International Journal of Engineering Applied Sciences and Technology, 2023
Vol. 8, Issue 02, ISSN No. 2455-2143, Pages 16-26
Published Online June 2023 in IJEAST (http://www.ijeast.com)

DEPRESSION DETECTION MODEL USINGWORD AND SENTENCE EMBEDDING WITH DIFFERENT CLASSIFIERS

KM Ankita Rai, Riya Goyal, and Monika Choudhary
Department of Computer Science and Engineering,
Indira Gandhi Delhi Technical University for Women, Delhi-110006, India

Abstract: **Determining depression is hugely important because it affects people's wellness. The common practice of diagnosing depression usually involves a physical session with a practitioner, which reduces the possibility of finding patients who are depressed. Instead, they investigate the possibility of detecting depression solely from textual data based on the posts people post on social media profiles. This research work consists of how Reddit users' text messages can be processed to perform a classification task to identify depression. In order to detect depression, we used Word and Sentence Embeddings. Different embeddings are tested with algorithms such as SVM classifiers, KNN classifiers, and decision tree classifiers, and the results are compared to those from various datasets. Evaluation metrics such as accuracy and precision are used to evaluate the performance of classifiers. USE embedding outperformed Word embeddings with a maximum accuracy of 92%, precision of 95%, recall of 96%, and F1 score of 92%. Results from the suggested method are compared to those from other methods and are found to be satisfactory.**

Keywords: **Depression, Word Embedding, Sentence Embedding, USE.**

I. INTRODUCTION

These days, depression is a widespread mental illness, but still, it is untreated and uncured. There are many symptoms of depression, i.e., depressed mood, lack of enjoyment or interest in activities, thinking about suicide, or feeling lonely or anxious. According to WHO, 5% of adults experience depression [2].

In this online world, social networking created a space for people to meet and communicate with each other. Our daily lives now include social networks like Twitter, Facebook, and Reddit, as a way to communicate our thoughts, emotions, and emotional status. According to one study, we can conclude that social media is one of the reasons for causing depression, anxiety, and psychological distress among adults and teenagers [3]. In most cases, depression also leads to suicide because they lose their trust in everyone including their family, friends, relatives, and doctors impact on them is very negative.

People's online activities gave researchers the chance to research depression detection at an early stage and find out the possible methods and algorithms for it. Using different types of embeddings such as word embedding and sentence embedding and different classifiers, we can achieve high performance [4]. Besides this, our techniques were put into practice utilizing openly accessible resources, making it easier to examine our findings [5].

In this work, we use two datasets for depression detection which are available publicly, i.e., a Reddit dataset [6] and a suicide depression detection dataset which is one of the subreddit on the Reddit platform [7]. In this study, we take into account how Reddit users' text messages can be processed to perform a classification task to identify depression.

The following could serve as a summary of both the important contributions and the suggested technique.

- Word embedding and Sentence embedding-based feature engineering: We provide a technique for using feature engineering based on word and sentence embedding for the classification of depression with a focus on identifying online users who are depressed based on their posts. In the selected dataset, we have used our method for diagnosing depression.
- Evaluation of performance: We evaluate and discuss the performance of several machine learning models typically used in tasks including sentiment classification, particularly to detect depression. Additionally, we have enhanced our research to include several words and sentence embeddings as well as hyper parameter tuning.

The following sections of this paper will discuss previously published studies, describe the research dataset we selected, present methodology and implementation for the depression detection problem, outline the study's purpose, look at the results of the models we used on our dataset, and then draw a conclusion and discuss further research.

II. RELATED WORK

Social media have gained enormous popularity over the past ten years, and billions of users now utilize them to post about their lives and thoughts while they are on the go. As a result, to monitor depression in individuals, researchers started looking at their online behavior. Table 1 is a list of their work, datasets, and algorithms that were reviewed.

By analyzing text conversations, [8] classify Reddit users in order to find depressive persons and also they compare their

International Journal of Engineering Applied Sciences and Technology, 2023
Vol. 8, Issue 02, ISSN No. 2455-2143, Pages 16-26
Published Online June 2023 in IJEAST (http://www.ijeast.com)

findings to the performance of CLEF/eRsik 2017 participants, we examine different baseline features in addition to more complicated features.

They examined the advantages of incorporating data from online social networks in clinical investigations of depression [9]. They looked at how language is used to describe negative emotions using real-time user mood data from the social media site Twitter.

