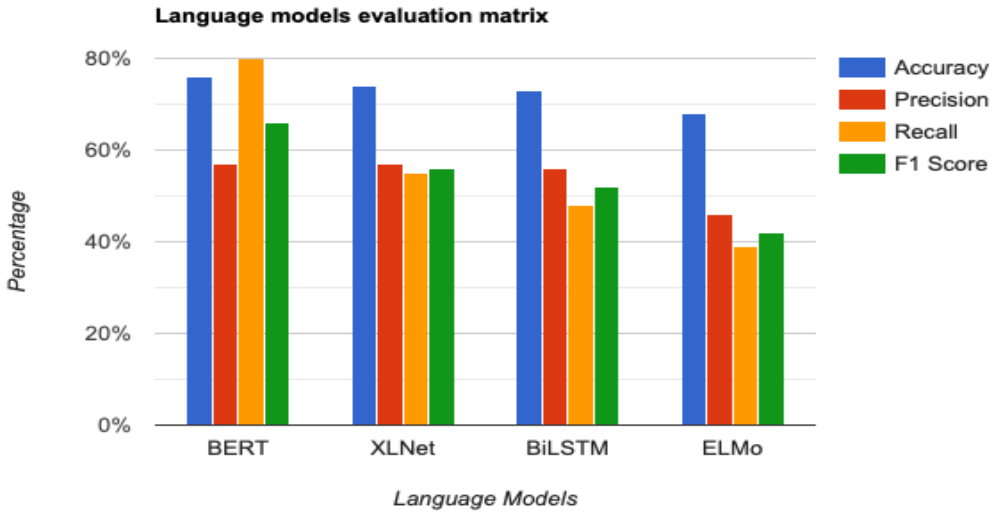


Figure 4.3 Evaluation matrix of language models with Shen et al's dataset

4.2.2 Case 2: Experiment with Eye's dataset

The following Table 4.2 shows the results of the models with the dataset from Eye (2020).

Model	Sampling Technique	Accuracy	Class	Precision	Recall	F1 score	Support	Confusion matrix
BERT		79%	0	0.99	0.71	0.83	350	[[249 101] [2 148]]
			1	0.59	0.99	0.74	150	
XLNet		**90 %**	0	0.95	0.93	0.94	350	[[326 24] [17 133]]
			1	**0.85**	**0.89**	**0.87**	150	
BiLSTM		73%	0	0.75	0.92	0.83	350	[[323 27] [107 43]]
			1	0.61	0.29	0.39	150	
ELMo		79%	0	0.96	0.73	0.83	350	[[254 96] [11 139]]
			1	0.59	0.93	0.72	150	
SVM		80%	0	1	0.71	0.83	350	[[249 101]

			1	0.6	0.99	0.74	150	[1 149]]
SVM	SMOTE	80%	0	1	0.71	0.83	350	[[249 101] [1 149]]
			1	0.6	0.99	0.74	150	
SVM	OVERSAMPLER	80%	0	1	0.71	0.83	350	[[249 101] [1 149]]
			1	0.6	0.99	0.74	150	
SVM	ADASYN	80%	0	1	0.71	0.83	350	[[249 101] [1 149]]
			1	0.6	0.99	0.74	150	
LR		0.79	0	0.95	0.73	0.83	350	[[257 93] [13 137]]
			1	0.6	0.91	0.72	150	
LR	SMOTE	0.79	0	0.98	0.72	0.83	350	[[252 98] [5 145]]
			1	0.6	0.97	0.74	150	
LR	OVERSAMPLER	0.79	0	0.98	0.71	0.83	350	[[250 100] [6 144]]
			1	0.59	0.96	0.73	150	
LR	ADASYN	0.79	0	0.98	0.71	0.83	350	[[250 100] [6 144]]
			1	0.59	0.96	0.73	150	
DT		80%	0	1	0.71	0.83	350	[[249 101] [1 149]]
			1	0.6	0.99	0.74	150	
DT	SMOTE	80%	1	1	0.71	0.83	350	[[250 100] [6 144]]
			0	0.6	0.99	0.74	150	
DT	OVERSAMPLER	80%	0	1	0.71	0.83	350	[[250 100] [6 144]]
			1	0.6	0.99	0.74	150	
DT	ADASYN	80%	0	1	0.71	0.83	350	[[250 100] [6 144]]
			1	0.6	0.99	0.74	150	
KNN		0.76	0	0.86	0.78	0.82	350	[[272 78] [43 107]]
			1	0.58	0.71	0.64	150	
KNN	SMOTE	0.56	0	0.99	0.38	0.55	350	[[132 218] [1 149]]
			1	0.41	0.99	0.58	150	

KNN	OVERSAMPLER	0.77	0	0.98	0.69	0.81	350	[[240 110] [6 144]]
			1	0.57	0.96	0.71	150	
KNN	ADASYN	0.55	0	0.99	0.37	0.53	350	[[128 222] [1 149]]
			1	0.4	0.99	0.57	150	

Table 4.2 Results of Eye's dataset

The statistical outcomes of each model that was trained using Eye's dataset are shown in Table 4.2 above. **XLNet** performed better than all the models, with a high **accuracy** score of **90%.** From this table, it is evident that language models outperformed traditional ones in predicting True positive values. XLNet outperformed all models with a **Precision score of 85%, Recall score of 89%, and F1 score of 87%.** This model misclassified only 17 depressed tweets.

The performance of Decision Tree and SVM is greater than LR and KNN in terms of positive case prediction with scores of 60% Precision, 99% Recall, and 74% F1. KNN with ADASYN, was the model with the lowest accuracy of 55%. The Figure 4.4 shows the comparison of evaluation in terms of traditional models.

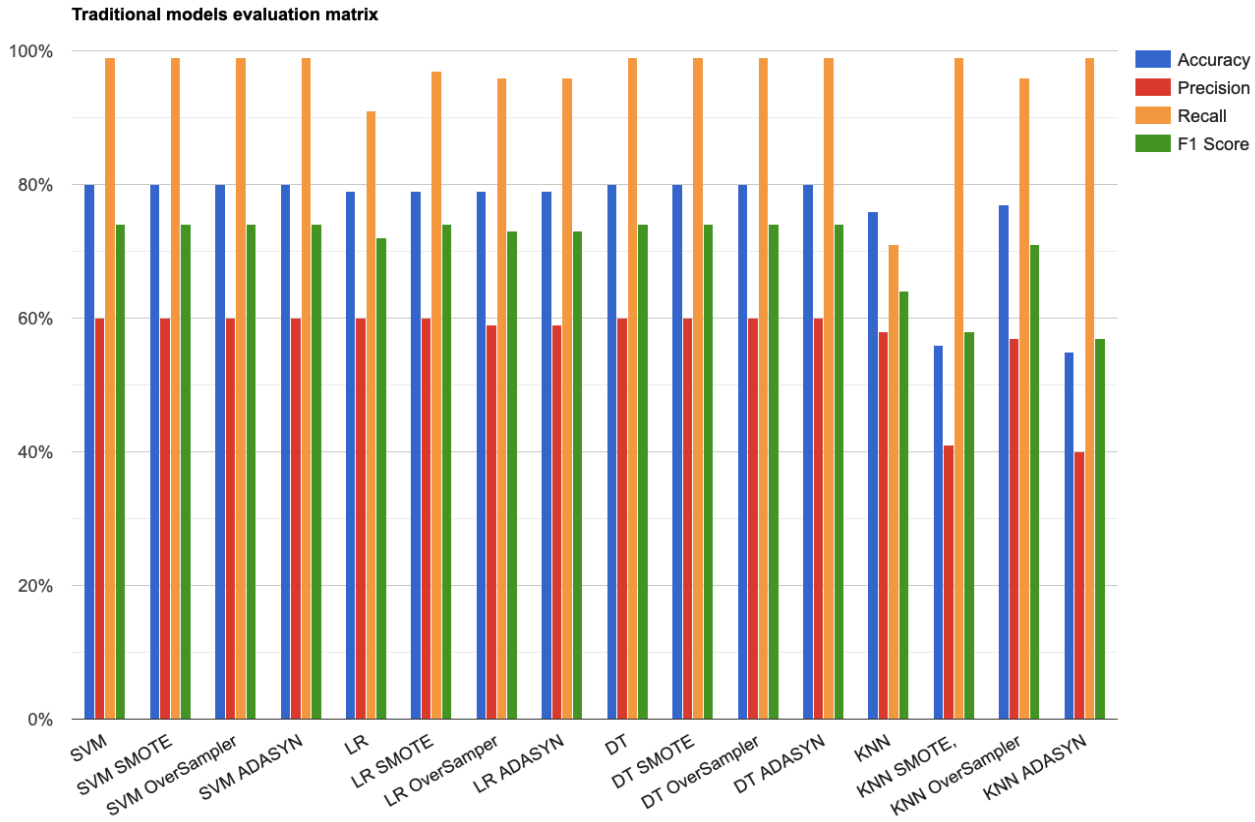


Figure 4.4. Evaluation matrix of traditional models with Eye's dataset

The comparison of language models reveals that XLNet predicted both negative and positive classes more accurately than Bert, ELMo, and BiLSTM, while BiLSTM was the model that did the worst, with an accuracy of 73%. It misclassified more than 75% of positive cases. More than 50% of depressed tweets were incorrectly labelled as non-depressed. The comparison between the language models for the positive class is shown in Figure 4.5.

