A PROJECT REPORT

ON

DETECTION OF DEPRESSION-RELATED POSTS IN REDITT SOCIAL MEDIA FORUM

Submitted in partial fulfillment of requirements to

ACHARYA NAGARJUNA UNIVERSITY

For the Award of the Degree
B. Tech in IT

BY

M. TEJA VARDHAN(Y18IT103)

Y. SAI TEJA(Y18IT090)

J. ASHOK(L19IT125)

NOVEMBER & 2021

RVR & JC COLLEGE OF ENGINEERING(AUTONOMOUS)
(NAAC A$^+$)
ACHARYA NAGARJUNA UNIVERSITY: GUNTUR 522 510

DETECTION OF DEPRESSION-RELATED POSTS IN REDITT SOCIAL MEDIA FORUM

A PROJECT REPORT

Submitted by

M.TEJA VARDHAN(Y18IT103)
Y.SAI TEJA(Y18IT090)
J. ASHOK (L19IT125)

---

in partial fulfillment for the award of the degree

of

BACHELOR OF TECHNOLOGY

in

INFORMATION TECHNOLOGY

DEPARTMENT OF INFORMATION TECHNOLGY
R V R & J C COLLEGE OF ENGINEERING, GUNTUR
(AUTONOMOUS)
NAGRJUNA UNIVERSITY: GUNTUR 522 510

NOVEMBER 2021

RVR & JC COLLEGE OF ENGINEERING (AUTONOMOUS) GUNTUR 522 019

BONAFIDE CERTIFICATE

Certified that this project report "**DETECTION OF DEPRESSION-RELATED POSTS IN REDITT SOCIAL MEDIA FORUM"** is the bonafide work of **"M. TEJA VARDHAN(Y18IT103), Y.SAI TEJA(Y18IT090), J. ASHOK(L19IT125)"** who carried out the project work under my supervision.

SIGNATURE	**SIGNATURE**
MRS. B. MANASA	**DR.A. SRIKRISHNA**
SUPERVISOR	HEAD OF DEPARTMENT
ASSISTANT PROFESSOR	PROFESSOR
INFORMATION TECHNOLOGY	INFORMATION TECHNOLOGY

TABLE OF CONTENTS

ABSTRACT

Depression is viewed as the largest contributor to global disability and a major reason for suicide. It has an impact on the language usage reflected in the written text. The key objective is to examine social media users' posts to detect any factors that may reveal the depression attitudes of relevant online users. For such purpose, we employ the Natural Language Processing (NLP) techniques and Machine Learning(ML) approaches to train the data and evaluate the efficiency of our proposed method.

We perform sentiment analysis to identify the polarity of the text with respect to our context "depression" using NLP. Later, we build models which predicts the polarity using ML.

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS, ABBREVIATIONS

S.No.	Abbrevation	Full Form
1	NLP	Natural Language Processing
2	LR	Logistic Regression
3	SVM	Support Vector Machine
4	RF	Random Forest
5	MLP	Multi-Layer Perceptron
6	BOW	Bag Of Words
7	LIWC	Linguistic Inquiry and Word Count
8	LDA	Latent Dirichlet Allocation
9	TFIDF	TermFrequency Inverse Document Frequency
10	CNN	Convolutional Neural Networks

1. INTRODUCTION

1.1. Background.

Depression is considered as one of the major reasons for suicides. According to the survey, people usually reveal their depression through their social media posts. These posts can be useful in analyzing the text and to predict the polarity. The depression in the text can be identified by using the Natural Language Processing (NLP).

NLP is a field that covers linguistics, computer understanding and manipulation of human language. It is a subset of artificial intelligence in which computers analyze, understand, and derive meaning from human language in a smart and useful way. NLP can be used to program the computers in order to process and analyze large amounts of natural language data. Natural language refers to the way that humans communicate with each other, that is either speech or text.

NLP is useful in many real word applications that are based on text. It is also very helpful in performing subtasks used in various applications. By utilizing NLP, developers can organize and structure knowledge to perform tasks such as automatic summarization, translation, named entity recognition, relationship extraction, sentiment analysis, speech recognition, and topic segmentation.

The overall NLP projects includes three segments. They are Preprocessing, Feature Extraction and Applying Supervised learning or Unsupervised learning algorithms. They use the machine learning classifiers for the purpose of classification.

1.1.1. ***Data Preprocessing.***

Data preprocessing is the process of removing the data which is not useful in categorizing the text. If the unwanted data is not removed, it may add redundancy and leads to the unwanted results. Some of the examples are URL’s (Uniform Resource Locator), mentions, punctuations, stop words, etc. They do not carry any significant meaning in classifying the text, so the clean text must be free from these type of categories.

Preprocessing transforms text into a more digestible form by discarding less informative words so that machine learning algorithms can perform better. It is an important step for tuning the text so that it is easy to extract important features. Pre-processing is important for two reasons. One is to remove the unwanted data in order to get consistent results. The other reason is Dimensionality Reduction. More number of words in the text results in more number of features. In order to reduce the dimensions, we have to trim unimportant words. On this pre-processed text, the selected feature extraction methods will be applied.

1.1.2. ***Feature Extraction.***

Feature Extraction is the most important step in NLP. The classification algorithm works only when it is provided with the features. In the text, the features are nothing but words. So, those words are to be extracted from the text. The next task is to know whether a particular word can be used as a feature or not. There are many statistical ways to determine these features. In a simple way, the important words in the text are treated as the features.

There are many feature extraction methods like Bag of Words(BOW)[38],[51], N-grams[2], LIWC[19], LDA[14],[44], TF-IDF, etc. Out of these, N-grams + TF-IDF, Linguistic Inquiry and Word Count (LIWC), Latent Dirichlet Allocation(LDA) are used.

N-Grams + TF-IDF.

The main function of feature extraction is selecting important words as features. A statistic is needed to determine whether a word is important or not. Here that statistic is nothing but TF-IDF. TF-IDF is used to rank words.

Tfidf is meant for rendering more importance to the rare words. It so happens that if you rely on word counts alone, the unimportant words like 'the', 'and', etc. will get more importance because they tend to get used more often. For example, a 1000-word document on data science will have the word data science only twice or thrice whereas it will have the word 'the' for at least 10 to 15 times. In order to arrest this problem, Tfidf is used.

LIWC.

Generally, a word can be used in multiple contexts. By understanding the sentence, a human can determine the context. But to a machine it is a difficult task. LIWC can be used to determine different contexts of the word. So that, we can determine which context leads to a positive sentiment or a negative sentiment.

Instead of using words as features, LIWC psycholinguistic categories are used as features in order to train models. The main advantage is; it considers different interpretations of words.

LDA.

Latent Dirichlet Allocation is a type of topic modelling. Here, instead of using words or categories as features, topics are used as features. A collection of words can be represented as a topic.

LDA is used to classify text in a document to a particular topic. It builds a topic per document model and words per topic model, modeled as Dirichlet distributions. As it extracts unique topics for each document, it is easy to use. One more advantage is that it is quick compared to remaining feature extraction methods.

1.1.3. ***Classification.***

Once the features are obtained, Machine Learning algorithms can be applied to identify the apt sentiment of the text. Usually multiple machine learning algorithms are preferred. The model which gives the better accuracy is the better model.

Supervised learning algorithms are used when the data contains class labels. The models are trained with the labelled data and predicts the class labels for the test data. Unsupervised learning algorithms can be used when the data does not contain any class labels or target attributes. In such cases, the algorithms try to establish the relations among the data.

1.2. **Problem Statement.**

Depression is a common mental health disorder. It often co-occurs with anxiety or other psychological and physical disorders [60]. It has an impact on the feelings and behavior of the affected individuals. When the depressed persons are not given any treatment, it can lead into a serious self-perception and at its worst, to suicide [6],[36].

With the development of Internet usage, people have started to share their experiences and challenges with mental health disorders through online forums, microblogs and tweets [47], [59].

1.3. **Objectives.**

The main objectives of this work are:

- To examine the relationship between depression and user's language usage.
- To design three LIWC features, N-grams, LIWC and LDA.
- To evaluate the power of N-grams probabilities, LIWC and LDA as single features for performance accuracy.
- To show the predictive power of both single and combined features with proposed classification approaches to achieve a higher performance in depression identification tasks.

1.4. **Need for Present Study.**

Using different Natural Language Processing techniques and text classification approaches, many techniques have been developed for text classification and depression detection. Some of them use single set features like Bag of Words(BOW), N-grams, LIWC or LDA to identify depression in the text [15],[23],[53],[57]. Some other techniques compare the performance of individual features with various machine learning classifiers. Recent studies examine the power of single features and their combinations such as N-grams + LIWC [30], BOW + LDA and TFIDF + LDA [64] to improve accuracy results.

The individual features have their own advantages and disadvantages. The main drawback of n-grams is that they are not designed to model linguistic knowledge. But n-grams are easy to calculate. LIWC has the advantage of obtaining the semantics in the text. But the dictionary need to be updated to the present scenario. Also, it is a time taking process. LDA has the advantage of text distribution based on high probability topics.

Similarly, different machine learning classifiers also have both advantages and disadvantages. Logistic Regression works only with linearly separable data and it may over fit the data. But LR is very easy to interpret and can quickly update to incorporate new data. Support Vector Machine can predict in variety of situations with low generalization errors. But SVM is computationally expensive, highly complex and requires more memory and time for training the model. Random Forest is highly efficient on large datasets and can deal with high dimensional data. But RF observed to be over fit for some datasets with noisy classification tasks. Also, more number of trees makes the algorithm slow for real-time prediction. Adaptive boosting is sensitive to noisy data and outliers but it is simple to implement and has fairly good generalization. Multi-Layer Perceptron is powerful and can model complex functions. Also it can adapt to unknown situations. But MLP can get stuck in local minima and the number of hidden layers are to be set very carefully.