In this work, [10] made an attempt to explain how web users talk about topics related to depression from the perspective of social media and linguistic patterns found in user chats.

Table 1. Previous research in the detection of depression via social media

Ref. No.	Approach	Dataset	Models
[11]	To conduct depression analysis on Mandarin text data, the BERT model is suggested.	WU3D Dataset	BERT model
[12]	Their research involves looking at Reddit users' postings to find any patterns that might indicate how depressed the relevant internet users are.	Reddit users dataset	SVM, LR, Random Forest, Ada Boost, MLP
[13]	Their research explored various ML-based strategies for the early diagnosis of MDDs using data from social media networks.	Reddit dataset	RF classifier
[14]	In their research, an effective method for locating texts describing one's self-perceived depressive symptoms is proposed, based on RNN built on LSTM.	Dataset froma public Norwegian information website: ung.no.	LR, Decision Trees, SVM, ANN, CNN, DBN
[15]	For determining the severity of depression among Twitter users, they have presented a novel short text clustering conceptual model.	Twitter Data	Novel short text clustering conceptual model different Frameworks are used.
[16]	Their research provides an updated dataset compiled using the most recent PHQ-9 underwriting phrases. Additionally, with textual entailment, the model can tell from tweets used in the training phase if a tweet is entailed or not, in which case it will adhere to the same class.	Twitter Data	GRU, Bi-GRU, LSTM, Bi-LSTM

International Journal of Engineering Applied Sciences and Technology, 2023
Vol. 8, Issue 02, ISSN No. 2455-2143, Pages 16-26
Published Online June 2023 in IJEAST (http://www.ijeast.com)

[17]	The goal of their study is to build a predictive model that can classify tweets as normal or depressed after being trained on the training corpus.	Twitter Dataset	Bi-LSTM and different ML and DL models.
[18]	In particular, when those communications do not explicitly contain particular terms like "depression" or "diagnostic," their study intends to discover whether ML may be used to detect indicators of depression in social media users by examining their social media posts.	Twitter Dataset, Facebook, Reddit, and an electronic diary	Adaptive Boosting,LR, MLP, Decision Tree, SVM, RF, Gradient Boosting, Bagging Predictors, LSVM
[1]	To automatically choose the indication posts from past user posts, they suggest using the reinforcement learning approach.	Twitter	NB, multiple social networking learning (MSNL) , Wasserstein dictionary learning (WDL), and multimodal depressive dictionary learning (MDL), CNN, LSTM
[19]	To predict anxious depression in real-time tweets, a novel model called the AD prediction model is put forward in this study.	Twitter	Gradient Boosting and Random Forest, Multinomial Naïve Bayes, Ensemble vote classifier.

III. METHODOLOGY

This research attempts to enhance the diagnosis of depression using textual data by combining several classifiers with Natural Language Processing (NLP) techniques. Such a classifier might be used as an initial screening tool before diagnosing depression. The objectives of this research are:

- Implementing a model which is used to detect and classify depression.
- Increase proficiency using machine learning algorithms to detect depression.

3.1 Dataset Description

Two sets of textual data were used to evaluate the effectiveness of Depression detection. The first dataset that uses to represent the Reddit data set. The acquisition of data for the detection model is essential for accurately detecting signs of mental health disorders. As an illustration, information from social media. There is a great amount of data available to the scientific community. from Facebook, Twitter, and Reddit to research the daily concerns and pressures of individuals throughout the world. The Reddit data set used in this study was obtained from a publicly accessible website that hosted 187,444 posts divided into five categories with both stressful and non-stressful content in each. 715 test data points (52.4% classified stressful) and 2,838 training data points (51.6%) were used to partition this data [6]. due to the relative nature of our data. Analysis of the dataset's content and model performance, which sheds light on the stress detection issue. The second dataset uses is suicide and depression detection. The dataset is a compilation of postings from the Reddit platform's "Suicide Watch" and "depression" subreddits. The posts made to "Suicide Watch" from its inception on December 16, 2008, through January 2, 2021, as well as the posts made to "Depression Watch" from January 1, 2009, through January 2, 2021, were all gathered [7]. The dataset is freely available on https://www.kaggle.com which contains 2,3,2074 text. Loading the dataset takes lots of time due to defining large text so we changed the dataset according to our requirements.