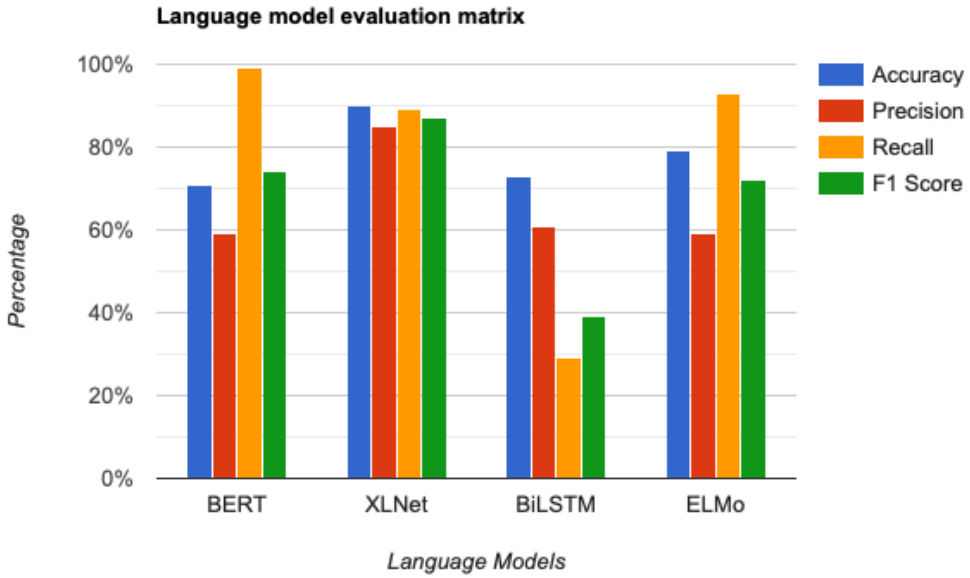


Figure 4.5: Evaluation matrix of language models with Eye's dataset

4.3 Misclassification analysis

In addition to find a best performing model, this study did a qualitative analysis of the projected tweets to enhance the depression classification. This is carried out to comprehend the causes of the incorrect classification of tweets. The test set was manually examined by contrasting the outcome of the accurate "manual" label with the "predicted" label from the model. The reasons found out on the anticipated tweets are given below.

- The tweets which discuss someone else's depression stage are classified as depressed due to the negative words present in it. Those tweets are not depressive, and it only represents someone else's mental stage. That does not mean the person who post the tweet is depressed. For e.g. "Her mother died in her childhood, and she started suffering from depression".
- The tweets about the depression support sessions are often misclassified as depressed due to the words present in it. For e.g. "The depression will worsen if you don't take the counselling sessions when needed. Contact this number if you need support".

- The tips about depression recovery are also misclassified as positive.
- Heavily depressed tweets such as “Depression loves me a lot” may be misclassified as negative due to the positive word like love.
- Any type of discussion about depression may be classified as depressed due to the presence of word depression in it. For e.g., ‘I’m shocked to find out that 300 million people suffer from depression which is so sadddd’.

4.4 Chapter Summary

This chapter summarises the performance evaluation of the models as well as the reason of misclassified tweets. The misclassification issues are identified such as the tweets about someone’s mental stage, depression related tips, general discussion on depression and the depression support agency’s posts. From the Performance evaluation, we understood that all models performed better on predicting negative cases than positive. As the aim of our study focuses on detecting depression symptoms which give weightage to predict positive classes. On this aspect, language models performed better than the traditional models. The prevalence of pre-trained models will be an asset to predict these symptoms correctly than traditional models.

The study showed that BERT and XLNet were the better models, with **XLNet** proved to be the best model given that it outperformed BERT by **90%** accuracy**.** Additionally, XLNet acquired very **high precision, recall and F1 score of 85%, 89%, 87%** respectively. In contrast, on the traditional models used, Decision Tree outperformed all other traditional models and showed comparatively good performance to predict positive cases. In conclusion, the language models performed better than the traditional models and **XLNet** is proved to be the best performing model, thereby chose for the EDDS framework.

CHAPTER 5: EVALUATION OF DELIVERABLE

5.0 Introduction

This chapter explains the validation of Early-stage Depression Detection System (EDDS) framework created for detecting depression in social media posts. EDDS was developed based on the sentiment analysis. This framework is created to help healthcare professionals to find people who have been suffering from depression by using their social media posts. Since depression is a mental illness, there is a high probability of people not knowing this condition. So, this framework intended to identify people with depressive symptoms, to offer needed support such as counselling, medication based on the condition. This reason makes this framework important to health sectors. The purpose of the validation procedure is to evaluate the dependability and applicability of EDDS framework to the healthcare professionals.

5.1 The concept of validation

Validity measures how well the research outcome can adapt to solve the real-world problem. As per McBurney & White, (2007), it is an integral part of framework development to evaluate at what extent the research findings are reliable and applicable. Golafshani (2003) mentioned that validation is a way to determine whether the research measures what it was intended to assess and how accurate the results are. To determine this, it is crucial to view the research findings through the researchers', stakeholders', and domain experts' viewpoints (Creswell & Miller, 2000). The method of validation used will determine how reliable and generalizable the research findings are. The research findings, which are discussed in the section below, were validated in this study using both internal and external validation.

5.2 Validation Approach

The aim of the study is to develop this EDDS framework so that healthcare professionals can use social media posts to identify people who are experiencing early-stage depression and prevent serious consequences like suicide. The evaluation of the study is done by medical professional through survey questionnaires. Due to the convenience, we chose convenience(non-profitability) sampling strategy to choose the relevant participants. It is a non-probability sampling technique to choose specific members of a target population who meet certain practical requirements, such as ease of accessibility, geographic proximity,

expediency, availability, or the willingness to participate (Edgar and Manz, 2017). The convenience sampling technique seems to be appropriate because it is easier to reach the target audience. Additionally, this method proved to be successful in situations where random sampling is impossible to perform (Ogunleye, 2021). A google form is used for the purpose of this survey which contains the participant consents and the survey questionnaires. The details of the survey form contain a detailed summary about the research such as nature of survey, aim, questions, objectives, and the proposed framework to avoid the issues caused by limited knowledge. Participants selected were medical professionals with minimum 2 years of experience from different hospital and even different countries, for getting a reasonable representation of the target audience. The online google survey form is emailed to the participants and asked relevant questions for validating EDDS framework.

5.3 Discussion of Validation Result

This section explains the details regarding the validation results obtained as part of the survey conducted on this research. Even though the sample used for validation is limited, the responses received were positive. The total of 11 participants who were Doctors, provided their consent to take part in survey by filling the questionnaire. The responses were analysed using the Google Form Response dashboard and Google Sheet generated from the same. The analysis found that there were 54.5% of males and 45.5% females who recorded their responses using the Google Form created for the survey. From the responses, it is evident that 90.9% of the users used social media more than once a day. It is interesting that this study included data from Twitter, which the respondents claimed was one of the most popular social networks of about 63.6%. All participants responded that they use social media for connecting with family and friends. On the top of it, all agreed the relevance of this framework for depression detection. The Figure 5.1 shows the frequency, preferred social media platform and purpose of social media respectively.

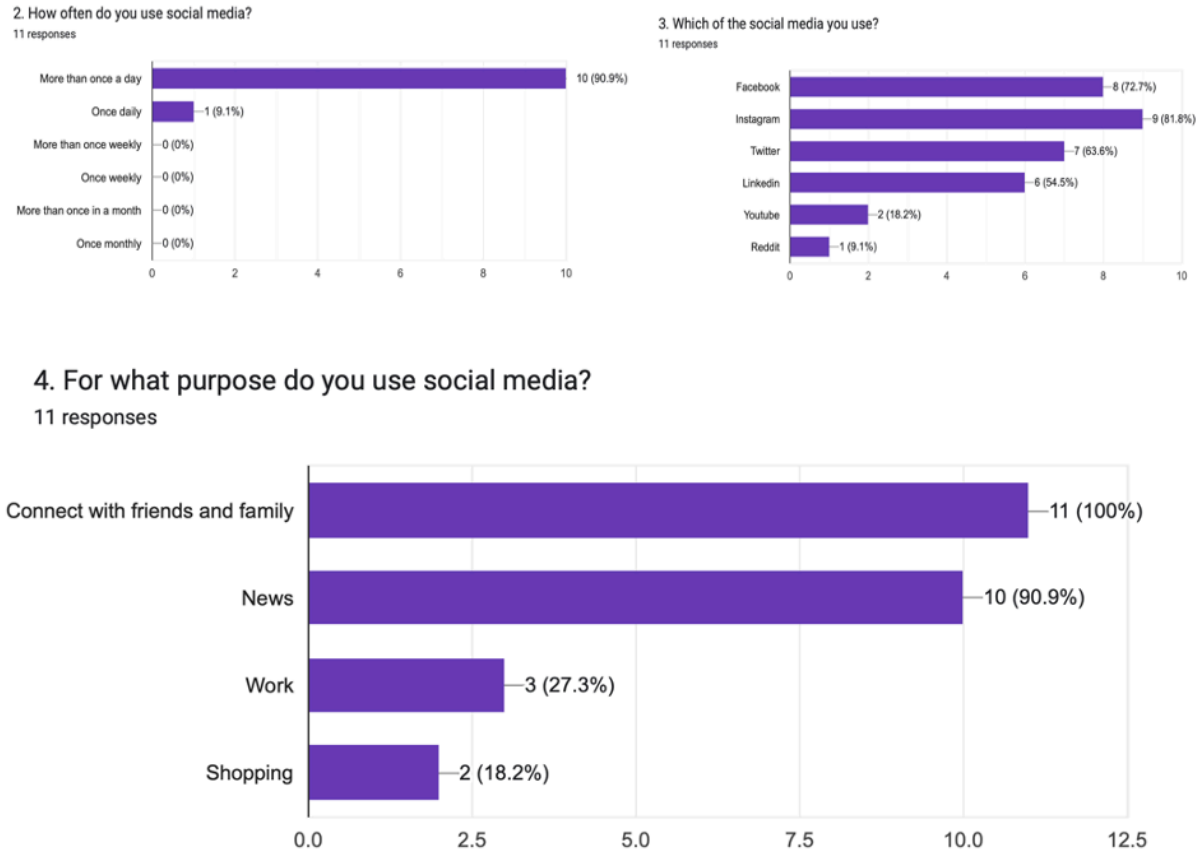


Figure 5.1. Illustration of social media usage

From these figures it is evident that social media became a necessary part of people's life that can be used to communicate and express their emotion and opinion.

To what extend do you agree with this?	People are not getting enough support and the increase in suicide make an urgency of a system to identify depression in its early stage.	Depression takes hold gradually, without a person realising he/she is depressed. It will result in a significant health deterioration and even the possibility of suicide is high	Conventional methods are not helpful to identify people with depression symptoms due to unawareness or self-denial.	People are expressing themselves in social media than interacting with people. Hence, social media can be used as a fruitful source to understand their state of mind.	People freely express their emotions and sentiments in social media platforms. These data can be efficient utilised to detect depression in its early stage.	Sentiment analysis system with machine learning models helps to classify people as depressed and non-depressed without any hassle, just by using their social media data.
Strongly agree	81.8%	63.6%	63.6%	63.6%	72.7%	63.6%
Agree	18.2%	36.4%	27.3%	27.3%	18.2%	27.3%
Neutral	0%	0%	9.1%	0%	9.1%	9.1%
Disagree	0%	0%	0%	9.1%	0%	0%
Strongly disagree	0%	0%	0%	0%	0%	0%

Table 5.1 EDDS Validation result-Part1

The table 5.1 shows the relevance of the EDDS due to the current increase in number of depressive people. Firstly, the survey focuses on the response on the prevalence of depression, inefficiency of conventional methods and the urgency of the system. The table shows that 81.8% people strongly agree to that and 18.2% agree that. Secondly, the questionnaire focuses on the analysis of selecting social media as a fruitful source of the research. 90.9% people believe that people's posts can be used to analyse the mental stage. Thirdly 90.9% strongly believe that the conventional methods are inefficient to identify the early-stage detection. This validation shows that the doctors believe this system is relevant and can be used as a better tool to identify depression with the use of social media data.