The present work aims to achieve higher accuracy and increase the performance of the model through a proper features selection and their multiple feature combinations.

2. **LITERATURE REVIEW**

2.1. **Depression and Social media.**

M. J. Friedrich [36]: The proportion of the global population living with depression is estimated to be 322 million people - 4.4% of the world's population - according to a new report, "Depression and Other Common Mental Disorders: Global Health Estimates", released by the World Health Organization. The report also includes data on anxiety disorders, which affect more than 260 million people - 3.6% of the global population. The prevalence of these common mental disorders is increasing, particularly in low and middle income countries, with many people experiencing both depression and anxiety disorders simultaneously.

Aaron Beck et al. [7]: It is highly likely that psychiatric historians writing in the next century will characterize our knowledge of depression in the 1960's as having been promising and imaginative but also relatively unsystematic and beset by ideological differences. If this characterization turns out to be a valid one, current reviews of the field of depression are limited by significant constraints. Within the boundaries of these barriers Beck has performed a useful service for mental health professionals by presenting a reasonably well-integrated summary of salient evidence, inference, and opinion regarding depression. In addition to providing us with general information from an historical and contemporary perspective, this book leans heavily upon the experimental work carried out by the author and his colleagues. The admixture of broad survey and research report vitalizes the book and, on the whole, makes it more readable.

Pyszvzynsky and Greenberg [58]: They applied theory and research on self-focused attention and self-regulatory processes to the problems of depression and use this framework to integrate the roles played by a variety of psychological processes emphasized by other theories of the development and maintenance of depression. They proposed that depression occurs after the loss of an important source of selfworth when an individual becomes stuck in a self-regulatory cycle in which no responses to reduce the discrepancy between actual and desired states are available.

Consequently, the individual falls into a pattern of virtually constant self-focus, resulting in intensified negative affect, self-derogation, further negative outcomes, and a depressive self-focusing style. Eventually, these factors lead to a negative self-image, which may take on value by providing an explanation for the individual's plight and by helping the individual avoid further disappointments. The depressive self-focusing style then maintains and exacerbates the depressive disorder. They reviewed findings from laboratory studies of mild to moderately depressed people, correlational studies of more severely depressed people, and clinical observations with respect to consistency with the theory.

Durkheim's [16]: Emile Durkheim's Suicide addresses the phenomenon of suicide and its social causes. Written by one of the world's most influential sociologists, this classic argues that suicide primarily results from a lack of integration of the individual into society. Suicide provides readers with an understanding of the impetus for suicide and its psychological impact on the victim, family, and society.

Masuda et al. [41]: Suicide explains the largest number of death tolls among Japanese adolescents in their twenties and thirties. Suicide is also a major cause of death for adolescents in many other countries. Although social isolation has been implicated to influence the tendency to suicidal behavior, the impact of social isolation on suicide in the context of explicit social networks of individuals is scarcely explored. To address this question, they examined a large data set obtained from a social networking service dominant in Japan.

The social network is composed of a set of friendship ties between pairs of users created by mutual endorsement. They carried out the logistic regression to identify users' characteristics, both related and unrelated to social networks, which contribute to suicide ideation. They defined suicide ideation of a user as the membership to at least one active user-defined community related to suicide. They found that the number of communities to which a user belongs to, the intransitivity (i.e., paucity of triangles including the user), and the fraction of suicidal neighbors in the social network, contributed the most to suicide ideation in this order. Other characteristics including the age and gender contributed little to suicide ideation. They also found qualitatively the same results for depressive symptoms.

Eichstaedt, J.C. et al. [27]: Depression, the most prevalent mental illness, is underdiagnosed and undertreated, highlighting the need to extend the scope of current screening methods. Here, they used language from Facebook posts of consenting individuals to predict depression recorded in electronic medical records. They accessed the history of Facebook statuses posted by 683 patients visiting a large urban academic emergency department, 114 of whom had a diagnosis of depression in their medical records. Using only the language preceding their first documentation of a diagnosis of depression, they could identify depressed patients with fair accuracy [area under the curve (AUC) = 0.69], approximately matching the accuracy of screening surveys benchmarked against medical records.

Restricting Facebook data to only the 6 months immediately preceding the first documented diagnosis of depression yielded a higher prediction accuracy (AUC =0.72) for those users who had sufficient Facebook data. Significant prediction of future depression status was possible as far as 3 months before its first documentation. They founded that language predictors of depression include emotional (sadness), interpersonal (loneliness, hostility), and cognitive (preoccupation with the self, rumination) processes. Unobtrusive depression assessment through social media of consenting individuals may be come feasible as a scalable complement to existing screening and monitoring procedures.

Moreno et al. [33]: Depression is common and frequently undiagnosed among college students. Social networking sites are popular among college students and can include displayed depression references. The purpose of this study was to evaluate college students' Facebook disclosures that met DSM (Diagnostic and Statistical Manual) criteria for a depression symptom or a major depressive episode (MDE). They selected public Facebook profiles from sophomore and junior undergraduates and evaluated personally written text: “status updates”. They applied DSM criteria to 1-year status updates from each prole to determine prevalence of displayed depression symptoms and MDE criteria.

Negative binomial regression analysis was used to model the association between depression disclosures and demographics or Facebook use characteristics. Two hundred proles were evaluated, and prole owners were 43.5% female with a mean age of 20 years. Overall, 25% of proles displayed depressive symptoms and 2.5% met criteria for MDE. Prole owners were more likely to reference depression, if they averaged at least one online response from their friends to a status update disclosing depressive symptoms, or if they used Facebook more frequently. College students commonly display symptoms consistent with depression on Facebook. Their findings suggest that those who receive online reinforcement from their friends are more likely to discuss their depressive symptoms publicly on Facebook. Given the frequency of depression symptom displays on public proles, social networking sites could be an innovative avenue for combating stigma surrounding mental health conditions or for identifying students at risk for depression.

Schwartz et al. [23]: Depression is typically diagnosed as being present or absent. However, depression severity is believed to be continuously distributed rather than dichotomous. Severity may vary for a given patient daily and seasonally as a function of many variables ranging from life events to environmental factors. Repeated population-scale assessment of depression through questionnaires is expensive. In this paper they used survey responses and status updates from 28,749 Facebook users to develop a regression model that predicts users’ degree of depression based on their Facebook status updates. Their user-level predictive accuracy was modest, significantly outperforming a baseline of average user sentiment. They used the model to estimate user changes in depression across seasons, and find, consistent with literature, users’ degree of depression most often increases from summer to winter. They then had shown the potential to study factors driving individuals’ level of depression by looking at its most highly correlated language features.

M. Nadeem. [38]: Social media has recently emerged as a premier method to disseminate information online. Through these online networks, tens of millions of individuals communicate their thoughts, personal experiences, and social ideals. They therefore explored the potential of social media to predict, even prior to onset, Major Depressive Disorder (MDD) in online personas. They employed a crowdsourced method to compile a list of Twitter users who profess to being diagnosed with depression. Using up to a year of prior social media postings, they utilized a Bag of Words approach to quantify each tweet. Lastly, they leveraged several statistical classifiers to provide estimates to the risk of depression.

This work posits a new methodology for constructing the classifier by treating social as a text-classification problem, rather than a behavioral one on social media platforms. By using a corpus of 2.5M tweets, they achieved an 81% accuracy rate in classification, with a precision score of 0.86. They believe that this method may be helpful in developing tools that estimate the risk of an individual being depressed, can be employed by physicians, concerned individuals, and healthcare agencies to aid in diagnosis, even possibly enabling those suffering from depression to be more proactive about recovering from their mental health.

De Choudhury et al. [35]: They consider social media as a promising tool for public health, focusing on the use of Twitter posts to build predictive models about the influence of childbirth on the forthcoming behavior and mood of new mothers. Using Twitter posts, they quantify postpartum changes in 376 mothers along dimensions of social engagement, emotion, social network, and linguistic style. They then constructed statistical models from a training set of observations of these measures before and after the reported childbirth, to forecast significant postpartum changes in mothers.

The predictive models can classify mothers who will change significantly following childbirth with an accuracy of 71%, using observations about their prenatal behavior, and as accurately as 80-83% when additionally leveraging the initial 2-3 weeks of postnatal data. The study is motivated by the opportunity to use social media to identify mothers at risk of postpartum depression, an underreported health concern among large populations, and to inform the design of low-cost, privacy-sensitive early-warning systems and intervention programs aimed at promoting wellness postpartum.

Choudhury et al. [34]: History of mental illness is a major factor behind suicide risk and ideation. However, research efforts toward characterizing and forecasting this risk is limited due to the paucity of information regarding suicide ideation, exacerbated by the stigma of mental illness. This paper fills gaps in the literature by developing a statistical methodology to infer which individuals could undergo transitions from mental health discourse to suicidal ideation. They utilized semi-anonymous support communities on Reddit as unobtrusive data sources to infer the likelihood of these shifts. They developed language and interactional measures for this purpose, as well as a propensity score matching based statistical approach. Their approach allows to derive distinct markers of shifts to suicidal ideation. These markers can be modeled in a prediction framework to identify individuals likely to engage in suicidal ideation in the future.

Guntuku et al. [48]: Although rates of diagnosing mental illness have improved over the past few decades, many cases remain undetected. Symptoms associated with mental illness are observable on Twitter, Facebook, and web forums, and automated methods are increasingly able to detect depression and other mental illnesses. In this paper, recent studies that aimed to predict mental illness using social media are reviewed. Mentally ill users have been identified using screening surveys, their public sharing of a diagnosis on Twitter, or by their membership in an online forum, and they were distinguishable from control users by patterns in their language and online activity. Automated detection methods may help to identify depressed or otherwise at-risk individuals through the large-scale passive monitoring of social media, and in the future may complement existing screening procedures.