3.2 Embeddings

There are two types of Embeddings for textual data: Word embeddings and Sentence embeddings. Word embeddings play a crucial role in deep models for providing input parameters in medium to late language tasks, like text categorization and sequence labeling. Over the past ten years, For this goal, a significant portion of word embedding techniques have been proposed, primarily falling into the areas of word embeddings based on context. Word embedding is the encoding of a word using fixed-length vectors of continuously real values. Using this method, a word is moved from a vocabulary to a latent vector space where terms in the area have similar contexts. It converts a word into a vector that includes all of the word's syntactic and semantic data. Therefore, it is believed to be particularly appropriate to use word embeddings as feature representations in neural network models for later NLP tasks, such as text categorization [20].Word embeddings represents in this paper are GloVe and Word2Vec. Contextual embeddings, like BERT and Elmo, surpass general word representations like Word2Vec and deliver outstanding performance on a variety of NLP applications. Contextual embeddings capture word usage in various settings and encode data that is transferable between languages by assigning each word a representation based on its context [21].

Sentence embeddings are essential for information retrieval services like search engines and recommender systems. With the growth of deep learning to improve the quality of language embeddings, DNN-based representation models have been frequently used as learning strategies [22]. Instead of computing embedding at the word level in the sentence embedding approach, vectors of embeddings reflect entire sentences and their semantic content. From word embeddings to sentence embeddings, there are two primary groups of sentence embedding approaches known as non-parameterized and parameterized models [23]. Universal Sentence Encoder embedding is used in this research paper for sentence embedding to the comparison of word embeddings with different classifiers such as RF, Decision Tree, SVM, KNN, AdaBoost, Log Reg, Gradient Boost classifier, XGB, and Naïve Bayes.

3.2.1 Sentence Embedding: Transformer-based USE

USE involves encoding the text into significant vectors that are capable of performing a variety of purposes, including semantic similarity, grouping, text categorization, etc. Tensor-Flow-hub has the pre-trained encoder available. The two variants of USE are the transformer encoder and the deep averaging network. Despite being more expensive to compute, the first one is more accurate. Even though the second one requires less processing, it is less accurate. A 512-dimensional vector is produced when different lengths of English text are provided[24]. Sentences were encoded into embedding vectors using two alternative methods. One uses Transformer [25] architecture, while the other uses a deep averaging network (DAN) [26]. The USE_T universal sentence encoder, which is more computationally costly than USE DAN and is based on the Transformers, has greater accuracy. We use the transformer encoder version of the USE in this study. Words, sentences, and documents can all be entered into USE_T [27]. When talking about universal sentence embeddings, the following three primary questions spring to mind:

(1) What distinguishes them from contextualized word embeddings,
(2) How useful are sentence embeddings?
(3) What exactly does the term "universal" mean?

First, there is insufficient contextualization for the current contextualized word embeddings to display the complete sentence. According to [28], higher layers of BERT cause word vectors from the same sentence to diverge from one another. This means that each word embedding from BERT still represents the local information from its position despite considering the context. [29], and [30] have demonstrated that neither the averaged BERT embedding nor the CLS embedding performs well downstream.

In light of this, it is important to understand how to properly create a sentence representation from powerful contextualized word embeddings. Nevertheless, it is still a topic worth debating. The requirement to integrate numerical global for sophisticated sentence phrases, and the presence of semantic/tax information is inevitable for artificial intelligence. Sentence embeddings are an efficient and effective way to retrieve such global information while word embeddings only encode local information.

Third, "universal" sentence embeddings strive to encode fundamental context-level information, much like BERT, which does so. Most tasks require sentence-level knowledge (semantics and grammar) and can be broadly used for sentence-level work in various fields, inspiring and promoting people everywhere. Universal sentence embeddings allow task-specific approaches to be further enhanced. Word and sentence embeddings serve as the foundation for document embeddings, which are hierarchically higher-level representations.

A comprehensive abstraction of the entire text should be the goal of a decent document representation. The key difficulty here is assembling the meaning of the individual pieces into a seamless and comprehensive meaning representation. In other words, a document has a complicated meaning, and there are an endless number of word and phrase combinations that can be used to create a document. Universal representations of documents are still a subject for further study, even though various task-specific document representations have so far demonstrated success in specific activities [31,32]. In this paper, we have used the pre-train model of USE [39].