To what extend do you agree with this?	The use of social media is increasing day by day. An automatic system which can classify their mental stage as depressed or non-depressed is considered beneficial to identify whether they need to seek support or not.	An early depression detection system is a COMPULSORY system for health sectors to provide necessary support to the people who suffer from depression.
Yes	90.9%	90.9%
No	0%	0%
Maybe	9.1%	9.1%

Table 5.2 EDDS Validation result-Part2

From table 5.2., the results shows that most of the participants, precisely 90.9% agrees that EDDS system is necessary and compulsory for the health sector to identify depressive people. In conclusion, most participants agrees that this research findings and the EDDS framework is relevant, pertinent, and applicable to the health sector for identifying depressive people to offer support.

CHAPTER 6: CONCLUSION

6.0 Conclusion

The most common mental illness, depression, is the primary factor in more than two-thirds of suicides each year(WHO, 2021). Unfortunately, a lot of cases go untreated due to self-denial or a failure to recognise them. Numerous research concurred that social media posts can be a useful tool for tracking a variety of mental health conditions, including depression, given the exponential rise in social media usage. This inspired to conduct a study on detecting depression in its early stage by using the latest natural language learning techniques with twitter data. The goal of this study is to discover the optimal framework for the health sector to identify the symptoms quickly and simply through text processing by doing a comparative analysis of currently utilised traditional models and the advanced language models.

The basis for detecting depression in an individual is their unrestrained sentiment, which they openly express on social media because so many people utilise it to portray a true picture of their private lives. The lack of research employing language models provides a window for further investigation. Additionally, the previous studies did not analyse and provide explanations for classification errors. These gaps led to the following study and answer the following research question.

How can Twitter sentiment analysis help detect early-stage depression?

6.1 Evaluation of Research Objectives & Questions

In this section, the objectives of the study will be reviewed to assess if the research question has been answered. The sections 5.1.1 to 5.1.3 discuss the evaluation of objectives 1 to 3.

6.1.1 Objective 1

To conduct a literature review to get a good background of the problem and to understand the methodologies normally used for this problem.

The literature review serves as a solid foundation for the study and aids in identifying the important contexts for sentiment analysis during the text classification process. The datasets (Chiong et al., 2021), methods, technologies, and traditional methods were selected based on

the previous studies. To summarise, this literature not only helps to identify the gaps in the research area but also helps to progress in a proper direction of the research.

6.1.2 Objective 2

To propose a comparative study of sentiment analysis to detect depression with language learning models as well as machine learning models using Twitter data.

This objective mainly helps to answer the literature review by conducting a comparative study between traditional and language models. By this implementation, we proved that the Twitter sentiment analysis using these models can detect depression.

6.1.3 Objective 3

To find the best performing models suitable to detect detection by evaluating the model's performance.

The best performing models were identified by evaluating the model performance using confusion matrix parameters such as Accuracy, Precision, F1-score, and Recall. The details of the confusion matrix parameters are added in section 4.2. The comparative result presented showed that the XLNet was the best performing classifier with 90% accuracy and thus it is selected as the best model for the depression detection prediction.

6.2 Contributions

This study contributes to existing knowledge in many ways. The significance of the findings and contributions of this study can be summarised as follows.

- The word cloud generated during EDA helps to identify the words majorly expressed by the depressed people.
- Most of the studies simply paid attention to confirming their research using a test set made from the same dataset. This may result in overfitting problems. Since the observations in validation dataset are manually annotated, that helps to reduce suspicion and expect more reliable outcomes.
- This comparative study of traditional vs language models will be a steppingstone to new studies.
- Proposed EDDS framework should be helpful to detect depression in its early stage.

6.3 Limitation & Future work

This study investigated the efficiency of detecting depression in its early stage using sentimental analysis and proposed a framework which can identify depression. However, the annotated datasets used in this study are 2 open datasets created by Eye (2020) and Shen et. al. (2017). In Eye's dataset the depressed classes are assigned by verifying 'diagnose' or 'depressed' words whereas Shen et al's dataset the depressed tweets are generated from the people who were diagnosed with depression. So, there is a high chance of bias in the datasets. Less strictly constructed datasets can be more beneficial especially when the models would be used for detecting depression in a real-life scenario. This will lead to a future study with a manually annotated/less strictly constructed datasets. More models including hybrid models can be developed and verified as part of future study. Finally, it should be noted that the approach presented in this study utilized supervised Machine Learning classifiers, and therefore, it is limited to use labelled datasets for training the models. Future study may focus on overcoming this limitation by including unsupervised classifiers.

The future study identified as follows.

- Study on datasets which are prepared less strictly and manually annotated by experts.
- Comparative study with more models and including unsupervised models also.
- Study on Hybrid models.

REFERENCE

AlSagri, H. S., & Ykhlef, M. (2020). Machine learning-based approach for depression detection in twitter using content and activity features. *IEICE Transactions on Information and Systems, 103*(8), 1825-1832.

Amaturo, E., & Punziano, G. (2017). Blurry Boundaries: Internet, Big-New Data, and Mixed-Method Approach. In Data Science and Social Research (pp. 35-55). Springer, Cham.

Angskun, J., Tipprasert, S., & Angskun, T. (2022). Big data analytics on social networks for real-time depression detection. Journal of Big Data, 9(1), 69. 10.1186/s40537-022-00622-2

Arslan, Y., Allix, K., Veiber, L., Lothritz, C., Bissyandé, T. F., Klein, J., & Goujon, A. (2021). A comparison of pre-trained language models for multi-class text classification in the financial domain. Paper presented at the Companion Proceedings of the Web Conference 2021, 260-268.

Bi, Y., Li, B., & Wang, H. (2021). Detecting Depression on Sina Microblog Using Depressing Domain Lexicon. Paper presented at the 2021 IEEE Intl Conf on Dependable, Autonomic and Secure Computing, Intl Conf on Pervasive Intelligence and Computing, Intl Conf on Cloud and Big Data Computing, Intl Conf on Cyber Science and Technology Congress, 965-970.

Biradar, A., & Totad, S. G. (2018). Detecting depression in social media posts using machine learning. Paper presented at the International Conference on Recent Trends in Image Processing and Pattern Recognition, 716-725.

Budhi, G. S., Chiong, R., & Wang, Z. (2021). Resampling imbalanced data to detect fake reviews using machine learning classifiers and textual-based features. Multimedia Tools and Applications, 80(9), 13079-13097.

Burdisso, S. G., Errecalde, M., & Montes-y-Gómez, M. (2019). A text classification framework for simple and effective early depression detection over social media streams. Expert Systems with Applications, 133, 182-197.

Cacheda, F., Fernandez, D., Novoa, F. J., & Carneiro, V. (2019). Early detection of depression: social network analysis and random forest techniques. Journal of Medical Internet Research, 21(6), e12554.

Cha, J., Kim, S., & Park, E. (2022). A lexicon-based approach to examine depression detection in social media: the case of Twitter and university community. Humanities and Social Sciences Communications, 9(1), 1-10.

Chiong, R., Budhi, G. S., Dhakal, S., & Chiong, F. (2021). A textual-based feature approach for depression detection using machine learning classifiers and social media texts. Computers in Biology and Medicine, 135, 104499.

Creswell, J. W., & Miller, D. L. (2000). Determining validity in qualitative inquiry. Theory into Practice, 39(3), 124-130.

Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: synthetic minority over-sampling technique. Journal of artificial intelligence research, 16, 321-357.

De Villiers, M. R. (2005). Three approaches as pillars for interpretive information systems research: development research, action research and grounded theory. Paper presented at the Proceedings of the 2005 Annual Research Conference of the South African Institute of Computer Scientists and Information Technologists on IT Research in Developing Countries, 142-151.
Devlin, J., Chang, M., Lee, K., & Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv Preprint arXiv:1810.04805,

Dey, S., Wasif, S., Tonmoy, D. S., Sultana, S., Sarkar, J., & Dey, M. (2020). A comparative study of support vector machine and Naive Bayes classifier for sentiment analysis on Amazon product reviews. In 2020 International Conference on Contemporary Computing and Applications (IC3A) (pp. 217-220). IEEE.

Eye, B. B. (2020). Depression Analysis.
Edgar, T., & Manz, D. (2017). Research methods for cyber security Syngress.

Fithriasari, K., Jannah, S. Z., & Reyhana, Z. (2020). Deep learning for social media sentiment analysis. MATEMATIKA: Malaysian Journal of Industrial and Applied Mathematics, 99-111.

Haibo He, Yang Bai, E. A. Garcia, & Shutao Li. (2008). ADASYN: Adaptive synthetic sampling approach for imbalanced learning. Paper presented at the - 2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence), 1322-1328. 10.1109/IJCNN.2008.4633969

He, H., & Garcia, E. A. (2009). Learning from imbalanced data. IEEE Transactions on knowledge and data engineering, 21(9), 1263-1284.

He, H., Bai, Y., Garcia, E. A., & Li, S. (2008). ADASYN: Adaptive synthetic sampling approach for imbalanced learning. In 2008 IEEE international joint conference on neural networks (IEEE world congress on computational intelligence) (pp. 1322-1328). IEEE.

Hesse-Biber, S., & Johnson, R. B. (2013). Coming at things differently: Future directions of possible engagement with mixed methods research.

Islam, M. R., Kamal, A. R. M., Sultana, N., Islam, R., & Moni, M. A. (2018). Detecting depression using k-nearest neighbors (knn) classification technique. Paper presented at the 2018 International Conference on Computer, Communication, Chemical, Material and Electronic Engineering (IC4ME2), 1-4.

Islam, M., Kabir, M. A., Ahmed, A., Kamal, A. R. M., Wang, H., & Ulhaq, A. (2018). Depression detection from social network data using machine learning techniques. Health Information Science and Systems, 6(1), 1-12.