Reece et al. [3]: They developed computational models to predict the emergence of depression and Post-Traumatic Stress Disorder(PTSD) in Twitter users. Twitter data and details of depression history were collected from 204 individuals (105 depressed, 99 healthy). They extracted predictive features measuring affect, linguistic style, and context from participant tweets (N = 279,951) and built models using these features with supervised learning algorithms. Resulting models successfully discriminated between depressed and healthy content, and compared favorably to general practitioners' average success rates in diagnosing depression, albeit in a separate population. Results held even when the analysis was restricted to content posted before first depression diagnosis. State-space temporal analysis suggests that onset of depression may be detectable from Twitter data several months prior to diagnosis. Predictive results were replicated with a separate sample of individuals diagnosed with PTSD (Nusers = 174, Ntweets = 243,775). A state-space time series model revealed indicators of PTSD almost immediately post-trauma, often many months prior to clinical diagnosis. These methods suggest a data-driven, predictive approach for early screening and detection of mental illness.

Tsugawa et al. [53]: In this paper, they extensively evaluated the effectiveness of using a user's social media activities for estimating degree of depression. As ground truth data, they used the results of a web-based questionnaire for measuring degree of depression of Twitter users. They extracted several features from the activity histories of Twitter users. By leveraging these features, they constructed models for estimating the presence of active depression. Through experiments, they showed that (1) features obtained from user activities can be used to predict depression of users with an accuracy of 69%, (2) topics of tweets estimated with a topic model are useful features, (3) approximately two months of observation data are necessary for recognizing depression, and longer observation periods do not contribute to improving the accuracy of estimation for current depression; sometimes, longer periods worsen the accuracy.

2.2. **Natural Language Processing.**

Coppersmith et al. [19]: Many significant challenges exist for the mental health field, but one in particular is a lack of data available to guide research. Language provides a natural lens for studying mental health - much existing work and therapy have strong linguistic components, so the creation of a large, varied, language-centric dataset could provide significant grist for the field of mental health research. They examined a broad range of mental health conditions in Twitter data by identifying self-reported statements of diagnosis. They systematically explored language differences between ten conditions with respect to the general population, and to each other. Their aim was to provide guidance and a roadmap for where deeper exploration is likely to be fruitful.

Sigmund Freud [49] wrote about Freudian slips or linguistic mistakes to reveal the secret thoughts and feelings of the writers. With the development of sociology and psycholinguistic theories, various approaches towards the relationship between depression and its language have been defined. Professor Freud developed his system of psychoanalysis while studying the so-called borderline cases of mental diseases, such as hysteria and compulsion neurosis. By discarding the old methods of treatment and strictly applying himself to a study of the patient's life he discovered that the hitherto puzzling symptoms had a definite meaning, and that there was nothing arbitrary in any morbid manifestation. Psychoanalysis always showed that they referred to some definite problem or conflict of the person concerned. It was while tracing back the abnormal to the normal state that Professor Freud found how faint the line of demarcation was between the normal and neurotic person, and that the psychopathologic mechanisms so glaringly observed in the psychoneuroses and psychoses could usually be demonstrated in a lesser degree in normal persons. With great ingenuity and penetration, the author throws much light on the complex problems of human behavior, and clearly demonstrates that the hitherto considered impassable gap between normal and abnormal mental states is more apparent than real.

Rude et al. [52]: Essays written by currently-depressed, formerly-depressed, and never-depressed college students were examined for differences in language that might shed light on the cognitive operations associated with depression and depression vulnerability. A text analysis program computed the incidence of words in predesignated categories.

Consistent with Beck's cognitive model and with Pyczsinski and Greenberg's self-focus model of depression, depressed participants used more negatively valence words and used the word," I" more than did never-depressed participants. Formerly depressed (presumably depression-vulnerable) participants did not die from never depressed participants on these indices of depressive processing.

However, consistent with prediction, formerly-depressed participants' use of the word" I" increased across the essays and was significantly greater than that of never depressed writers in the final portion of the essays.

Zinken et al. [31]: This study investigated whether an analysis of narrative style (word use and cross-clausal syntax) of patients with symptoms of generalized anxiety and depression disorders can help predict the likelihood of successful participation in guided self-help. Texts by 97 people who had made contact with a primary care mental health service were analyzed. Outcome measures were completion of the guided self-help programme, and change in symptoms assessed by a standardized scale (CORE-OM).

Regression analyses indicated that some aspects of participants' syntax helped to predict completion of the programme, and that aspects of syntax and word use helped to predict improvement of symptoms. Participants using non-finite complement clauses with above-average frequency were four times more likely to complete the programme (95% confidence interval 1.4 to 11.7) than other participants. Among those who completed, the use of causation words and complex syntax (adverbial clauses) predicted improvement, accounting for 50% of the variation in well-being benefit.

Nguyen et al. [57]: A large number of people use online communities to discuss mental health issues, thus oaring opportunities for new understanding of these communities. This paper aims to study the characteristics of online depression communities (CLINICAL) in comparison with those joining other online communities (CONTROL).

They used machine learning and statistical methods to discriminate online messages between depression and control communities using mood, psycholinguistic processes and content topics extracted from the posts generated by members of these communities. All aspects including mood, the written content and writing style are found to be significantly different between two types of communities.

Sentiment analysis shows the clinical group have lower valence than people in the control group. For language styles and topics, statistical tests reject the hypothesis of equality on psycholinguistic processes and topics between two groups.

They had shown good predictive validity in depression classification using topics and psycholinguistic clues as features. Clear discrimination between writing styles and contents, with good predictive power is an important step in understanding social media and its use in mental health.

Tyshchenko et al. [64] suggested categorizing the stop words and adding LIWC like word categories as an extra feature to an already designed method (BOW + TFIDF + LIWC). In addition, he applied multiple feature combinations to increase the performance using Convolutional Neural Networks (CNN) which consist of neurons with learnable weights and differ in terms of their layers. CNNs are very similar to simple feed-forward neural networks and state of the art method in the text and sentence classification tasks.

Calvo et al. [45]: Natural language processing (NLP) techniques can be used to make inferences about peoples' mental states from what they write on Facebook, Twitter and other social media. These inferences can then be used to create online pathways to direct people to health information and assistance and also to generate personalized interventions. Regrettably, the computational methods used to collect, process and utilize online writing data, as well as the evaluations of these techniques, are still dispersed in the literature.

This paper provides a taxonomy of data sources and techniques that have been used for mental health support and intervention. Specifically, they review how social media and other data sources have been used to detect emotions and identify people who may be in need of psychological assistance; the computational techniques used in labeling and diagnosis; and finally, they discuss ways to generate and personalize mental health interventions.

The overarching aim of this scoping review is to highlight areas of research where NLP has been applied in the mental health literature and to help develop a common language that draws together the fields of mental health, human-computer interaction and NLP.

Stir man and Penne baker [55]: The purpose of this study was to determine whether distinctive features of language could be discerned in the poems of poets who committed suicide and to test two suicide models by use of a text-analysis program. Approximately 300 poems from the early, middle, and late periods of nine suicidal poets and nine non suicidal poets were compared by use of the computer text analysis program, Linguistic Inquiry and Word Count (LIWC).

Language use within the poems was analyzed within the context of two suicide models. In line with a model of social integration, writings of suicidal poets contained more words pertaining to the individual self and fewer words pertaining to the collective than did those of non-suicidal poets. In addition, the direction of effects for words pertaining to communication was consistent with the social integration model of suicide.

The study found support for a model that suggests that suicidal individuals are detached from others and are preoccupied with self. Furthermore, the findings suggest that linguistic predictors of suicide can be discerned through text analysis.

A. Hern´andez-Castan˜eda and H. Calvo [4]: They identified deceptive text by using different kinds of features: A continuous semantic space model based on latent dirichlet allocation topics (LDA), one-hot representation (OHR), syntactic information from syntactic n-grams (SN), and lexicon-based features using the linguistic inquiry and word count dictionary (LIWC). Several combinations of these features were tested to assess the best source(s) for deceptive text identification. By selecting the appropriate features, they were able to obtain a benchmark-level performance using a Naive Bayes classifier.

They tested on three different available corpora: A corpus consisting of 800 reviews about hotels, a corpus consisting of 600 reviews about controversial topics, and a corpus consisting of 236 book reviews. They found that the merge of both LDA features and OHR yielded the best results, obtaining accuracy above 80% in all tested datasets.

Additionally, this combination of features has the advantage that language specific-resources are not required (e.g. SN, LIWC), compared to other reference works. Additionally, they presented an analysis on which features lead to either deceptive or truthful texts, finding that certain words can play different roles (sometimes even opposing ones) depending on the task being evaluated.

Maupom at al. [14]: As part of the eRisk2018 shared task on depression, which consists in the early assessment of depression risk in social media users, they implemented a system based on the topic extraction algorithm, Latent Dirichlet Allocation and simple neural networks. The system uses uni-gram, bi-gram and tri-gram frequency to extract 30 latent topics in an unsupervised manner. Once transformed onto this feature space, the users are given a diagnostic probability by a Multilayer Perceptron. Finally, a decision algorithm based on an absolute threshold of probability, which shrinks with time, classifies every user.

Resnik et al. [44]: Topic models can yield insight into how depressed and non-depressed individuals use language differently. In this paper, they explored the use of supervised topic models in the analysis of linguistic signal for detecting depression, providing promising results using several models.

2.3. Machine Learning.

Inna Pirina et al. [26]: This paper presents a set of classification experiments for identifying depression in posts gathered from social media platforms. In addition to the data gathered previously by other researchers, they collected additional data from the social media platform Reddit. Their experiments show promising results for identifying depression from social media texts. More importantly, however, they had shown that the choice of corpora is crucial in identifying depression and can lead to misleading conclusions in case of poor choice of data.