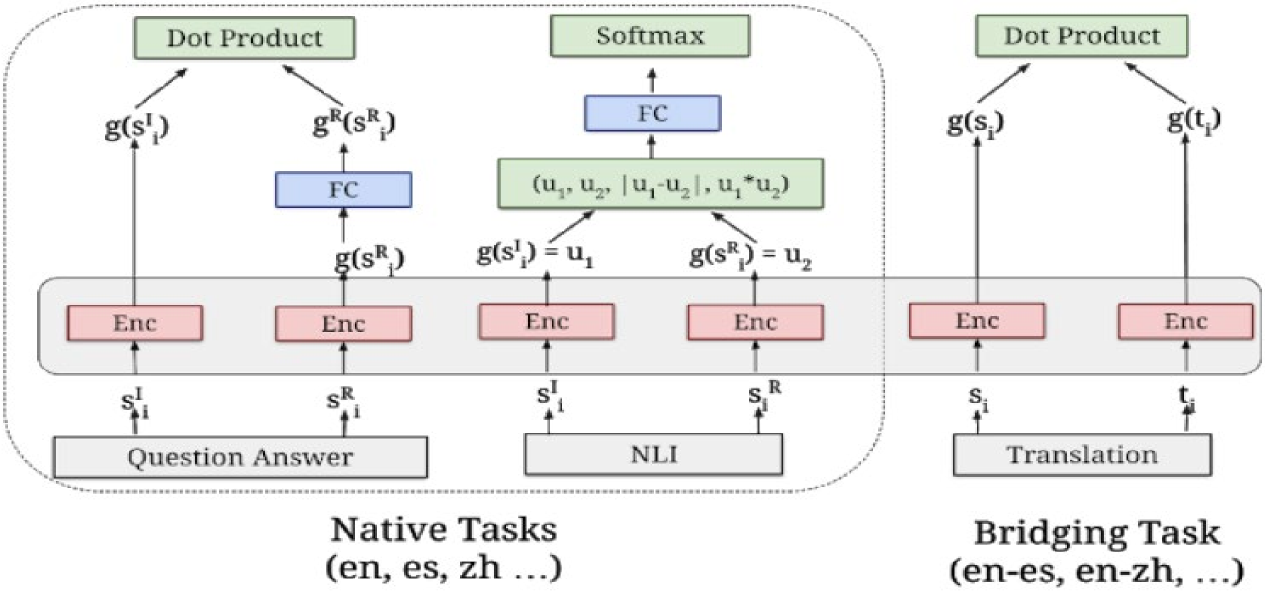

Fig. 1. Architecture of USE Transformer [39]

3.2.2 Word Embedding: Word2Vec

The two approaches used by Word2vec are Continuous Bag of Words (CBOW) and Skip-gram[33]. A vector-distributed numerical representation of word characteristics is produced by Word2vec. The goal and advantage of Word2vec groups vectors of related words into one large vector space. With enough information, word2vec can precisely be based on their history, and words' meanings can be predicted occurrence. These forecasts can be utilized to establish a word's relationship with words that are similar to it one another. The Skip-Gram word2vec model is employed. The benefit of Skip Gram is that it is better suited for words. they are far less common than CBOW. Since there are several word variations, the main justification for picking the Skip-Gram model is that it is believed that words that are infrequently used may appear. A continuous skip-gram makes predictions for a specific range by classifying terms before and after the center word as much as possible. Less weight is given to such distant words in training because they are less connected to the central word of a sentence. Contrarily, CBoW employs a succession of words from the past and future intending to correctly identify the target word in the middle. It functions by positioning all past or future words within a selected window in the same place, and their vector averages [34].

3.2.3 Word Embedding: GloVe

Global Vectors word embeddings, which Stanford addressed in 2014 and are typically used in the NLP people group prepared on Twitter, were accepted for the experiment's purposes[35]. The pre-made GloVe Twitter embeddings also come in four different sizes: 25, 50, 100, and 200. The co-occurrence matrix is produced using a 50-size symmetric window. Distributional representations, as represented by GloVe, 2and discrete representations, as represented by One-hot encoding, make up the majority of the ways that word embeddings are currently expressed. According to several studies, GloVe performs better on analogy tasks than the skip-gram and Word2Vecmodels when used in conjunction with matrix factorization techniques that can fully exploit global statistical data. Instead of focusing simply on the word's local meaning, GloVe word embedding takes the word's overall context into account. The context is established using a co occurrence matrix [36].

IV. EXPERIMENT AND RESULT

4.1 Proposed Methodology

Numerous methods are being developed to identify depression in postings. We use NLP and text classification algorithms to identify depression, and we include a detailed description of these methods in our work. The Fig. 1 framework includes data preprocessing, and feature extraction proceeded by machine learning classifiers, features analysis, and experimental findings.