Joshi, M. L., & Kanoongo, N. (2022). Depression detection using emotional artificial intelligence and machine learning: A closer review. Materials Today: Proceedings, 58, 217-226. https://doi.org/10.1016/j.matpr.2022.01.467

Jain, A. P., & Katkar, V. D. (2015). Sentiment analysis of Twitter data using data mining. In 2015 International Conference on Information Processing (ICIP) (pp. 807-810). IEEE.

Kim, J., Lee, J., Park, E., & Han, J. (2020). A deep learning model for detecting mental illness from user content on social media. Scientific Reports, 10(1), 1-6.

Kim, Y. (2014). Convolutional neural networks for sentence classification. arXiv preprint arXiv:1408.5882.

Li, G., Li, B., Huang, L., & Hou, S. (2020). Automatic construction of a depression-domain lexicon based on microblogs: text mining study. JMIR Medical Informatics, 8(6), e17650.

Lim, Y. Q., Lim, C. M., Gan, K. H., & Samsudin, N. H. (2020). Text sentiment analysis on Twitter to identify positive or negative context in addressing inept regulations on social media platform. Paper presented at the 2020 IEEE 10th Symposium on Computer Applications & Industrial Electronics (ISCAIE), 96-101.

Pachouly, S. J., Raut, G., Bute, K., Tambe, R., Bhavsar, S., & Students, U. (2021). Depression Detection on Social Media Network (Twitter) using Sentiment Analysis. Int.Res.J.Eng.Technol., 8, 1834-1839.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., & Vanderplas, J. (2011). Scikit-learn: Machine learning in Python. the Journal of machine Learning research, 12, 2825-2830.

Pirina, I., & Çöltekin, Ç. (2018). Identifying depression on reddit: The effect of training data. Paper presented at the Proceedings of the 2018 EMNLP Workshop SMM4H: The 3rd Social Media Mining for Health Applications Workshop & Shared Task, 9-12.

Ogunleye, B. O. (2021). Statistical learning approaches to sentiment analysis in the nigerian banking context Available from ProQuest Dissertations & Theses Global: The Humanities and Social Sciences Collection. Retrieved from https://search.proquest.com/docview/2734703680

Queirós, A., Faria, D., & Almeida, F. (2017). Strengths and limitations of qualitative and quantitative research methods. European Journal of Education Studies.

Rajaraman, P. V., Nath, A., Akshaya, P. R., & Bhuja, G. C. (2020). Depression detection of tweets and A comparative test. International Journal of Engineering Research, 9(03), 422-425.

Rehurek, R., & Sojka, P. (2011). Gensim—statistical semantics in python. Retrieved from genism. org.

Robinson, D. (2017). The incredible growth of Python. Stack Overflow—Sep, 6.

Ruder, S., Ghaffari, P., & Breslin, J. G. (2016). A hierarchical model of reviews for aspect-based sentiment analysis. arXiv preprint arXiv:1609.02745.

Saha, A., Al Marouf, A., & Hossain, R. (2021). Sentiment analysis from depression-related user-generated contents from social media. Paper presented at the 2021 8th International Conference on Computer and Communication Engineering (ICCCE), 259-264.

Salton, G., Wong, A., & Yang, C. S. (1975). A vector space model for automatic indexing. Communications of the ACM, 18(11), 613-620.

Saunders, M. N. K. (2015). Research methods for business students. Pearson Education.

Shah, F. M., Ahmed, F., Joy, S. K. S., Ahmed, S., Sadek, S., Shil, R., & Kabir, M. H. (2020). Early depression detection from social networks using deep learning techniques. Paper presented at the 2020 IEEE Region 10 Symposium (TENSYMP), 823-826.

Shen, G., Jia, J., Nie, L., Feng, F., Zhang, C., Hu, T., Chua, T., & Zhu, W. (2017). Depression detection via harvesting social media: A multimodal dictionary learning solution. Paper presented at the Ijcai, 3838-3844.

Shetty, N. P., Muniyal, B., Anand, A., Kumar, S., & Prabhu, S. (2020). Predicting depression using deep learning and ensemble algorithms on raw twitter data. International Journal of Electrical and Computer Engineering, 10(4), 3751.

Statista. (2022). Number of Twitter users worldwide from 2019 to 2024 Retrieved Dec 31 2022, from https://www.statista.com/statistics/303681/twitter-users-worldwide/

Stephen, J. J., & Prabu, P. (2019). Detecting the magnitude of depression in Twitter users using sentiment analysis. International Journal of Electrical and Computer Engineering, 9(4), 3247.

Tadesse, M. M., Lin, H., Xu, B., & Yang, L. (2019). Detection of Depression-Related Posts in Reddit Social Media Forum. IEEE Access, 7, 44883-44893. 10.1109/ACCESS.2019.2909180

Trotzek, M., Koitka, S., & Friedrich, C. M. (2018). Utilizing neural networks and linguistic metadata for early detection of depression indications in text sequences. IEEE Transactions on Knowledge and Data Engineering, 32(3), 588-601.

Twitter usage statistics. Twitter Usage Statistics - Internet Live Stats. (n.d.). Retrieved December 12, 2022, from https://www.internetlivestats.com/twitter-statistics/

Wang, R., Li, Z., Cao, J., & Chen, T. (2019). Chinese Text Feature Extraction and Classification Based on Deep Learning. In Proceedings of the 3rd Inter

Wedderburn, R. W. (1974). Quasi-likelihood functions, generalized linear models, and the Gauss—Newton method. Biometrika, 61(3), 439-447.

WHO. (2021). Depression https://www.who.int/news-room/fact-sheets/detail/depression

WHO. (2022). COVID-19 pandemic triggers 25% increase in prevalence of anxiety and depression worldwide https://www.who.int/news/item/02-03-2022-covid-19-pandemic-triggers-25-increase-in-prevalence-of-anxiety-and-depression-worldwide

Woo, S. E., O'Boyle, E. H., & Spector, P. E. (2017). Best practices in developing, conducting, and evaluating inductive research.

Xu, G., Meng, Y., Qiu, X., Yu, Z., & Wu, X. (2019). Sentiment analysis of comment texts based on BiLSTM. Ieee Access, 7, 51522-51532.

Yang, Z., Dai, Z., Yang, Y., Carbonell, J., Salakhutdinov, R. R., & Le, Q. V. (2019). Xlnet: Generalized autoregressive pre training for language understanding. Advances in Neural Information Processing Systems, 32

Appendix A – Research Project Plan

Task 2: Research Project Plan: Depression Detection using Sentimental Analysis

Devi Priya M

c1056531

1. INTRODUCTION AND JUSTIFICATION

Depression is a mental illness that affects the quality of life which can lead to suicide. The major problem of the illness is that the chance of not knowing about the symptoms are very high. As per WHO, 280 million people world-wide suffer with this illness. The illness got worse due to the pandemic situation and a lot of people committed suicide which makes the importance to detect it in the early stages (Stephen & Prabu, 2019). It can be quite challenging for health care organisations to identify depression in the early stages because people are reluctant to seek care when necessary or uninformed of their mental state. In this era, people are very comfortable expressing their emotions in social media as posts, messages, and comments rather than speaking or sharing their emotions to other individuals (Burdisso et al., 2019). Social networking platforms represent a person's personal life (Stephen & Prabu, 2019) which can be used for identifying the symptoms of mental illness such as depression by using natural language processing (NLP) with machine learning(ML) techniques.

2. RESEARCH QUESTION, AIMS & OBJECTIVES

The project's goal is to provide a framework to help healthcare institutions identify depression early on, reducing the likelihood that the illness would worsen because of ignorance. The development of a framework including sentiment analysis and machine learning models makes it simpler for health organisations to recognise people who are experiencing depression in its early stages and offer them with appropriate treatment by utilising the social networking data.

2.1 Research question

How can Twitter sentiment analysis help detect early-stage depression?

2.2 Objectives

1. Conduct research on social media sentiment analysis for depression detection.
2. Identify relevant project management methodology for the framework development.
3. Decide suitable datasets for the machine models.
4. Decide suitable pre-processing methods.
5. Decide suitable ML and NLP techniques.
6. Develop a framework to detect depression using the twitter dataset.
7. Statistical model evaluation using confusion matrix parameters.
8. Result evaluation from the health sector and decide the future study.

2.3 Deliverable

A framework that can use an aid of detecting the initial symptoms of depression.

3 LITERATURE REVIEW

3.1 Why do sentiment analysis for depression detection?

Diagnoses and subsequent treatment for depression are often delayed or missed entirely because of the late detection (Joshi & Kanoongo in 2022). Stephen & Prabu, (2019) and Joshi & Kanoongo (2022) argue that the sentiments expressed in the tweets became an efficient method of understanding the deeper emotions of the users. Thereby, the sentimental analysis of social media is an efficient method to detect the symptoms of depression in an early stage.

3.2 Machine Learning Models

3.2.1 Traditional Machine Learning Models

In 2022, Joshi & Kanoongo have conducted a comparative study on Twitter sentiment analysis. The dataset contains 43,000 tweets which divides training and test data in a 70:30 ratio. In addition to emoji extraction, stop-word removal, spelling correction and lemmatization, the author employed lengthy and efficient pre-processing techniques. Further, the model is trained using a bag of words models to identify the frequency of the term on the text for the predictive model. They believe that that will improve the accuracy of prediction. The 30 % of test data is pre-processed to use to train the model by adding the tweet positive or negative column. The performance of the model is evaluated using a confusion matrix and they analysed that the Multinomial Naive Bayes works better than Support Vector Machine (SVM) on their analysis. Although the study appears to be extremely effective, it does not provide the model performance time, which is essential for an early stage detection.

In order to identify depression in tweets, Rajaraman et al. did a study in 2020 using TF-IDF predictions. The research makes use of the Python libraries NumPy, Matplot, Scikit Learn, Natural Language Toolkit(NLTK), WordCloud, and Kera. By categorising the tweets as normal or depressive, the pre-processed Twitter data is used to train the system and conduct testing. The long short-term memory networks (LSTM) Recurrent neural network(RNN), a deep learning classifier, is discovered to be the most accurate classifier after a comparison analysis of several techniques including TF-IDF, Naive-bayers, LSTM, Logistic Regression, and Linear Support Vectors. Using the parameters of the confusion matrix, the statistical verification is carried out. Even though this study appears to be thorough, it is still unclear how quickly the models can identify depression.