H. Almeida, A. Briand, and M.-J. Meurs [24]: This paper presents the systems developed by the UQAM team for the CLEF eRisk Pilot Task 2017. The goal was to predict as early as possible the risk of mental health issues from user-generated content in social media. Several approaches based on supervised learning and information retrieval methods were used to estimate the risk of depression for a user given the content of its posts in reddit. Among the five systems evaluated, the experiments show that combining information retrieval and machine learning approaches gives the best results.

M. Trotzek, S. Koitka, and C. M. Friedrich [40]: Depression is ranked as the largest contributor to global disability and is also a major reason for suicide. Still, many individuals suffering from forms of depression are not treated for various reasons. Previous studies have shown that depression also has an effect on language usage and that many depressed individuals use social media platforms or the internet in general to get information or discuss their problems.

This paper addresses the early detection of depression using machine learning models based on messages on a social platform. In particular, a convolutional neural network based on different word embedding's is evaluated and compared to a classification based on user-level linguistic metadata. An ensemble of both approaches is shown to achieve state-of-the-art results in a current early detection task.

Table 1 provides an overview of research papers that have mined textual data for insights into the relationship between depression and language usage. The first column describes the source of data used in social media, Live journal and blogposts. The features are based on a linguistic analysis of the text followed by different machine.

Data Source	Features	Methods	Best Results	References
Reddit	BOW, UMLS	Ada Boost & SVM	F1-score 0.75 & 0.98	Sayanta Paul et al. [51]
	N-grams, Topic Modeling	MLP	F1-score 0.64	Maupom et al. [14]
	LIWC+N-grams	SVM	Acc. 0.82	JT Wolohan et al. [30]
Tweets	LIWC, sentiment,	RF	AUC 0.87	Reece et al. [3]
	time series LIWC, n-grams, topic modeling, sentiment	LR	AUC 0.85	Preotiuc-Pietro et al. [15]
	N-grams, topic modeling	SVM	Acc. 0.69	Tsugawa et al. [54]
	N-grams	Neural Network	AUC 0.76	Benton et al. [2]
Facebook	LIWC, N-grams, topic modeling	LR	AUC 0.72	Eichstaedt, J. C. et al. [27]
Live Journal	LIWC, topics modeling, mood tags	Regression Models	Acc. 0.93	Nguyen et al. [57]
Blog posts	TFIDF, topic modeling, BOW	CNN	Acc. 0.78	Tyshchenko et al. [64]

Table 1. Overview of recent depression detection studies in social media learning approaches to obtain the best performance for the particular experiment conducted by the authors mentioned in the last column.

3. METHODOLOGIES

3.1. Architecture.

The architecture of the proposed system gives the basic idea about the tasks that are to be performed. The architecture consists of the different stages that are to be considered.

Firstly, we need to have a dataset. The data is then pre-processed and different feature extraction methods like N-grams, LIWC and LDA are applied on that data. After extracting the features, the machine learning classifiers are trained. Results can be obtained by performing the testing. Once the results are obtained, they can be analyzed accordingly.

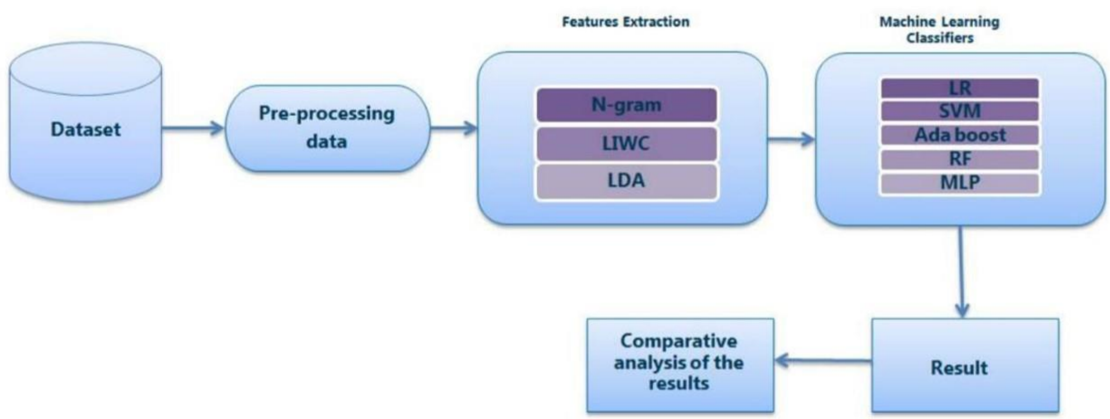

Figure 1. Depression Detection Framework

3.2. Modules to be Implemented.

The modules to be implemented are

1. Importing Dataset
2. Data Pre-Processing
3. Features Extraction
4. Text Classification
5. Metrics Evaluation

3.3. **Module Description.**

3.3.1. ***Importing Dataset.***

Importing the dataset is the initial task for any project that is related to data analysis. After importing the dataset, we need to check the dimensionality of the dataset. The number of dimensions needs to be reduced if possible by removing any unnecessary attributes that are present in the dataset. Also we need to ensure that the dataset does not contain any missing values. Since we are dealing with the text processing and considering the labelled dataset, only two attributes are important that are the text and the label of the text. So if any missing values are encountered, simply the value of the other attribute is also discarded. Finally, the dataset is converted to a data frame for efficient processing of the data.

3.3.2. ***Data Pre-Processing.***

Data pre-processing is an important step in performing the text classification.
The data can be classified only after obtaining the clean text.

- Removal of URL's, Mentions and Punctuations.

  The social media posts usually contain URL's and mentions. These URL's are the web addresses which act as a reference to a web resource. An example URL is ”https://www.google.com”. Mentions are used to refer to a particular social media user. Mentions start with the character”@”. Punctuations are the marks which are used to separate the sentences and their elements to clarify meaning. Some of the punctuation marks are comma, full stop, apostrophe, etc. The clean text should be free from these URL's, mentions and the punctuations.

- Removal of stop words.

  Stops words are the most commonly used words in a language. Stop words should be removed before processing the natural language data. Some of the example stop words in English are a, an, the, i, me, etc.

  The list of stop words can be obtained from the natural language toolkit library for the python language. Stop words in the text can be removed by using that list.

- Stemming.

  Stemming is the process of converting the words into their root form. This step can be performed by using either Stemming or Lemmatization. Lemmatization returns root forms of actual language. For example, consider the word studies. Stemming returns studi, whereas lemmatization returns study. Consider another word better. Stemming returns better, whereas lemmatization returns good. So, for better results first lemmatization is applied and then stemming is performed.

3.3.3. ***Features Extraction.***

In order to explore the language usage of the users, n-grams modeling, liwc and topic modeling will be used. These methods will be helpful in extracting the features from the text.

N-gram Modeling:

N-gram modeling is widely used in text mining and NLP as a feature for depression detection [15],[20]. It is used to calculate the probability of co-occurrence of each input sentence as n-grams. Basically these n-grams can be viewed as unigrams, bigrams, trigrams and so on. Unigrams are simply the individual words that appear in the text. Bigrams are the set of two words that appear in a sequence in the text.

These unigrams and bigrams can be used as the features for classifying the text. Since they are in the string format, they cannot be used directly. They can be used for classification if they are converted to numerical values. In order to convert unigrams and bigrams into numerical values, TF-IDF is used [22].

TF-IDF means Term Frequency Inverse Document Frequency. By using this statistic, the less informative words which occur frequently can be removed so that the more informative words that occur in small fraction can be pertained. If TF-IDF value is high, the importance of the n-gram is high and vice versa. In addition, for the bigrams PMI can be used to filter the infrequent bigrams. PMI means Pointwise Mutual Information [10]. It is defined as the ratio of probability of the co-occurrence of words to the product of probabilities of the individual occurrences of the word. If PMI is high, there is a high chance for the co-occurrence of the words and vice versa.

LIWC:

LIWC, or the Linguistic Inquiry and Word Count dictionary, is widely used in computational linguistics as a source of features for psychological and psycholinguistic analysis. It works as a baseline measure with a set of words and a behavioral link. It is often presented in several mental health projects [9],[11],[18],[23],[27],[28],[39],[54].

To accomplish this experiment, the features are extracted from different features in view of psycholinguistic measures and change every depressive and non-depressive post into numerical values. This way scores are obtained for three higher level categories considering standard linguistic dimensions, psychological processes and personal concerns.

The standard linguistic processes are one of the largest parts of the LIWC psycholinguistic vocabulary package [62]. Some of the standard linguistic features are verbs, pronouns, prepositions, articles, adverbs, conjunctions, negations, etc. Some of the sub-categories of the psychological processes are effective processes, biological processes, social processes. Effective processes can be further divided into features related to anxiety, sadness, positive and negative emotion. Biological processes are divided into features like body, health, ingestion, etc. Personal concerns [37] can be viewed as the features related to work, home, money, religion, etc.

The text is converted into the numerical values based on the different categories of words that are present in the text [42]. The LIWC2015 dictionary [29] is implemented in order to extract the linguistic features from the text.

Topic Modeling:

Topic modeling is an effective tool in computational linguistics to reduce the input of textual data feature space to a fixed number of topics [43]. Topic modeling means generating topics from the given text in an unsupervised fashion. It automatically generates the group of non-labeled words. The choice of the words is based on the probability. Each post is viewed as a document which deals with different topics. In order to derive topic distributions for each post in the dataset, LDA is used. LDA is an acronym for Latent Dirichlet Allocation, which is a probabilistic generative model for discretion of data collections. It is helpful in identifying the underlying topic structures [13]. LDA implementation is provided by the Mallet toolkit [5].

The number of features is based on the number of topics chosen to generate in each document. While generating the topics, the words that appear in more than 10 posts are considered. Also, the words that are present in more than half of the posts can be eliminated.

3.3.4. ***Text Classification Techniques.***

To estimate the presence of depression, classifying approaches are employed to estimate the likelihood of depression within the users. The proposed framework is developed using Logistic Regression [1], [50], Support Vector Machine [61], Random Forest [8], Adaptive Boosting [46], [63] and Multi-Layer Perceptron [21], [25].