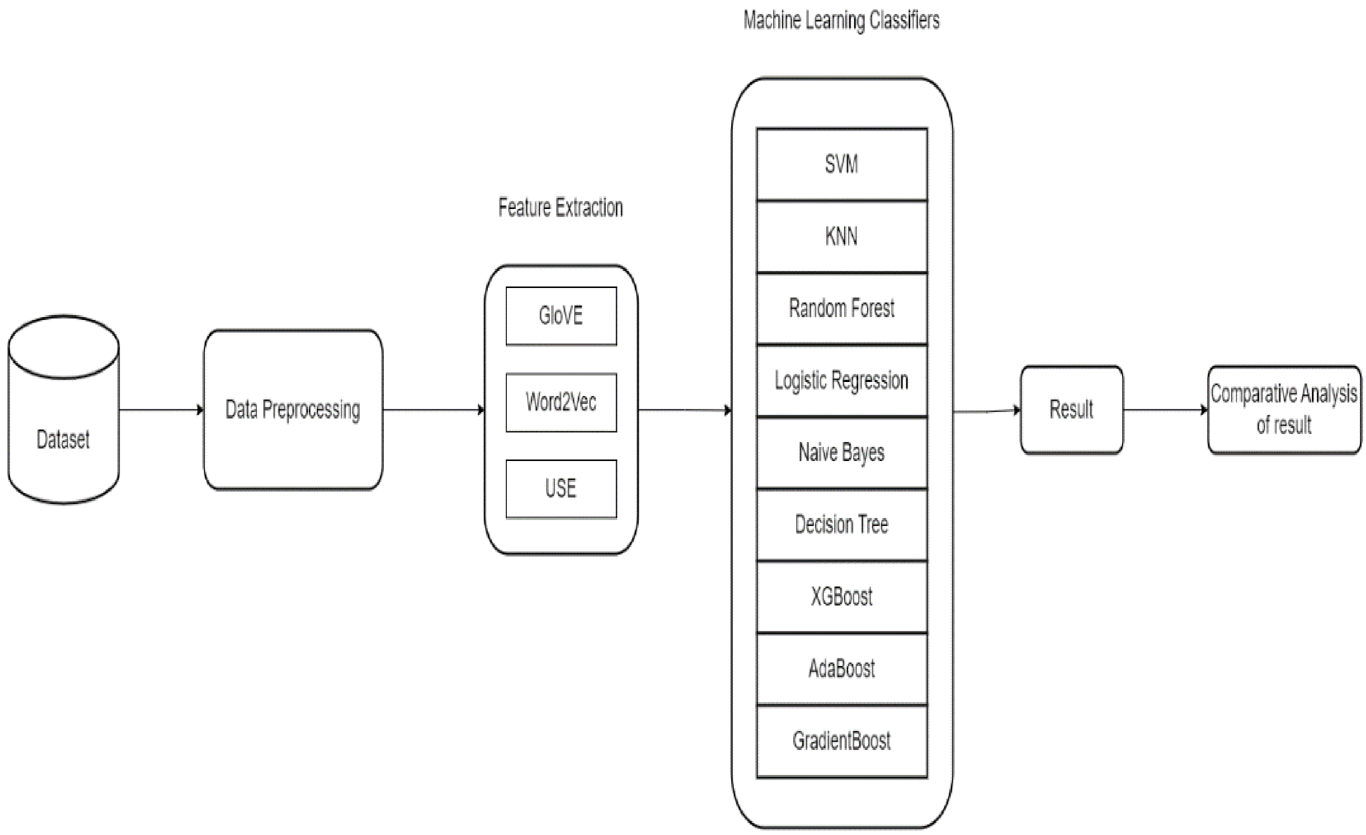

Fig. 2. Proposed framework for Depression Detection

4.1.1 Data Pre-processing
Before the dataset moves on to the stage of feature selection and training, we pre-process it using NLP tools. To begin with, we tokenize the posts to create separate tokens. We then eliminate any URLs, punctuation, or stop words that could result in unpredictable outcomes if left in.

4.1.2 Feature Extraction
Following data pre-processing, models are fed with features that correspond to users' Reddit forum language preferences. We make advantage of the word and sentence embedding characteristics to examine the language the users have used in the posts. These text encoding techniques are used to encode the words that various classifiers will come after.

4.1.3 Classification Techniques
We use classification techniques to evaluate the likelihood of depression among users to determine whether depression is present. SVM, RF, LR, KNN, Naïve Bayes, Decision Tree, XGBoost, Adaptive Boosting, and Gradient Boost Classifier are used to create the suggested framework.