In the same year, AlSagri & Ykhlef have done similar studies of various ML-base approaches using tweets. The researchers argue that the accuracy and F1 score increase as the number of features increases. They employed SVM, Naive-Bayers, and Decision Tree classification models, and due to its effectiveness, they recommended SVM-linear classifiers. After pre-processing, features are extracted from the text. Self-Center, TF-IDF, Feature Selector, Sentiment, Use-words, and Synonyms are the characteristics that are utilised. The researchers succeed to make their model as an outperformed one due to the diversity and richness of its feature set. Even if they insist that the selected candidates were manually labelled by two distinct psychologists in unanimity, the special selection gave us the impression that the dataset was highly biased. Furthermore, they have not yet offered any proof of sampling methods used to find and correct biases.

In 2018, Islam, Md et al and Islam, Md Rafiqul et al conducted studies to identify depression using Facebook comments with different models. Despite the fact that this study was conducted on Facebook, ML models and sentiment analysis approaches were nonetheless useful for our investigation. While Islam, Md, et al worked on k-Nearest Neighbours (KNN) classification Technique, the same researcher and Islam, Md Rafiqul et al focused on 4 other machine learning models. The data collected is divided into 2, depressive and non-depressive. Both the studies used Linguistic Inquiry and Word Count (LIWC) to investigate emotional (5 emotional:- positive, negative, sad, anger, anxiety), temporal(3-present, past, future) and linguistic style factors (e.g., articles, prepositions, pronouns, verbs and negations)to train the model each one alone and together. The second study additionally used 10-fold cross validation to raise the effectiveness of the models. Performance was calculated using evaluation parameters. The authors concur that additional research in that area is possible because the study lacked the methods to extract passphrases from a wider variety of emotional traits.

3.2.2. Deep Learning Models

A system that uses Twitter as a source of data and predicts depression using the Back Propagation Neural Network (BPNN) model was developed in 2018 by Biradar & Totad. The sentimental analysis is done using Senti Strength which has a sentimental value and is used as the training data for the BPNN model. Additionally, the data is gathered using the Twitter API, and each tweet is assigned a sentiment score using BPNN, which is then utilised to train the BPNN model. Even

though this model seems to be a single, centric predictive model which uses twitter activities and predicts them into depressed or not depressed, the literature lacks the performance evaluation methods.

Using Twitter datasets, Shetty et al., 2020 studied ML classifiers to recognise depression. Sentiment analysis and ML classifiers are the two parts of the analysis. Twitter postings from specific individuals are subjected to sentiment analysis to predict binary classes, such as depressed/not depressed. In the initial stage, classifiers from LSTM and Convolutional Neural Networks (CNN) are utilised. Classifiers like support vector classifiers, multinomial naive-bayes classifiers, Bernoulli naive-bayes classifiers, logistic regression classifiers, random forest classifiers, and ensembles are used in the second stage. The result is then compared to the first stage prediction using the weighted mean after being vectorized using the count vectorizer, TF-IDF, and n-grams. These weights are assigned based on the model's accuracy with the data provided. In the same year, Rajaraman et conducted a study using the LSTM RNN, a deep learning classifier and found that it is the highest accurate classifier than other traditional classifiers he used. Both researchers argue that the deep learning models are having high accuracy compared to traditional methods. The limitation of these two researchers is that the studies did not provide any sampling techniques to analyse the sampling issue.

4. RESEARCH DESIGN

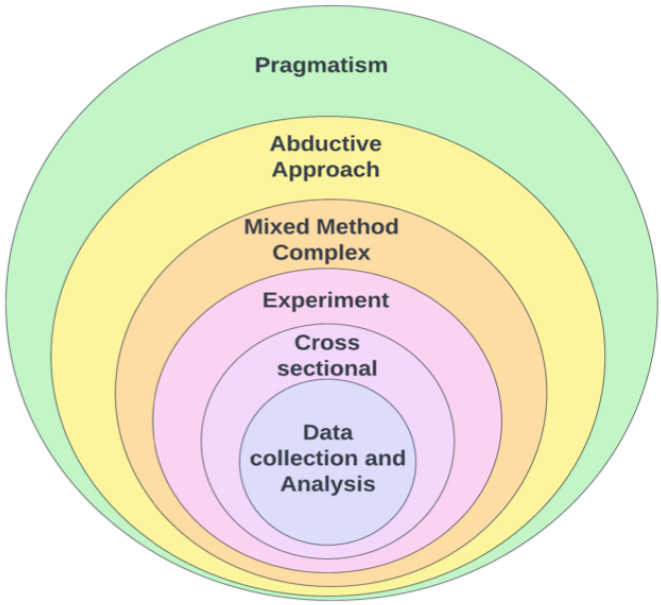
4.1 Research Philosophy, Approach & Methodology

To accomplish the goals and address the research challenge, the machine learning model implementation needs to adhere to the pragmatic philosophy. Pragmatic philosophy is more appropriate for this framework development because it uses a variety of methodologies, quantitative data, and action based on useful ideas and results (Saunders, 2015). For this study, the abductive method appears to be more accurate for the following reasons. First, testable results are produced using well-established technology like machine learning models. The same data gathering is also utilised to find the conceptual framework and test this framework by understanding trends and patterns. Finally, the earlier research will serve as a foundation for the development or modification of new frameworks. The abductive model research approach aids in planning the study considering the restrictions and limitations noted in the earlier approach (Saunders, 2015).

The methodology of choice for this quantitative study will be the Mixed-method complex design. By analysing massive, categorised tweets for sentiment analysis, the study uses quantitative analysis techniques to analyse qualitative data, which complicates the mixed-method design paradigm. Because the dataset is used to statistically examine the frequency of occurrences of various patterns of data, the data gathering relies on indirect observation. The same dataset can split to train and test the models. The experiment research strategy seems to be relevant for this project to create a systematic approach of procedure to test the specific theory. The literature review provides a good background to the system to identify and manipulate the set of variables to achieve a specific result.

It is very crucial to test the performance of the model developed using a certain set of specific techniques such as confusion matrix. The specific timeline of the project is to console the time horizon into cross-sectional and to complete within the timeline.

Figure1 : Research Onion for the proposed project



4.2 Techniques and Procedures

4.2.1 Data Collection:

Twitter is one of the most popular social networking sites because of its unique features. It is recognised as a basic microblogging platform with user interfaces that allow the posting of 140-character maximum short stories (K. A. Govindasamy & N. Palanichamy, 2021). Moreover, major languages such as R and python have support to fetch the data by downloading tweets. For this research, an open source dataset is taken from kaggle and the main columns of the dataset used are unique_id, date & time, user_id and tweets.

The source and the screenshot of the sample data is given below;

Link: Twitter_data

Figure 2 : Sample data

1467810369	Mon Apr 06 22:19:45 PDT	NO_QUERY	_TheSpecialOne_	@switchfoot http://twitpi
1467810672	Mon Apr 06 22:19:49 PDT	NO_QUERY	scotthamilton	is upset that he can't upd
1467810917	Mon Apr 06 22:19:53 PDT	NO_QUERY	mattycus	@Kenichan I dived many t
1467811184	Mon Apr 06 22:19:57 PDT	NO_QUERY	ElleCTF	my whole body feels itchy
1467811193	Mon Apr 06 22:19:57 PDT	NO_QUERY	Karoli	@nationwideclass no, it's
1467811372	Mon Apr 06 22:20:00 PDT	NO_QUERY	joy_wolf	@Kwesidei not the whole
1467811592	Mon Apr 06 22:20:03 PDT	NO_QUERY	mybirch	Need a hug
1467811594	Mon Apr 06 22:20:03 PDT	NO_QUERY	coZZ	@LOLTrish hey long time
1467811795	Mon Apr 06 22:20:05 PDT	NO_QUERY	2Hood4Hollywood	@Tatiana_K nope they dic
1467812025	Mon Apr 06 22:20:09 PDT	NO_QUERY	mimismo	@twittera que me muera
1467812416	Mon Apr 06 22:20:16 PDT	NO_QUERY	erinx3leannexo	spring break in plain city...
1467812579	Mon Apr 06 22:20:17 PDT	NO_QUERY	pardonlauren	I just re-pierced my ears

4.2.2 Tools and Techniques

A quantitative study is planned to develop a framework with multiple classifiers to identify early-stage depression using tweets. The selection of a programming language done based on the developer's level of proficiency with it, is a key component in many machine learning systems. Most used are Python, R or Java. For this implementation, Python language is chosen because of the familiarity and wide community support. Moreover, python has a significant role in data analytics, machine learning, data science, data engineering, and even artificial intelligence due to its library support, especially NLP, Scikit learn etc. This research makes use of the Python libraries NumPy, Matplot, Scikit Learn, NLTK, WordCloud, and Kera.

The proposed system design is given below.

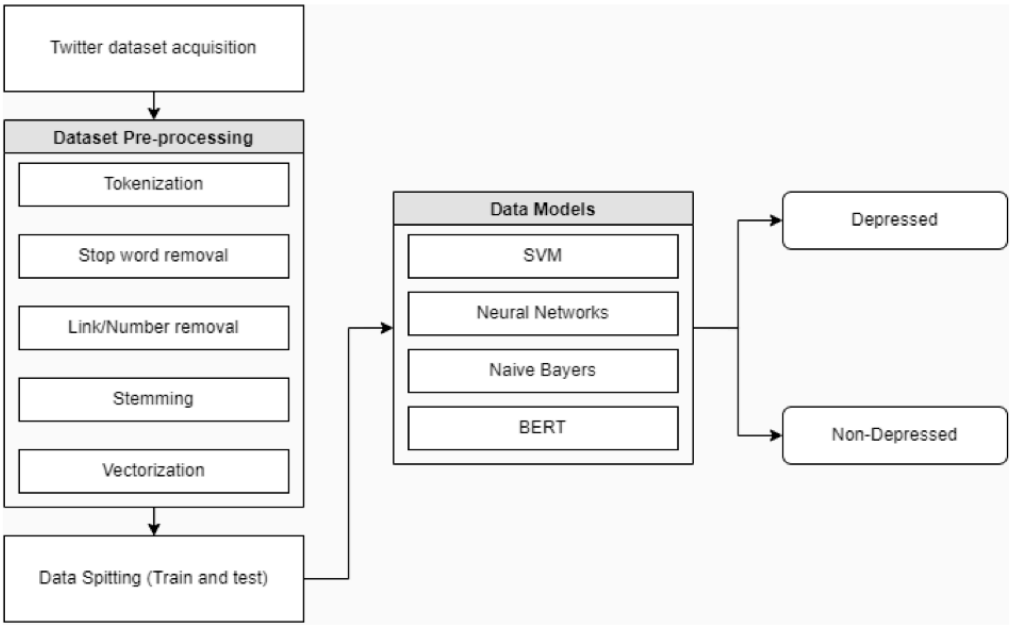


Figure 3: System Design

In the pre-processing stage, the emoji extraction, timestamp/digits/symbols/pronounce removal, spelling correction, lemmatization, and stop word removal are planned to be conducted. A thorough data exploration can be performed to analyse the initial trends and patterns of the large dataset. The

models will then be trained using a bag of words models to identify the frequency of the term on the text for the predictive model.