Logistic Regression:

Logistic Regression estimates the probability that a data item belongs to a particular category. Here, we will use the logistic function or sigmoid function to model the probabilities. The sigmoid function gives output between 0 and 1 for all the input values.

$$p(X) = \frac{e^{(b_0+b_1*X)}}{1+e^{(b_0+b_1*X)}} \tag{3.1}$$

The logistic function will always produce an S-shaped curve regardless of the input variable *X*. The above equation can be reframed as

$$\log\left(\frac{p(X)}{(1-p(X))}\right) = b_0 + b_1 * X \tag{3.2}$$

24

The quantity $\frac{p(X)}{(1-p(X)}$ is called the odds ratio, and can take any value between 0 and ∞. Values of the odds ratio very close to 0 and ∞ indicate very low and very high probabilities of p(X) respectively. The quantity $\log(\frac{p(X)}{(1-p(X))})$ is called the logit.
The coefficients b_0 and b_1 are unknown, and must be estimated based on the available training data. For this purpose, we can use the maximum likelihood, a powerful statistical technique. Maximum likelihood is a very good approach for to fit non-linear models.
The mathematical function for the likelihood can be given as:

$$l\,(b_0, b_1) = p(X) * (1 - p(X)) \tag{3.3}$$

The estimates b_0 and b_1 are chosen to maximize the likelihood function. Once the coefficients have been estimated, we can calculate the probability $p(X)$, which yields a value 1 for depressed individuals and 0 for non-depressed individuals.

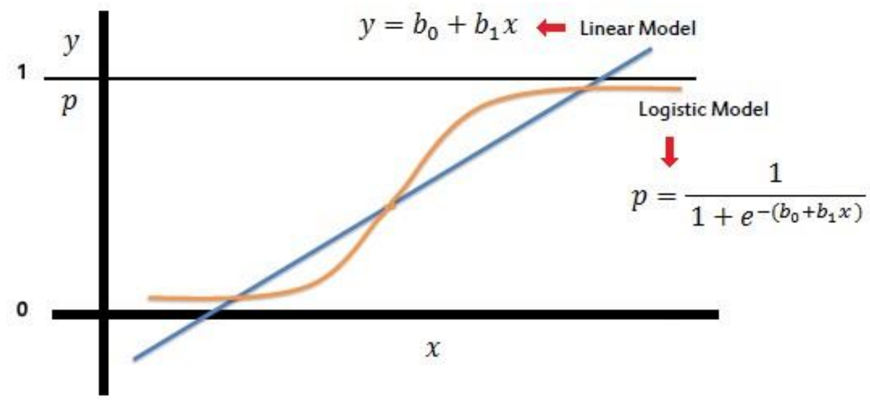

Figure 2. Comparison of linear model and logistic model

Support Vector Machine:

The main objective of SVM is to select a hyperplane with the maximum possible margin between the support vectors in the given dataset. Support vectors are the data points which are closest to the hyperplane. Hyperplane is a decision plane that separates between a set of objects having different class memberships. Margin is the perpendicular distance between the hyperplane and the support vectors.

Generate the hyperplanes which segregates the classes in the best way. Then, identify and select the right hyperplane that classifies the data points correctly and has the good margin.

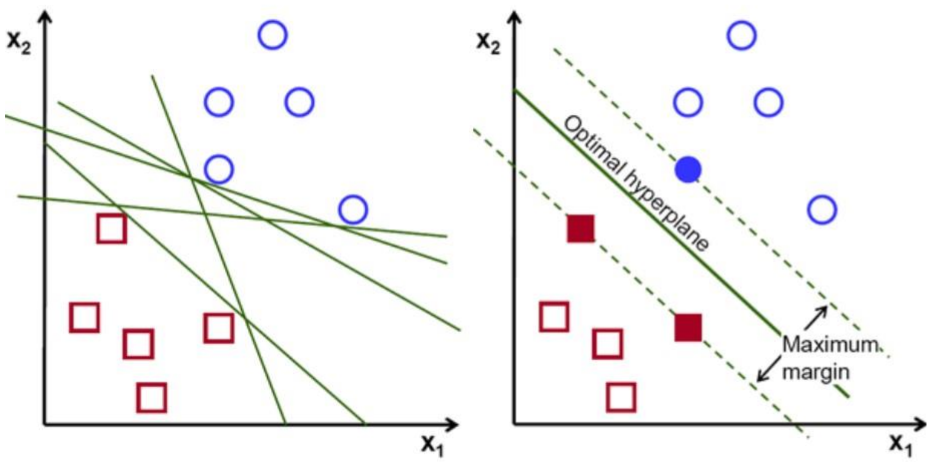

Figure 3. SVM-Finding the optimal hyper line

The SVM algorithm is implemented by using a kernel. A kernel transforms an input data space into the required form. SVM uses a technique called kernel trick, which takes a low-dimensional input space and transforms it into a higher dimensional space. It is most useful in the non-linear separation problem.

A linear kernel can be used as normal dot product of any two given vectors.

$$K(x,x_i) = sum(x * x_i) \tag{3.4}$$

where, x is the input and x_i is the support vector.

A polynomial kernel is a more generalized form of the linear kernel. The polynomial kernel can distinguish curved or nonlinear input space.

$$K(x,x_i) = 1 + sum(x * x_i)^d \tag{3.5}$$

where, d is the degree of the polynomial.

A Radial Basis Function(RBF) kernel is a popular kernel function used in SVM classification. RBF can map an input space in infinite dimensional space.

$$K(x, x_i) = exp(-gamma * sum(x - x_i^2)) \tag{3.6}$$

Here, *gamma* is a parameter whose value ranges from 0 to 1. Lower gamma value will loosely fit the training dataset, whereas higher value exactly fits the training dataset resulting in over-fitting. In other words, low value of gamma considers only nearby points in calculating the separating line, while the higher gamma value considers all the data points.

Also, we have the regularization parameter, called the penalty parameter, which represents misclassification or error term. It tells the SVM how much the error is bearable on each training example. A smaller regularization value creates a small-margin hyperplane; a larger value creates a larger-margin hyperplane.

Finally, we need to tune the parameters such as the kernel function, gamma and regularization to the appropriate values corresponding to the characteristics of the input dataset.

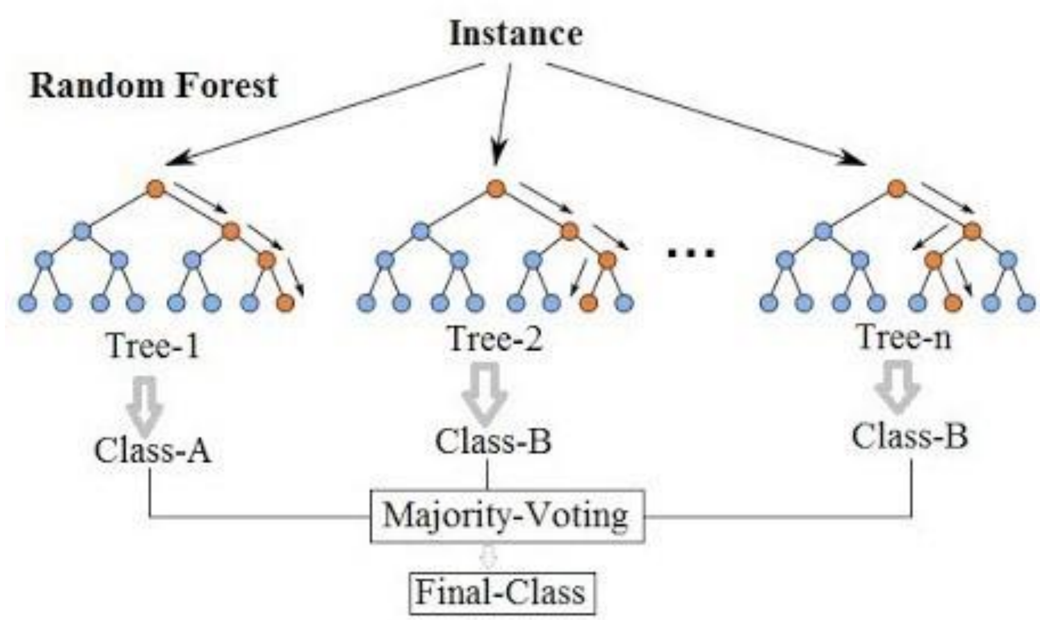

Figure 4. Random Forest Classification Algorithm

Random Forest:

Random Forest (RF) is an ensemble of decision tree classifiers trained with the bagging method where a combination of learning models increases the overall result. The training algorithm of random forest is called bootstrap algorithm or bagging technique.

The Steps are given below.

(1) Load the data where it consists of m features representing the behavior of the dataset.

(2) It selects n features randomly from m features, i.e. to create random samples with replacement from the original data. About 1/3 rows of the original data are left out known as the Out of Bag(OOB) samples. The model trains the new samples. OOB samples are used to determine the unbiased OOB error.

(3) Calculate the node d using the best split. Split the node into sub-nodes.

(4) Repeat the steps, to find n number of trees.

(5) Calculate the total number of votes of each tree for predicting the target. The highest voted class is the final prediction of the random forest.

Adaptive Boosting:

The concept of adaboost is about correcting the mistakes of the previous classifiers. A meta classifier ensemble combines a large number of classifiers to produce more accurate and robust predictions than the predictions by each individual classifier.
The most common algorithm used with Adaboost are decision trees with one level. Because these trees are so short and only contain one decision for classification, they are often called as decision stumps.
Each instance in the training dataset is weighted. The initial weight is set to:

$$weight(x_i) = \frac{1}{n} \tag{3.7}$$

where, x_i is the i^{th} training instance, and n is the number of training instances.