4.2 Result
We tested various models with both the embeddings and decided to use 9 classifiers for classification, i.e. RF, SVM, KNN, Naïve Bayes, LR, Decision Tree, XGBoost, Adaptive Boosting, and Gradient Boost Classifier. The performances of classifiers are measured using precision, accuracy, F1 score, and recall which are shown in Table 2 and Table 3.

Word2Vec with ML classifiers:
For dataset 1, the best accuracy achieved is 60% for LR and XGBoost, and for dataset 2, it is 73% for the SVM. Logistic Regression and Random Forest also performed well compared to other models.

GloVe with ML classifiers:
For dataset 1, the best accuracy achieved is 61% for the gradient boost classifier, and for dataset 2, it is 76% for the random forest. Random Forest also performed well compared to other models.

USE with ML classifiers:
For dataset 1, the best accuracy achieved is 80% for LR, and for dataset 2, it is 92% for SVM and LR. Gradient Boost classifier, Random Forest, and XGBoost also performed well compared to other models.

Table 2. Accuracy Comparison of Word Embedding

Model	Dataset 1				Dataset2			
	Acc.	F1	P	R	Acc.	F1	P	R
Word2Vec+SVM	0.58	0.62	0.61	0.64	0.73	0.72	0.69	0.74
Word2Vec+LR	0.60	0.64	0.63	0.65	0.72	0.70	0.68	0.73
Word2Vec+KNN	0.55	0.62	0.58	0.66	0.66	0.61	0.65	0.57
Word2Vec+RF	0.57	0.63	0.60	0.66	0.72	0.71	0.68	0.76
Word2Vec+Bayes	0.57	0.61	0.60	0.61	0.68	0.66	0.65	0.67
Word2Vec+GBM	0.57	0.63	0.60	0.68	0.72	0.71	0.69	0.74
Word2Vec+DT	0.50	0.54	0.54	0.54	0.61	0.60	0.57	0.62
Word2Vec+XGB	0.60	0.65	0.61	0.70	0.71	0.70	0.68	0.73
Word2Vec+AdaBoost	0.56	0.61	0.59	0.63	0.71	0.70	0.68	0.72
GloVe+SVM	0.54	0.63	0.55	0.73	0.64	0.65	0.59	0.72
GloVe+LR	0.55	0.62	0.56	0.70	0.57	0.60	0.52	0.69
GloVe+KNN	0.48	0.55	0.51	0.61	0.58	0.58	0.54	0.64
GloVe+RF	0.60	0.63	0.59	0.67	0.76	0.75	0.72	0.78
GloVe+Bayes	0.55	0.61	0.57	0.66	0.53	0.59	0.49	0.74
GloVe+GBM	0.61	0.65	0.62	0.67	0.75	0.74	0.72	0.77
GloVe+DT	0.54	0.56	0.55	0.57	0.64	0.63	0.60	0.60
GloVe+XGB	0.56	0.61	0.64	0.58	0.75	0.74	0.71	0.77

GloVe+AdaBoost	0.57	0.61	0.64	0.59	0.74	0.72	0.71	0.73

Table 3. Accuracy Comparison of Sentence Embedding

Model	**Dataset 1**				**Dataset2**			
	Acc.	F1	P	R	Acc.	F1	P	R
USE+SVM	0.78	0.79	0.79	0.79	0.92	0.92	0.93	0.92
USE+LR	0.80	0.81	0.81	0.82	0.92	0.92	0.94	0.91
USE+KNN	0.72	0.76	0.69	0.86	0.90	0.90	0.85	0.96
USE+RF	0.77	0.78	0.77	0.78	0.91	0.91	0.95	0.88
USE+Bayes	0.75	0.76	0.75	0.77	0.90	0.90	0.94	0.86
USE+GBM	0.79	0.80	0.78	0.82	0.91	0.91	0.94	0.89
USE+DT	0.68	0.69	0.70	0.67	0.84	0.84	0.86	0.83
USE+XGB	0.78	0.79	0.78	0.79	0.91	0.91	0.93	0.89
USE+AdaBoost	0.74	0.75	0.76	0.74	0.89	0.89	0.90	0.88