Based on the literature reviews, the Sentimental Analysis techniques are planned to be done on the pre-processed data using LIWC. In 2018, Islam, Md et al used LIWC to investigate emotional, temporal, and linguistic style factors to train the model each one and together. They have used 27 libraries and still 23 libraries are there to explore and we can analyse and implement required ones from those set in the implementation phase.

Based on the literature reviews, for the framework implementation SVM, Naive Bayers, Neural Networks, BERT ML models are used due to their high performance and accuracy. The performance of the models is evaluated using the Confusion matrix. The further tuning methods for improving the accuracy can be decided during the project development stage.

4.2.3 Project Management Methodology

The agile and waterfall models are the most used project management models. In the waterfall model, the development process is in a linear sequential flow, while in the agile model, it is in iterative form. The proposed research is time bound and has been carried out under the supervision. So the Agile-Waterfall hybrid model can be more appropriate for this.

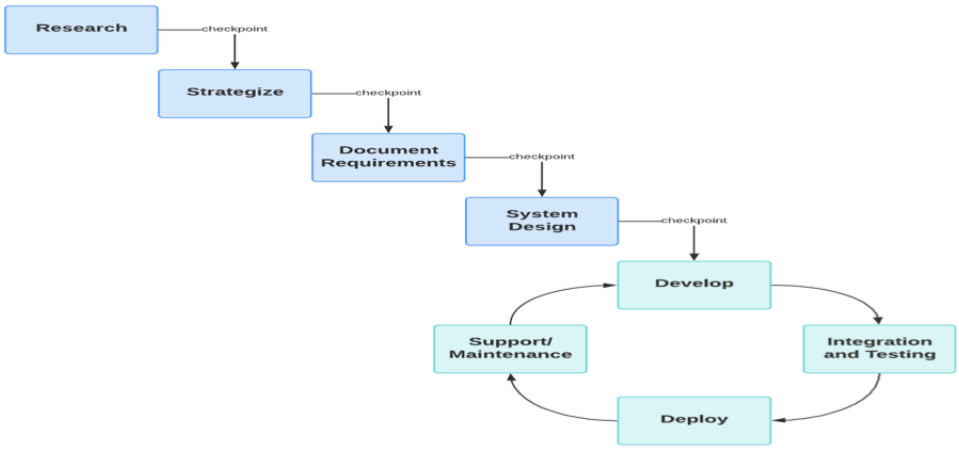


Figure 4: Agile-Waterfall Hybrid

As defined by Erick Bergmann and Andy Hamilton, the Agile-Waterfall hybrid typically allows the developing software to work within the Agile methodology which needs continuous integration and testing, while the tasks should be completed in a specific timeline ("Agile-Waterfall Hybrid: Is It Right for Your Team? | Lucidchart Blog", 2022).

5. ETHICS, RISKS AND ISSUES

5.1 ETHICS

Ethics in research refers to the moral ideals that guide research.

- **Beneficence & Non-Malfeasance** - This research is very useful for the health sector to identify and offer needed support to improve mental health.
- **Integrity -** This research will be conducted based on proven methods for sentiment analysis.
- **Impartiality:** This research will be conducted with honest and knowledgeable intention.

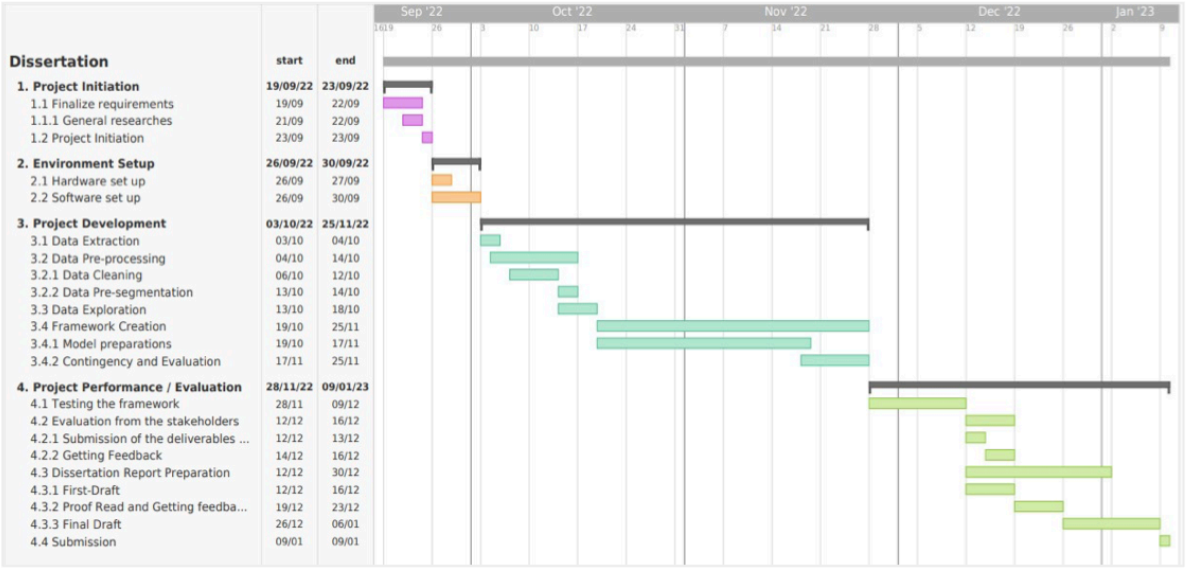
The Ethics Checklist form is attached to this document in the Appendix A section.

5.2 RISKS AND POTENTIAL ISSUES

The below table represents the potential risks that may be encountered during the development stage.

Table 1 : Risk table

ID	Risk description	Probability	Impact	Risk category	Response and Mitigation Plan
1	Resource risk due to health issues	Medium	High	Quite High	Accept
2	Communication Issue	Low	Medium	Medium	Reduce - Maintain healthy communications
3	Incompatible system configuration	High	High	High	Avoid - Choose appropriate software versions
4	File Loss Issue	High	High	High	Avoid - Daily backup to a cloud such as git Hub
5	Poor materials or journals	Medium	Medium	Medium	Reduce - Selection of highly cited and relevant trusted journals.
6	Data authorization issue	High	High	High	Avoid - Already identified good open source data
7	Lack of Meeting deadline	High	High	High	Avoid - Try to finish each milestone as per the plan.
8	Lack of academic integrity	High	High	High	Avoid - confirm the research will conduct on the academic integrity



REFERENCE

AlSagri, H. S., & Ykhlef, M. (2020). Machine learning-based approach for depression detection in twitter using content and activity features. IEICE Transactions on Information and Systems, 103(8), 1825-1832.

Biradar, A., & Totad, S. G. (2018). Detecting depression in social media posts using machine learning. Paper presented at the International Conference on Recent Trends in Image Processing and Pattern Recognition, 716-725.

Burdisso, S. G., Errecalde, M., & Montes-y-Gómez, M. (2019). A text classification framework for simple and effective early depression detection over social media streams. Expert Systems with Applications, 133, 182-197.

Islam, M. R., Kamal, A. R. M., Sultana, N., Islam, R., & Moni, M. A. (2018). Detecting depression using k-nearest neighbours (knn) classification technique. Paper presented at the 2018 International Conference on Computer, Communication, Chemical, Material and Electronic Engineering (IC4ME2), 1-4.

Islam, M., Kabir, M. A., Ahmed, A., Kamal, A. R. M., Wang, H., & Ulhaq, A. (2018). Depression detection from social network data using machine learning techniques. Health Information Science and Systems, 6(1), 1-12.

Joshi, M. L., & Kanoongo, N. (2022). Depression detection using emotional artificial intelligence and machine learning: A closer review. Materials Today: Proceedings, 58, 217-226. https://doi.org/10.1016/j.matpr.2022.01.467

Rajaraman, P. V., Nath, A., Akshaya, P. R., & Bhuja, G. C. (2020). Depression detection of tweets and A comparative test. International Journal of Engineering Research, 9(03), 422-425.

Saunders, M. N. K. (2015). Research methods for business students. Pearson Education.

Shetty, N. P., Muniyal, B., Anand, A., Kumar, S., & Prabhu, S. (2020). Predicting depression using deep learning and ensemble algorithms on raw twitter data. International Journal of Electrical and Computer Engineering, 10(4), 3751.

Stephen, J. J., & Prabu, P. (2019). Detecting the magnitude of depression in Twitter users using sentiment analysis. International Journal of Electrical and Computer Engineering, 9(4), 3247.

Agile-Waterfall Hybrid: Is It Right for Your Team? | Lucidchart Blog. (2022). Retrieved 4 September 2022, from https://www.lucidchart.com/blog/is-agile-waterfall-hybrid-right-for-your-team

Appendix B – Ethics Checklist and Publication form

Sheffield Hallam University

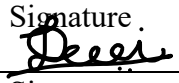
College of Business, Technology and Engineering

Research Skills and Dissertation Module (55-706556).

PUBLICATION PROCEDURE FORM

In this module, while you create your own research question or topic area, your supervisor makes a significant intellectual contribution to this work as the research progresses. Your supervisor will make the decision on whether your work merits publication based on the quality of the work you have produced. Your supervisor will co-author the paper for publication with you and your supervisor will both be listed as authors. You are required to sign the declaration below to confirm that you understand and will follow this procedure.

Declaration:

I ..Devi Priya Bijosh Mohan… confirm that I understand will comply with the Publication Procedure outlined in the Module Handbook and the Blackboard Site. G5		
Student:	Signature 	Date 01/11/2022
Supervisor:	Signature BayodeOgunleye	Date 01/11/2022

UREC2 RESEARCH ETHICS PROFORMA FOR STUDENTS UNDERTAKING LOW RISK PROJECTS WITH HUMAN PARTICIPANTS

This form is designed to help students and their supervisors to complete an ethical scrutiny of proposed research. The University Research Ethics Policy (https://www.shu.ac.uk/research/excellence/ethics-and-integrity/policies) should be consulted before completing the form. The initial questions are there to check that completion of the UREC 2 is appropriate for this study. The final responsibility for ensuring that ethical research practices are followed rests with the supervisor for student research.