A decision stump is prepared on the training data using the weighted samples. Since our problem is a binary classification problem, each decision stump makes one decision on one input variable and outputs a +1.0 or -1.0. In our case, +1.0 indicates depression and -1.0 indicates the absence of depression.

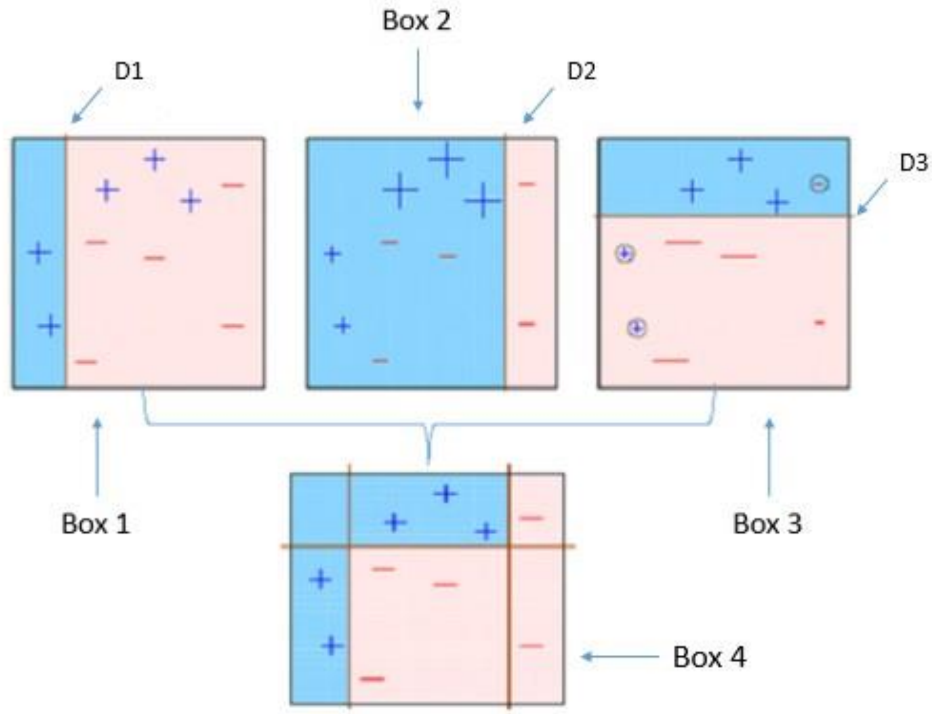

Figure 5. Ada Boost Classification Algorithm

The misclassification rate is calculated for the trained model. It can be calculated as:

$$error = \frac{(correct - N)}{N} \tag{3.8}$$

where, *error* is the misclassification rate, *correct* is the number of training instances predicted correctly by the model, and *N* is the total number of training instances.

We would modify the misclassification rate to use the weighting of the training instances and calculate the weighted sum of the misclassification rate. It is given as

$$error = \frac{sum(w(i) * terror(i))}{sum(w)} \tag{3.9}$$

where, *w* is the weight for training instance *i*, and *terror* is the prediction error for the training instance i which is 1 is if misclassified and 0 if correctly classified.

A stage value is calculated for the trained model which provides a weighting for any predictions that the model makes. The stage value is calculated as follows:

$$stage = \ln(\frac{(1 - error)}{error}) \tag{3.10}$$

where, *stage* is the stage value used to weight predictions from the model, *ln*() is the natural logarithm, and error is the misclassification error for the model.

The effect of the stage weight is that more accurate models have more weight or contribution to the final prediction. The training weights are updated giving more weight to incorrectly predicted instances, and less weight to correctly predicted instances.

The weight of one training instance (w) is updated using:

$$w = w * e^{(stage*terror)} \tag{3.11}$$

where, w is the weight for a specific training instance, *stage* is the misclassification rate, and terror is the prediction error.

Weak models are added sequentially, trained using the weighted training data. The process continues until a pre-set number of learners have been created or no further improvement can be made on the training dataset.

Multi-Layer Perceptron:

Multi-Layer Perceptron is a type of network where multiple layers of a group of perceptron are stacked together to make a model. Generally, perceptron or neuron is a linear function that takes multiple inputs and produce an output.

In the case of MLP, perceptron is a linear model which takes bunch of inputs, multiply them with weights and add a bias term to generate an output.

$$Z = \overrightarrow{w} * X + b \tag{3.12}$$

where, w is the weight, b is the bias.

Bias is like the intercept added in a linear equation. It is an additional parameter in the neural network which is used to adjust the output along with the weighted sum of the inputs to the neuron. Bias is a constant that helps the model to fit best for the given data.

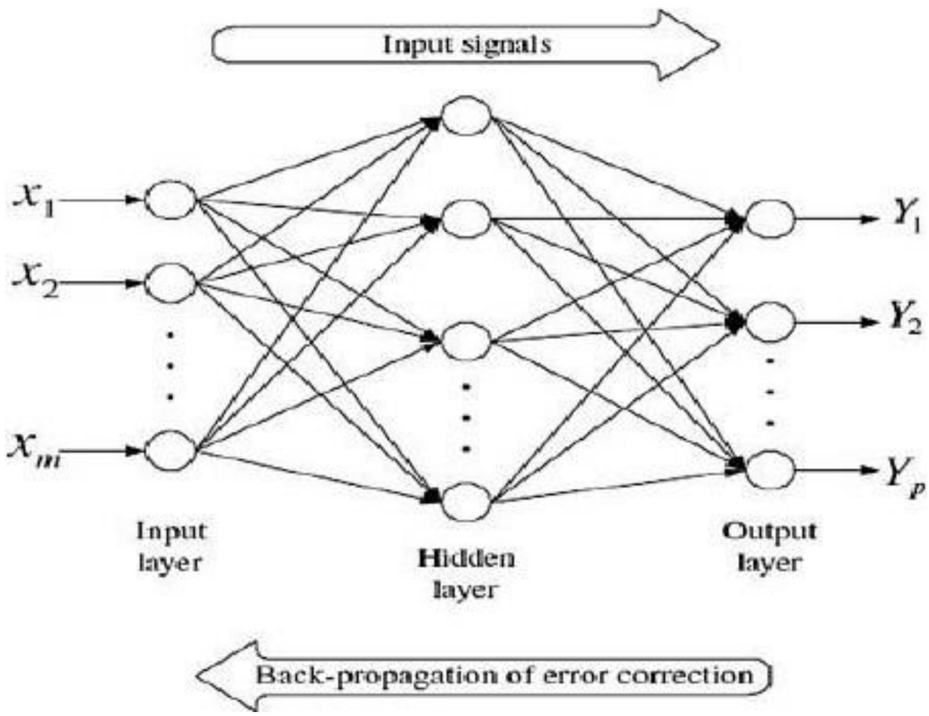

Figure 6. Multi-Layer Perceptron Classification Algorithm

First, we need to specify the model input, the number of the input attributes.
Then, we need to specify the number and the type of the layers. We have many types of the layers. Some of the useful types of layers are:
Dense: It is a fully connected layer and the most common type of layer used on MLP models.
Dropout: It applies dropout to the model, setting a fraction of inputs to zero in an effort to reduce overfitting.
Merge: Combines the inputs from multiple models into a single model. The dense layer can be represented as

#»

$$f(X) = W * X + b \tag{3.13}$$

Each layer consists of a number of nodes and each node has a weight associated with it. Some of the common types of weight initialization are:
Uniform: Weights are initialized to small random values between 0 and 0.05.
Normal: Weights are initialized to small Gaussian random values.
Zero: All weights are initialized to zero.

If we want our MLP model to be flexible and learn non-linear decision boundaries, we need to introduce non-linearity into the network. The non-linearity can be achieved by adding the activation function. Some of the standard neuron activation functions are softmax, rectifier, and sigmoid.
The rectifier function, also known as ReLU (Rectified Linear Unit) outputs the input value directly if it is positive, otherwise, it will output zero.

$$f(X) = max\ (0, X) \tag{3.14}$$

We need to specify a loss function, also called the objective function for evaluation of the model and to navigate the weight space. Some of the commonly used loss functions are:
MSE: For mean squared error.
Binary cross entropy: For binary logarithmic loss.
Categorical cross entropy: For multi-class logarithmic loss.

Every layer should have a feed-forward loop and backpropagation loop. Feedforward loop takes an input and generates output for making a prediction. Backpropagation loop helps in training the model by adjusting weights in the layer to lower the output loss. In backpropagation, weight update is done by using the chain rule and optimized using an optimization algorithm. Some of the commonly used optimizers for this purpose are:
SGD: Stochastic Gradient Descent, with support for momentum.
RMSprop: Adaptive learning rate optimization method.
Adam: Adaptive Moment Estimation, also uses adaptive learning rates.

3.3.5. ***Metrics Evaluation.***

Metrics Evaluation is the final and the important of all the modules that needs to be implemented. This module tests how efficiently all the above modules have been working. The final result of the work can be analyzed by calculating some metrics. These metrics are used to understand and compare the efficiencies of the different models.

3.4. **Dataset Description.**

To identify the depression, two datasets are used. One dataset consists of reddit users posts, the other dataset consists of twitter users posts.

The reddit dataset was created by the Inna Pirina et al. [26]. It consists of the list of depressed and non-depressed user's posts. The data corpus contains 1293 depression-indicative posts and 548 standard posts. The twitter dataset is downloaded from the Kaggles website. It consists of 5235 depression-indicative tweets and 6199 standard tweets.

Table 2. Example posts from reddit dataset

Depressed Posts	Normal Posts
I hate the way I look I've pretty much always hated the way I looked. I just recently got a really bad haircut and I'm considering shaving my head. I'll be ugly no matter what so I guess hair or no hair won't really make a difference. I can't afford plastic surgery or else I would. It sucks being an ugly girl.	Top 15 Inspiring Friendship Quotes A friend is seemed to be a person who takes a great place in someone's life. They share all happy and sad moments in the same manner. To make your friends realize that they have taken an important place in your life. Here are 15 friendship quotes which show your feeling towards your friends.