Table 4. Model Processing Time

Model	**# Time(seconds)**	
	Dataset 1	**Dataset 2**
Word2Vec+SVM	0.25	2.56
Word2Vec+LR	0.02	0.05
Word2Vec+KNN	0.00	0.00
Word2Vec+RF	1.32	3.46
Word2Vec+Bayes	0.00	0.03
Word2Vec+GBM	3.04	8.56
Word2Vec+DT	0.12	0.33
Word2Vec+XGB	0.62	1.92
Word2Vec+AdaBoost	0.70	1.98
GloVe+SVM	0.37	1.23
GloVe+LR	0.21	0.22

International Journal of Engineering Applied Sciences and Technology, 2023
Vol. 8, Issue 02, ISSN No. 2455-2143, Pages 16-26
Published Online June 2023 in IJEAST (http://www.ijeast.com)

GloVe+KNN	0.10	0.17
GloVe+RF	0.91	1.38
GloVe+Bayes	0.03	0.05
GloVe+GBM	1.52	2.53
GloVe+DT	0.06	0.10
GloVe+XGB	0.36	0.53
GloVe+AdaBoost	0.54	0.55
USE+SVM	0.83	4.20
USE+LR	0.25	0.38
USE+KNN	0.04	0.03
USE+RF	4.05	12.07
USE+Bayes	0.06	0.12
USE+GBM	40.11	93.15
USE+DT	1.83	7.08
USE+XGB	8.25	15.67
USE+AdaBoost	8.90	18.12

Table 5. Comparison of our method with existing methods

References	Model	Precision(%)	Recall(%)	F1-Score(%)
Present study	LogReg w/ pre-trained USE + features	81	82	81
Elsbeth Turcan et al. 2019 [6]	LogReg w/ pre-trained Word2Vec + features	73.46	81.03	77.06

From Table 5, we can conclude that Logistic Regression works better with Universal Sentence Encoder as compared to Word2Vec. Since the Universal Sentence Encoder can be used to encode text into high-dimensional vectors for semantic similarity, text classification, clustering, and other tasks using natural language. Additionally, USE is so effective that it may be used for sentences, phrases, or brief paragraphs even with high batch sizes and an encoder that is optimized for text longer than a word. Prior to analysis, USE should also remove all duplicate sentences or phrases. They discover that sentence-level transfer learning typically outperforms transfer learning utilizing sentence embeddings [38]. They observe remarkably strong performance for a transfer task with text embeddings, requiring just small quantities of supervised training data.

V. CONCLUSION AND FUTURE SCOPE

In this work, we investigated various classifier sets with various embeddings for social media depression prediction. We used the text messages from both datasets of people who post on the social site to classify depression and looked at which word and sentence embedding method performed better. According to our research, the results of sentence embedding with all of the classifiers outperform that of word embedding. The Universal Sentence Encoder has a maximum accuracy of 80% for dataset 1 and 92% for dataset 2.

Future research will use semantic role labeling to evaluate various classification models, examine additional features, and produce better results. We will also create a general model that can be used with any dataset.

VI. REFERENCES

[1]. Gui, T., Zhang, Q., Zhu, L., Zhou, X., Peng, M. and Huang, X., 2019. Depression detection on social media with reinforcement learning. In Chinese Computational Linguistics: 18th China National Conference, CCL 2019, Kunming, China, October 18–20, 2019, Proceedings 18 (pp. 613-624). Springer International Publishing.

[2]. WHO, "Depression," 2021.

[3]. Keles, B., McCrae, N. and Grealish, A., 2020. A systematic review: the influence of social media on depression, anxiety and psychological distress in adolescents. International Journal of Adolescence and Youth, 25(1), pp.79-93.

[4]. Keles, B., McCrae, N. and Grealish, A., 2020. A systematic review: the influence of social media on depression, anxiety and psychological distress in adolescents. International Journal of Adolescence and Youth, 25(1), pp.79-93.

[5]. Añel, J.A., 2011. The importance of reviewing the code. Communications of the ACM, 54(5), pp.40-41.

[6]. Turcan, E. and McKeown, K., 2019. Dreaddit: A Reddit dataset for stress analysis in social media. arXiv preprint arXiv:1911.00133.

[7]. https://www.kaggle.com/datasets/nikhileswarkomati/suicide-watch

[8]. Stankevich, M., Isakov, V., Devyatkin, D. and Smirnov, I.V., 2018. Feature Engineering for Depression Detection in Social Media. In ICPRAM (pp. 426-431).

[9]. Park, M., Cha, C. and Cha, M., 2012. Depressive moods of users portrayed in Twitter. In Proceedings of the 18th ACM International Conference on Knowledge Discovery and Data Mining, SIGKDD 2012 (pp. 1-8).