Note that students and staff are responsible for making suitable arrangements to ensure compliance with the General Data Protection Act (GDPR). This involves informing participants about the legal basis for the research, including a link to the University research data privacy statement and providing details of who to complain to if participants have issues about how their data was handled or how they were treated (full details in module handbooks). In addition, the act requires data to be kept securely and the identity of participants to be anonymized. They are also responsible for following SHU guidelines about data encryption and research data management. Guidance can be found on the SHU Ethics Website https://www.shu.ac.uk/research/excellence/ethics-and-integrity

Please note that it is mandatory for all students to only store data on their allotted networked drive space and not on individual hard drives or memory sticks etc.

The present form also enables the University and College to keep a record confirming that research conducted has been subjected to ethical scrutiny.

The form must be completed by the student and the supervisor and independently reviewed by a second reviewer or module leader (additional guidance can be obtained from your College Research Ethics Chair[1]). In all cases, it should be counter-signed and kept as a record showing that ethical scrutiny has occurred. Some courses may require additional scrutiny. Students should retain a copy for inclusion in their research projects, and a copy should be uploaded to the relevant module Blackboard site.

Please note that it may be necessary to conduct a health and safety risk assessment for the proposed research (SECTION B). Further information can be obtained from the University's Health and Safety Website

SECTION A

1. Checklist questions to ensure that this is the correct form:

Health Related Research within the NHS, or Her Majesty's Prison and Probation Service (HMPPS), or with participants unable to provide informed consent check list.

Question	Yes/No
Does the research involve? • Patients recruited because of their past or present use of the NHS	No
• Relatives/carers of patients recruited because of their past or present use of the NHS	No
• Access to data, organs, or other bodily material of past or present NHS patients	No

Question	Yes/No
• Foetal material and IVF involving NHS patients	No
• The recently dead in NHS premises	No
• Prisoners or others within the criminal justice system recruited for health-related research	No
• Police, court officials, prisoners, or others within the criminal justice system	No
• Participants who are unable to provide informed consent due to their incapacity even if the project is not health related	No
• Is this an NHS research project, service evaluation or audit? *For NHS definitions please see the following website* http://www.hra.nhs.uk/documents/2013/09/defining-research.pdf	No

2. Checks for research with human participants

Question	Yes/No
1. Will any of the participants be vulnerable? *Note: Vulnerable people include children and young people, people with learning disabilities, people who may be limited by age or sickness, pregnancy, people researched because of a condition they have, etc. See full definition on ethics website in the document* ***Code of Practice for Researchers Working with Vulnerable Populations*** *(under the Supplementary University Polices and Good Research Practice Guidance)*	No
2. Are drugs, placebos, or other substances (e.g., food substances, vitamins) to be administered to the study participants or will the study involve invasive, intrusive, or potentially harmful procedures of any kind?	No
3. Will tissue samples (including blood) be obtained from participants?	No
4. Is pain or more than mild discomfort likely to result from the study?	No
5. Will the study involve prolonged or repetitive testing?	No
6. Is there any reasonable and foreseeable risk of physical or emotional harm to any of the participants? *Note: Harm may be caused by distressing or intrusive interview questions, uncomfortable procedures involving the participant, invasion of privacy, topics relating to highly personal information, topics relating to illegal activity, or topics that are anxiety provoking, etc.*	No
7. Will anyone be taking part without giving their informed consent?	No
8. Is it covert research? *Note: 'Covert research' refers to research that is conducted without the knowledge of participants.*	No
9. Will the research output allow identification of any individual who has not given their express consent to be identified?	No

3. General project details

Details	
Name of student	Devi Priya Bijosh Mohan
SHU email address	c1056531@hallam.shu.ac.uk
Department/College	Computing (MSc Big Data Analytics)/BTE
Name of supervisor	Bayode Ogunleye
Supervisor's email address	b.ogunleye@shu.ac.uk
Title of proposed research	Depression Detection using Sentimental Analysis
Proposed start date	19-09-2022
Proposed end date	09-01-2023
Background to the study and the rationale (reasons) for undertaking the research (500 words)	Depression is a mental illness that affects the quality of life which can lead to suicide. The major problem of the illness is that the chance of not knowing about the symptoms are very high. The illness got worse due to the pandemic situation and a lot of people committed suicide which makes the importance to detect it in the early stages. It can be quite challenging for health care organisations to identify depression in the early stages because people are reluctant to seek care when necessary or uninformed of their mental state. Several methods are proposed to help people who cannot receive adequate diagnosis and treatment. Depression can be diagnosed by medical methods such as medical history, physical exams, lab tests or psychological evaluation that answers questions about the thoughts, emotions, and behaviour. Though the methods mentioned before are effective for depression, early diagnosis, and the gap between who can and cannot receive the treatment is still relevant. This method is inadequate in the case of patients who are unaware about their mental stage. The issue is proven by the fact that only 4.6% of the world population suffer from depression, 43.3% of them do not take their symptoms seriously and do not care to be treated professionally. If the symptoms could be identified early, either by therapy session or medications, the affected people can come back to normal life. People became very comfortable expressing their ideas, sentiments, goals and other accomplishments in social media such as posts, messages, and comments rather than speaking or sharing the emotions to others. This can be used in a positive way such as identifying the symptoms of mental illness like depression by using natural language processing with machine learning techniques. The project's goal is to provide a framework to help healthcare institutions identify depression early on, reducing the likelihood that the illness would worsen because of ignorance. The

Details	
	development of a framework including sentiment analysis and machine learning models makes it simpler for health organisations to recognise people who are experiencing depression in its early stages and offer them with appropriate treatment by utilising the social networking data. Deliverable: A framework that can use an aid of detecting the initial symptoms of depression.
Aims & research question(s)	**Aim** The project's goal is to develop a framework using sentiment analysis techniques to detect early-stage depression. **Question** How can Twitter sentiment analysis help detect early stage depression?
Methods to be used for: 1. Recruitment of participants 2. Data collection 3. Data analysis	An open source Kaggle dataset is selected for this project. A quantitative analysis is planning to conduct using different Machine learning models. The research outcome will be evaluated by beneficiaries (the selection of beneficiaries will be non-random and will be done online)
Outline the nature of the data held, details of anonymization, storage and disposal procedures as required.	Not Applicable

4. Research in external organizations

Question	**Yes/No**
1. Will the research involve working with/within an external organization (e.g., school, business, charity, museum, government department, international agency, etc.)?	No

Question	Yes/No
2. If you answered YES to question 1, do you have granted access to conduct the research from the external organization? *If YES, students please show evidence to your supervisor. You should retain this evidence safely.*	
3. If you do not have permission for access is this because: A. you have not yet asked B. you have asked and not yet received an answer C. you have asked and been refused access. *Note: You will only be able to start the research when you have been granted access.*	

5. Research with products and artefacts

Question	Yes/No
1. Will the research involve working with copyrighted documents, films, broadcasts, photographs, artworks, designs, products, programs, databases, networks, processes, existing datasets, or secure data?	Yes
2. If you answered YES to question 1, are the materials you intend to use in the public domain? *Notes: 'In the public domain' does not mean the same thing as 'publicly accessible'.* • *Information which is 'in the public domain' is no longer protected by copyright (i.e., copyright has either expired or been waived) and can be used without permission.* • *Information which is 'publicly accessible' (e.g., TV broadcasts, websites, artworks, newspapers) is available for anyone to consult/view. It is still protected by copyright even if there is no copyright notice. In UK law, copyright protection is automatic and does not require a copyright statement, although it is always good practice to provide one. It is necessary to check the terms and conditions of use to find out exactly how the material may be reused etc.* *If you answered YES to question 1, be aware that you may need to consider other ethics codes. For example, when conducting Internet research, consult the code of the Association of Internet Researchers; for educational research, consult the Code of Ethics of the British Educational Research Association.*	Yes
3. If you answered NO to question 2, do you have explicit permission to use these materials as data? *If YES, please show evidence to your supervisor.*	
4. If you answered NO to question 3, is it because: A. you have not yet asked permission B. you have asked and not yet received and answer C. you have asked and been refused access. *Note You will only be able to start the research when you have been granted permission to use the specified material.*	

SECTION B

HEALTH AND SAFETY RISK ASSESSMENT FOR THE RESEARCHER

1. **Does this research project require a health and safety risk assessment for the procedures to be used?** Discuss this with your supervisor and consult the Risk Assessment Toolkit for teaching research.

   ☐ Yes
   ☒ No

   (If **YES** the completed Health and Safety Risk Assessment form should be attached). You can find a Blank/Sample Risk Assessment Form at the Checklist, Generic and TORS Risk Assessments on the Risk Assessment Toolkit

2. **Will the data be collected fully online (no face-to-face contact with participants)?**

   ☒ Yes (See the safety guidance for online research[2] and **go to question 8b**).
   ☐ No (Go to question 3)

3. **Will the proposed data collection take place on campus?**

   ☐ Yes (Please answer questions 5 to 8)
   ☒ No (Please complete all questions and consult with your supervisor or HoD for current guidance and permission for face-to-face research outside the university)

4. **Where will the data collection take place?**
   (Tick as many as apply if data collection will take place in multiple venues)

	Location	**Please specify**
☐	Researcher's Residence	
☐	Participant's Residence	
☐	Education Establishment	
☐	Other e.g., business/voluntary organisation, public venue	
☐	Outside UK	

5. **If face-to-face contact with participants is required for your study? Please stipulate below how you will comply with any government requirements related to Covid-19 and social distancing or other limitations on contact.**

6. How will you travel to and from the data collection venue?

☐ On foot ☐ By car ☐ Public Transport
☐ Other (Please specify)

Please outline how you will ensure your personal safety when travelling to and from the data collection venue (include any Covid-19 related precautions)

N/A

7. How will you ensure your own personal safety whilst at the research venue?

N/A

8. Are there any potential risks to your health and wellbeing associated with either (a) the venue where the research will take place and/or (b) the research topic itself?

☒ None that I am aware of
☐ Yes (Please outline below including steps taken to minimise risk)

9. If you are carrying out research off-campus, you must ensure that each time you go out to collect data you ensure that someone you trust knows where you are going (without breaching the confidentiality of your participants), how you are getting there (preferably including your travel route), when you expect to get back, and what to do should you not return at the specified time.