I don’t need help or pills, I need to die. Nothing like that has or ever will help. No platitudes, nothing is ever going to help. I’m stupid, useless, weird, and have no personality. I’m a waste of space and I HAVE to die. There’s no other option.	Family tree related question, not sure where to post Ok so I have a cousin from my aunt and a guy she used to date. What is the title for that guy’s sister? I consider him a sort of uncle, but what would I call his sister? I’m sure it’s an aunt of some sort, but I need specifics.

The datasets consist of only two attributes. They are text and target. The text attribute contains the posts written by the users. The target attribute has two values namely 0 and 1. 0 indicates standard post whereas 1 indicates the depression-indicative post.

Example tweets from twitter dataset

Depressed Posts	Normal Posts
Argh. I hate my life	Like I’m so proud of myself
RT @ youreWelcome: I think I’m depressed	I am easily satisfied ... with the very best
I hate forcing myself to sleep	Anyways I just felt like showing off how much I love my family
RT @emmataskerr: really miss my old happy self. don’t know who I am anymore	RT @Libras R Us: As a #Libra, I get revenge by living a great life
Just feel so empty and sad. Had enough of everything now. All I ever wanted was to be happy with my family and friends and its ruined	Went shopping with my mama and aunt... God I love my life!!(: wouldn’t change it for the world<333

4.DISCUSSION AND PROPOSED OUTPUT

The main task is to detect depression of each of the users in the chosen data. Execution of the text classifying techniques will be started by using the entire dimension feature space extracted from the dataset. For baseline features, LIWC categories, N-gram probabilities, LDA model and their multiple feature combinations built on the training data will be used.

The aim of combining the distinct NLP techniques is to find out what combination of the features best favors the performance accuracy for depression detection. For estimating the presence of depression within the posts, five classifiers are applied. For implementation of each classifier, Scikit-learn library for the Python language will be used [17].

To evaluate the classification techniques, we apply the evaluation metrics, such as accuracy of estimations (Acc.) and F-score (F1) consisting of precision (P) and recall (R). It relies on a confusion matrix incorporating the information about each test sample prediction outcome. In the evaluation metrics, we find that there is a number of true positive predictions(TP), true negative predictions (TN), false positive predictions (FP) and false negative predictions (FN)[56].

- True Positive is an outcome where the model correctly predicts the positive class.
- True Negative is an outcome where the model correctly predicts the negative class.
- False Positive is an outcome where the model incorrectly predicts the positive class.
- False Negative is an outcome where the model incorrectly predicts the negative class.

Accuracy:

Accuracy is the rate of correct classification. It is the ratio of the number of correctly classified instances to the total number of instances.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{4.1}$$

Precision:

Precision estimates how many positively identified samples are correct.

$$Precision = \frac{TP}{TP + FP} \tag{4.2}$$

Recall:

Recall estimates what proportion of positive samples was correctly identified.

$$Recall = \frac{TP}{TP + FN} \tag{4.3}$$

F1-score:

F1 score is a harmonic average of the precision and recall. The closer both values of precision and recall, the higher the F1 score is.

$$F1 = \frac{2 * precision * recall}{precision + recall} \tag{4.4}$$

Each classified corpus contains an accuracy, F-score, precision and recall result value. The predictive power for a better performance is hidden in proper feature selections and their multiple feature combinations.

Table 3: Results of Classification models

FEATURE	LR				SVM				RF				Ada Boost				MLP			
	Acc.	F1	P	R	Acc.	F1	P	R	Acc.	F1	P	R	Acc.	F1	P	R	Acc.	F1	P	R
LIWC	69%	0.80	0.95	0.69	74%	0.74	0.75	0.72	78%	0.84	0.77	0.95	67%	0.81	0.68	0.99	70%	0.72	0.74	0.71
LDA	77%	0.83	0.82	0.84	72%	0.80	0.75	0.88	68%	0.67	0.68	0.68	66%	0.74	0.61	0.95	75%	0.74	0.75	0.72
unigram	68%	0.80	0.93	0.79	70%	0.81	0.70	0.89	71%	0.81	0.71	0.96	72%	0.81	0.73	0.92	70%	0.81	0.71	0.95
bigram	80%	0.79	0.81	0.78	80%	0.80	0.79	0.82	74%	0.73	0.75	0.72	71%	0.68	0.75	0.63	79%	0.78	0.80	0.76
LIWC+LDA+ unigram	80%	0.84	0.88	0.81	79%	0.81	0.81	0.81	83%	0.84	0.82	0.86	73%	0.72	0.87	0.62	78%	0.81	0.84	0.79
LIWC+LDA+ bigram	89%	0.89	0.89	0.92	90%	0.91	0.89	0.93	85%	0.85	0.83	0.87	79%	0.81	0.72	0.93	91%	0.93	0.90	0.92

[a] Acc. represents accuracy; LR=Logistic Regression, SVM=Support Vector Machine, RF=Random Forest, Ada Boost=Adaptive Boosting, MLP=Multilayer Perceptron

5. CONCLUSION AND FUTURE WORK

The work tries to identify the presence of depression in Reddit and Twitter social media and searched for affective performance increase solutions of depression detection. The work characterizes a closer connection between depression and language usage by applying NLP and text classification techniques.

To measure the signs of depression, we examine the performance of both single feature and combined feature sets using various text classifying methods. For the purpose of classification, five machine learning classifiers are used. They are Logistic Regression, Support Vector Machine, Random Forest, Adaptive Boosting and Multi-Layer Perceptron. The proper outcome of the work can be known only after conducting the experiment and evaluating the metrics. Analysis of the result parameters can be used to evaluate the performance of the work.

REFERENCES

[1] A. Agresti, An Introduction to Categorical Data Analysis. Hoboken, NJ, USA: Wiley, 2018.

[2] A. Benton, M. Mitchell, and D. Hovy. (2017). "Multi-task learning for mental health using social media text." [Online]. Available: https://arxiv.org/abs/1712.03538

[3] A. G. Reece, A. J. Reagan, K. L. Lix, P. S. Dodds, C. M. Danforth, andE. J. Langer, "Forecasting the onset and course of mental illness withtwitter data," Sci. Rep., vol. 7, no. 1, p. 13006, Oct. 2017.

[4] A. Hern´andez-Castan˜eda and H. Calvo, "Deceptive text detection using continuous semantic´ space models," Intell. Data Anal., vol. 21, no. 3,pp. 679–695, Jan. 2017.

[5] A. K. McCallum. (2002). MALLET: A Machine Learning for Language Toolkit. [Online]. Available: http://mallet.cs.umass.edu

[6] A. Li, X. Huang, B. Hao, and B. O'Dea, H. Christensen, T. Zhu, and A. F. Jorm, "Attitudes towards suicide attempts broadcast on social media: An exploratory study of chinese microblogs," PeerJ, vol. 8, no. 3, p. e1209,Sep. 2015.

[7] A. T. Beck, Depression: Clinical, Experim., Theoretical Aspects. Philadelphia, PA, USA: Univ. Pennsylvania Press, 1967.

[8] B. Xu, Y. Ye, and L. Nie, "An improved random forest classifier for image classification," in Proc. IEEE Int. Conf. Inf. Autom., Jun. 2012, pp. 795–800.

[9] C. Chung and J. W. Pennebaker, "The psychological functions of functionwords," Social Commun., vol. 1, pp. 343–359, Sep. 2007.

[10] C. D. Manning and H. Schu¨tze, Foundations of Statistical Natural Language Processing, vol. 999. Cambridge, MA, USA: MIT Press, 1999.

[11] C. M. Homan, N. Lu, X. Tu, M. C. Lytle, and V. Silenzio, "Social structure and depression in trevorspace," in Proc. 17th ACM Conf. Comput.Supported Cooperat. Work Social Comput., Feb. 2014, pp. 615–625.

[12] D. E. Losada, F. Crestani, and J. Parapar, "eRISK 2017: CLEF lab on earlyrisk prediction on the internet: Experimental foundations," in Proc. Int.Conf. Cross-Lang. Eval. Eur. Lang., 2017, pp. 346–360.

[13] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent Dirichlet allocation,"J. Mach. Learn. Res., vol. 3, pp. 993–1022, Mar. 2003.

[14] D. Maupom´es and M. Meurs, "Using topic extraction on social media content for the early detection of depression," in Proc. CLEF (Working Notes), vol. 2125, Sep. 2018. [Online]. Available: https://CEUR-WS.org

[15] D. Preotiuc-Pietro et al., "The role of personality, age, and gender in tweeting about mental illness," in Proc. 2nd Workshop Comput. Linguistics Clin. Psychol. Linguistic Signal Clin. Reality, 2015, pp. 21–30.

[16] E. Durkheim and A. Suicide, A Study in Sociology. Abingdon, U.K.:Routledge, 1952.

[17] F. Pedregosa et al., “Scikit-learn: Machine learning in Python,” J. Mach.Learn. Res., vol. 12, pp. 2825–2830, Oct. 2011.

[18] G. Coppersmith, M. Dredze, and C. Harman, “Quantifying mental health signals in twitter,” in Proc. Workshop Comput. Linguistics Clin. Psychol.Linguistic Signal Clin. Reality, 2014, pp. 51–60.

[19] G. Coppersmith, M. Dredze, C. Harman, and K. Hollingshead, “From ADHD to SAD: Analyzing the language of mental health on twitter through self-reported diagnoses,” in Proc. 2nd Workshop Comput. Linguistics Clin. Psychol. Linguistic Signal Clin. Reality, 2015, pp. 1–10.

[20] G. Coppersmith, M. Dredze, C. Harman, K. Hollingshead, and M. Mitchell, “Clpsych 2015 shared task: Depression and ptsd on twitter,” in Proc. 2nd Workshop Comput. Linguistics Clin. Psychol. Linguistic Signal Clin. Reality, 2015, pp. 31–39.