Please outline here the procedure you propose using to do this.

Adherence to SHU policy and procedures

Ethics sign-off	
Personal statement	
I can confirm that: • I have read the Sheffield Hallam University Research Ethics Policy and Procedures • I agree to abide by its principles.	
Student	
Name: Devi Priya Bijosh Mohan	Date: 05-09-2022
Signature: 	
Supervisor or another person giving ethical sign-off	

Ethics sign-off	
I can confirm that completion of this form has not identified the need for ethical approval by the TPREC/CREC or an NHS, Social Care, or other external REC. The research will not commence until any approvals required under Sections 4 & 5 have been received and any necessary health and safety measures are in place.	
Name: Dr. Bayode Ogunleye	Date: 01/11/2022
Signature: BayodeOgunleye	
Additional Signature if required by course leader:	
Name: .	Date:
Signature:	

Please ensure that you have attached all relevant documents. Your supervisor must approve them before you start data collection:

Documents	**Yes**	**No**	**N/A**
Research proposal if prepared previously	☒	☐	☐
Any recruitment materials (e.g., posters, letters, emails, etc.)	☐	☐	☒
Participant information sheet[3]	☒	☐	☐
Participant consent form[4]	☒	☐	☐
Details of measures to be used (e.g., questionnaires, etc.)	☐	☐	☒
Outline interview schedule / focus group schedule	☐	☐	☒
Debriefing materials	☐	☐	☒
Health and Safety Risk Assessment Form	☐	☐	☒

Appendix C – Links to Dataset

https://github.com/dpriyabijosh/Depression_detection/Data

Appendix D – Information Sheet and Consent form

Early Depression Detection System(EDDS) Validation

* Required

Participant Information Section

Dear Sir/Madam,

I am a Masters student at Sheffield Hallam University (SHU), United Kingdom and am conducting a research on depression detection using Transformer models utilising twitter data. The aim is to develop a framework with sentiment analysis and machine learning models to detect early-stage depression. To achieve this, pre-trained Language Models are used as a tool which is capable of analysing the sentiment behind people's posts and predict whether they are depressed or not. Depression is one of the most prevalent mental illnesses which adversely affects people's lives. According to WHO, approximately 280 million people worldwide experience depression, which is a leading cause of disability and a major contributor to the overall burden of diseases. More over, it is proved to be one of the main causes of suicide so it is crucial to diagnose as soon as possible. Peopl became very comfortable expressing their emotions in social media such as posts, messages, and comments rather than speaking or sharing the emotions to others. From the previous researches, it is evident that the social media datasets are efficient in detecting users' emotional statements and potential mental illness All the above-mentioned reasons influence me to conduct a research to detect depression using already available Twitter dataset.
The Early Stage Depression Detection System(EDDS) is a classification framewor that can use an aid to detect depression detection in its early stages by analysing the social media user's posts/comments of a user. This is a classification system that classifies the post into depressed or non-depressed using Machine Learning models. It is important to know your views regarding the EDDS research outcome in terms of relevance and usefulness. It will helps to establish the relevance of the reattach findings and formulate recommendations.

I would like to thank you in advance for your values and kind consideration. At SHU, confidentiality and anonymity are guaranteed as all the information gathered will conform to the University's Ethical procedure (https://www.shu.ac.uk/research/excellence/ethics-and-integrity/policies). The University undertakes research as part of its function for the community under its legal status. Data protection allows us to use personal data for research with appropriate safeguards in place under the legal basis of public tasks that are in the public interest. A full statement of your rights can be found at https://www.shu.ac.uk/about-this-website/privacy-policy/privacy-notes/privacy-notice-for-research. All University research is reviewed to ensure that participants are treated appropriately and their rights respected. If you would like to receive

further information about the research, please free to contact me (c1056531@hallam.shu.ac.uk).

Devi Priya Bijosh Mohan
MSc Big Data Analytics
Sheffield Hallam University
Sheffield
United Kingdom
Email: c1056531@hallam.shu.ac.uk

Participant Consent Form

1. I have read the Participant Information Section for this study and have had details of the study explained to me. I understand that I am free to withdraw from the study, without giving a reason for my withdrawal or to decline to answer any particular questions in the study without any consequences to my future treatment by the researcher. I consent to the information collected for the purposes of this research study, once anonymized (so that I cannot be identified), to be used for any other research purposes.

   Mark only one oval.

   - ◯ No
   - ◯ Yes

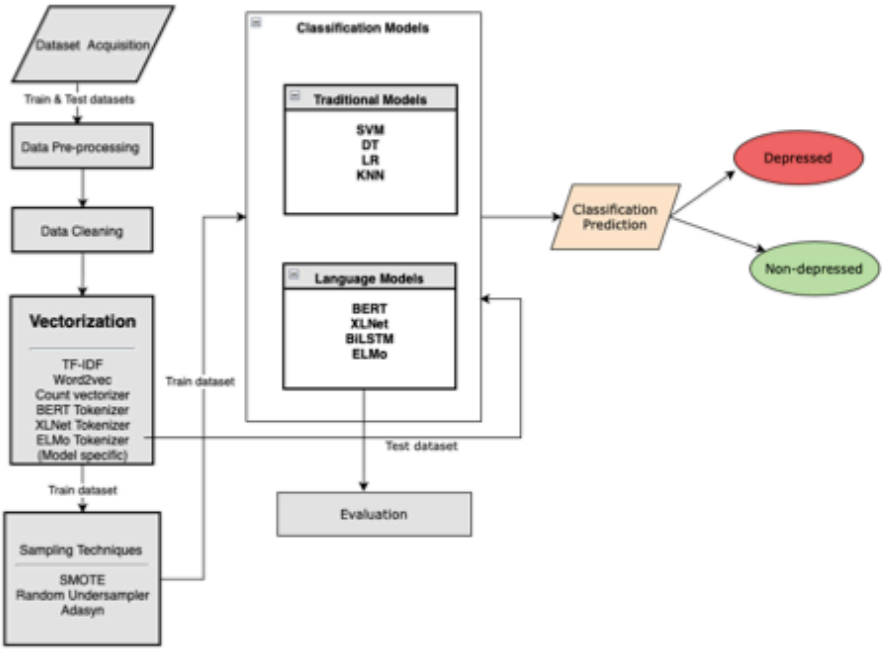
2. Please indicate your willingness to participate in this exercise by clicking on your preferred option below. I assure you that the data and information provided will remain strictly confidential and will be used for the purpose of this research.

   Mark only one oval.

   - ◯ No, I am not willing to participate
   - ◯ Yes, I am willing to participate

Appendix E – Framework and Questionnaire for gathering Validation Data

Proposed EDDS Framework



Research Feedback Form

Please provide responses to the questions

3. 1. Do you use Social media?

Mark only one oval.

- ◯ No
- ◯ Yes

4. 2. How often do you use social media?

Check all that apply.

- [ ] More than once a day
- [ ] Once daily
- [ ] More than once weekly
- [ ] Once weekly
- [ ] More than once in a month
- [ ] Once monthly
- [ ] Other: _____

5. 3. Which of the social media you use?

Check all that apply.

- [ ] Facebook
- [ ] Instagram
- [ ] Twitter
- [ ] Linkedin
- [ ] Other: _____

6. 4. For what purpose do you use social media?

Check all that apply.

- [ ] Connect with friends and family
- [ ] News
- [ ] Work
- [ ] Shopping
- [ ] Other: _____

7. 5. What is your gender?

Mark only one oval.

- ◯ Male
- ◯ Female
- ◯ Prefer not to say

8. 6. Are you a doctor? *

Mark only one oval.

- ◯ Yes
- ◯ No

Research Feedback Form 2

Please provide response on how valid the research claims and findings are with regards to the experience in medical field.

9. 7. Depression became a major problem nowadays and the pandemic situation also increas in prevalence of anxiety and depression worldwide. To what extend do you agree or disagre with this?

Mark only one oval.

- ◯ Strongly disagree
- ◯ Disagree
- ◯ Neutral
- ◯ Agree
- ◯ Strongly agree

10. 8. People are not getting enough support and the increase in suicide make an urgency of system to identify depression in its early stage. Do you agree on it?

Mark only one oval.

- Strongly disagree
- Disagree
- Neutral
- Agree
- Strongly agree

11. 9. Depression takes hold gradually, without a person realising he/she is depressed. It will result in a significant health deterioration and even the possibility of suicide is high. To wha extend do you agree/disagree with this?

Mark only one oval.

- Strongly disagree
- Disagree
- Neutral
- Agree
- Strongly agree

12. 10. Conventional methods are not helpful to identify people with depression symptoms du to unawareness or self denial. To what extend do you agree/disagree with this?

Mark only one oval.

- Strongly disagree
- Disagree
- Neutral
- Agree
- Strongly agree

13. 11. People are expressing themselves in social media than interacting with people. Hence social media can be used as a fruitful source to understand their state of mind. To what extend do you agree/disagree with this?

Mark only one oval.

- ◯ Strongly disagree
- ◯ Disagree
- ◯ Neutral
- ◯ Agree
- ◯ Strongly agree

14. 12. People freely express their emotions and sentiments in social media platforms. These data can be efficient utilised to detect depression in its early stage. To what extent do you agree/disagree with this?

Mark only one oval.

- ◯ Strongly disagree
- ◯ Disagree
- ◯ Neutral
- ◯ Agree
- ◯ Strongly agree

15. 13. Sentiment analysis system with machine learning models helps to classify people as depressed and non-depressed without any hassle, just by using their social media data. D you agree this will helps to the health sector?

Mark only one oval.

- ◯ Strongly disagree
- ◯ Disagree
- ◯ Neutral
- ◯ Agree
- ◯ Strongly agree

16. 14. The use of social media is increasing day by day. An automatic system which can classify their mental stage as depressed or non-depressed is considered beneficial to identify whether they need to seek support or not. To what extend do you agree or disagre this?

Mark only one oval.

- ◯ Yes
- ◯ No
- ◯ Maybe

17. 15. An early depression detection system is a COMPULSORY system for health sectors t provide necessary support to the people who suffer from depression. To what extend do y agree/disagree this?

Mark only one oval.

- ◯ Yes
- ◯ No
- ◯ Maybe

18. Please provide if you have any additional comments.

Appendix n – Codes and Outputs

The code generated for this study is in the below GitHub link.

https://github.com/dpriyabijosh/Depression_detection