[21] G. Cybenko, “Approximation by superpositions of a sigmoidal function,” Math. Control, Signals Syst., vol. 2, no. 4, pp. 303–314, 1989.

[22] G. Salton and C. Buckley, “Term-weighting approaches in automatic textretrieval,” Inf. Process. Manage., vol. 24, no. 5, pp. 513–523, 1988.

[23] H. A. Schwartz et al., “Towards assessing changes in degree of depression through facebook,” in Proc. Workshop Comput. Linguistics Clin. Psychol. Linguistic Signal Clin. Reality, 2014, pp. 118–125.

[24] H. Almeida, A. Briand, and M.-J. Meurs, “Detecting early risk of depression from social media user-generated content,” in Proc. CLEF, 2017,pp. 1–10.

[25] H. Ramchoun, M. A. J. Idrissi, Y. Ghanou, and M. Ettaouil, “Multilayer perceptron: Architecture optimization and training,” IJIMAI, vol. 4, no. 1, pp. 26–30, 2016.

[26] I. Pirina and C. C. C¸olt´ekin, “Identifying depression on reddit: The effect of training data,” in Proc. EMNLP Workshop, 2018, pp. 9–12.

[27] J. C. Eichstaedt et al., “Facebook language predicts depression in medical records,” Proc. Nat. Acad. Sci., vol. c, no. 44, pp. 11203–11208, Oct. 2018.

[28] J. S. Buyukdura, S. M. McClintock, and P. E. Croarkin, “Psychomotor retardation in depression: Biological underpinnings, measurement, and treatment,” Progr. Neuro-Psychopharmacol. Biol. Psychiatry, vol. 35,no.2, pp. 395–409, Mar. 2011.

[29] J. W. Pennebaker, R. J. Booth, R. L. Boyd, and M. E. Francis, “Linguistic Inquiry and Word Count: LIWC2015,” Pennebaker Conglomerates, Austin, TX, USA, 2015. [Online]. Available: www.LIWC.net

[30] J. Wolohan, M. Hiraga, A. Mukherjee, Z. A. Sayyed, and M. Millard, “Detecting linguistic traces of depression in topic-restricted text: Attending to self-stigmatized depression with nlp,” in Proc. 1st Int. Workshop Lang. Cognition Comput. Models, 2018, pp. 11–21.

[31] J. Zinken, K. Zinken, J. C. Wilson, L. Butler, and T. Skinner, “Analysis of syntax and word use to predict successful participation in guided self-help for anxiety and depression,” Psychiatry Res., vol. 179, no. 2, pp. 181–186,Sep. 2010.

[32] L. van der Maaten and G. E. Hinton, “Visualizing data using T-SNE,”J. Mach. Learn. Res., vol. 9, pp. 2579–2605, Nov. 2008.

[33] M. A. Moreno et al., “Feeling bad on Facebook: Depression disclosures by college students on a social networking site,” Depression Anxiety, vol. 28,no. 6, pp. 447–455, Jun. 2011.

[34] M. D. Choudhury, E. Kiciman, M. Dredze, G. Coppersmith, and M. Kumar,“Discovering shifts to suicidal ideation from mental health content insocial media,” in Proc. CHI Conf. Hum. Factors Comput. Syst., May 2016,pp. 2098–2110.

[35] M. De Choudhury, S. Counts, and E. Horvitz, “Predicting postpartum changes in emotion and behavior via social media,” in Proc. SIGCHI Conf.Hum. Factors Comput. Syst., Apr. 2013, pp. 3267–3276.

[36] M. J. Friedrich, “Depression is the leading cause of disability around the world,” JAMA, vol. 317, no. 15, p. 1517, Apr. 2017.

[37] M. M. Tadesse, H. Lin, B. Xu, and L. Yang, “Personality predictions based on user behavior on the facebook social media platform,” IEEE Access, vol. 6, pp. 61959–61969, 2018.

[38] M. Nadeem. (2016). “Identifying depression on twitter.”[Online]. Available: https://arxiv.org/abs/1607.07384

[39] M. Park, D. W. McDonald, and M. Cha, “Perception differences between the depressed and non-depressed users in twitter,” in Proc. 7th Int. AAAIConf. Weblogs Social Media, Jun. 2013, pp. 45–65.

[40] M. Trotzek, S. Koitka, and C. M. Friedrich, “Utilizing neural networks and linguistic metadata for early detection of depression indications in text sequences,” IEEE Trans. Knowl. Data Eng., to be published.

[41] N. Masuda, I. Kurahashi, and H. Onari, “Suicide ideation of individuals in online social networks,” PloS One, vol. 8, no. 4, Apr. 2013, Art. no. e2262.

[42] N. Ram´ırez-Esparza, C. K. Chung, E. Kacewicz, and J. W. Pennebaker, “The psychology of word use in depression forums in english and in spanish: Texting two text analytic approaches,” in Proc. ICWSM, Mar. 2008,pp. 1–10

[43] P. Resnik, A. Garron, and R. Resnik, “Using topic modeling to improve prediction of neuroticism and depression in college students,” in Proc. Conf. Empirical Methods Natural Lang. Process., Sep. 2013,pp. 1348–1353.

[44] P. Resnik, W. Armstrong, L. Claudino, T. Nguyen, V.-A. Nguyen, and J. Boyd-Graber, “Beyond LDA: Exploring supervised topic modeling for depression-related language in twitter,” in Proc. 2nd Workshop Comput. Linguistics Clin. Psychol. Linguistic Signal Clin. Reality, 2015, pp. 99–107.

[45] R. A. Calvo, D. N. Milne, M. S. Hussain, and H. Christensen, “Natural language processing in mental health applications using non-clinical texts,”Natural Language Eng., vol. 23, no. 5, pp. 649–685, 2017.

[46] R. E. Schapire, Y. Singer, and A. Singhal, “Boosting and rocchio applied to text filtering,” SIGIR, vol. 98, pp. 215–223, Aug. 1998.

[47] S. Bagroy, P. Kumaraguru, and M. De Choudhury, “A social media based index of mental well-being in college campuses,” in Proc. CHI Conf. Hum.Factors Comput. Syst., May 2017, pp. 1634–1646.

[48] S. C. Guntuku, D. B. Yaden, M. L. Kern, L. H. Ungar, and J. C. Eichstaedt, “Detecting depression and mental illness on social media: An integrative review,” Current Opinion Behav. Sci., vol. 18, pp. 43–49, Dec. 2017.

[49] S. Freud, The Psychopathology of Everyday Life. London, U.K.: Hogarth,1901.

[50] S. L. Gortmaker, “Theory and methods–applied logistic regression by david w. hosmer jr and stanley lemeshow,” Contemp. Sociol., vol. 23, no. 1,p. 159, Jun. 1994.

[51] S. Paul, S. K. Jandhyala, and T. Basu, “Early detection of signs of anorexia and depression over social media using effective machine learning frameworks,” in Proc. CLEF, Aug. 2018, pp. 19. [52] S. Rude, E.-M. Gortner, and J. Pennebaker, “Language use of depressed and depressionvulnerable college students,” Cognition Emotion, vol. 18,no. 8, pp. 1121–1133, Dec. 2004.

[53] S. Tsugawa, Y. Kikuchi, F. Kishino, K. Nakajima, Y. Itoh, and H. Ohsaki, “Recognizing depression from twitter activity,” in Proc. 33rd Annu. ACM Conf. Hum. Factors Comput. Syst., Apr. 2015, pp. 3187–3196.

[54] S. Tsugawa, Y. Kikuchi, F. Kishino, K. Nakajima, Y. Itoh, and H. Ohsaki,“Recognizing depression from twitter activity,” in Proc. CHI, Apr. 2015,pp. 3187–3196.

[55] S. W. Stirman and J. W. Pennebaker, “Word use in the poetry of suicidal and non-suicidal poets,” Psychosomatic Med., vol. 63, no. 4, pp. 517–522,Jul. 2001.

[56] T. Basu and C. Murthy, “A feature selection method for improved document classification,” in Advanced Data Mining and Applications. New York, NY, USA: Springer, 2012, pp. 296–305.

[57] T. Nguyen, D. Phung, B. Dao, S. Venkatesh, and M. Berk,“Affective and content analysis of online depression communities,” IEEE Trans. Affect.Comput., vol. 5, no. 3, pp. 217–226, Jul. 2014. [58] T. Pyszczynski and J. Greenberg, “Self-regulatory perseveration and thedepressive self-focusing style: A self-awareness theory of reactive depression,” Psychol. Bull., vol. 102, no. 1, p. 122, Jul. 1987.

[59] T. S. Portal Statistics and Studies. (2019). Social Media Usage Worldwide.[Online]. Available: https://www.statista.com/statistics/272014/global-social-networks-ranked-by-number-of-users/

[60] W. H. Organization. (2017). Depression and Other Common Mental Disorders: Global Health Estimates. Geneva: World Health Organization. [Online]. Available:

http://www.who.int/en/newsroom/fact-sheets/detail/depression

[61] W. S. Noble, “What is a support vector machine?,” Nature Biotechnol., vol. 24, no. 12, p. 1565, May 2006.

[62] X. Huang, T. Liu, A. Li, Z. Chen, and T.Zhu, “Using linguistic features to estimate suicide probability of chinese microblog users,” in Proc. HCC, Sep. 2014, pp. 145–156.

[63] Y. Freund, R. Schapire, and N. Abe, “A short introduction to boosting,” J.-Jpn. Soc. Artif. Intell., vol. 14, nos. 771–780, p. 1612, 1999.

[64] Y. Tyshchenko, “Depression and anxiety detection from blog posts data,” Nature Precis. Sci., Inst. Comput. Sci., Univ. Tartu, Tartu, Estonia, 2018.