[10]. Xu, R. and Zhang, Q., 2016. Understanding online health groups for depression: social network and linguistic perspectives. Journal of medical Internet research, 18(3), p.e63.

[11]. Teck Kiong, Y., 2022, June. An Initial Study of Depression Detection on Mandarin Textual through BERT Model. In 14th ACM Web Science Conference 2022 (pp. 459-463).

[12]. Tadesse, M.M., Lin, H., Xu, B. and Yang, L., 2019. Detection of depression-related posts in reddit social media forum. IEEE Access, 7, pp.44883-44893.

[13]. Cacheda, F., Fernandez, D., Novoa, F.J. and Carneiro, V., 2019. Early detection of depression: social network analysis and random forest techniques. Journal of medical Internet research, 21(6), p.e12554.

[14]. Uddin, M.Z., Dysthe, K.K., Følstad, A. and Brandtzaeg, P.B., 2022. Deep learning for prediction of depressive symptoms in a large textual dataset. Neural Computing and Applications, 34(1), pp.721-744.

[15]. Aalam, P. and Siddiqui, T., SHORT TEXT CLUSTERING CONCEPTUAL MODEL TO DETECT TWITTER USERS'DEPRESSION LEVELS. Journal of SEYBOLD Report, 15(7).

[16]. Elshazly, M., Haggag, M.H. and Ehssan, S.A., 2021. A Depression Detection Model using Deep Learning and Textual Entailment. International Journal of Computer Science and Information Security (IJCSIS), 19(12).

[17]. Ahmad, H., Asghar, M.Z., Alotaibi, F.M. and Hameed, I.A., 2020. Applying deep learning technique for depression classification in social media text. Journal of Medical Imaging and Health Informatics, 10(10), pp.2446-2451.

[18]. Chiong, R., Budhi, G.S., Dhakal, S. and Chiong, F., 2021. A textual-based featuring approach for depression detection using machine learning classifiers and social media texts. Computers in Biology and Medicine, 135, p.104499.

[19]. Kumar, A., Sharma, A. and Arora, A., 2019. Anxious depression prediction in real-time social data. arXiv preprint arXiv:1903.10222.

[20]. Wang, C., Nulty, P. and Lillis, D., 2020, December. A comparative study on word embeddings in deep learning for text classification. In Proceedings of the 4th International Conference on Natural Language Processing and Information Retrieval (pp. 37-46).

[21]. Liu, Q., Kusner, M.J. and Blunsom, P., 2020. A survey on contextual embeddings. arXiv preprint arXiv:2003.07278.

[22]. Yang, J., Liu, Z., Xiao, S., Lian, J., Wu, L., Lian, D., Sun, G. and Xie, X., 2022. A Mutually Reinforced Framework for Pretrained Sentence Embeddings. arXiv preprint arXiv:2202.13802.

[23]. B. Wang and C. . -C. J. Kuo, "SBERT-WK: A Sentence Embedding Method by Dissecting BERT-Based Word Models," in IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 28, pp. 2146-2157, 2020, doi: 10.1109/TASLP.2020.3008390.

[24]. Deepthi, G. and Sowjanya, A.M., 2021. Query-Based Retrieval Using Universal Sentence Encoder. Rev. d'IntelligenceArtif., 35(4), pp.301-306.

[25]. Iyyer, M., Manjunatha, V., Boyd-Graber, J. and Daumé III, H., 2015, July. Deep unordered composition rivals syntactic methods for text classification. In Proceedings of the 53rd annual meeting of the association for computational linguistics and the 7th international joint conference on natural language processing (volume 1: Long papers) (pp. 1681-1691).

[26]. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł. and Polosukhin, I., 2017. Attention is all you need. Advances in neural information processing systems, 30.

[27]. Alibadi, Z., Du, M. and Vidal, J.M., 2019, May. Using Pre-trained Embeddings to Detect the Intent of an Email. In Proceedings of the 7th ACIS International Conference on Applied Computing and Information Technology (pp. 1-7).

[28]. Ethayarajh, K., 2019. How contextual are contextualized word representations? comparing the geometry of BERT, ELMo, and GPT-2 embeddings. arXiv preprint arXiv:1909.00512.

[29]. Li, B., Zhou, H., He, J., Wang, M., Yang, Y. and Li, L., 2020. On the sentence embeddings from pre-trained language models. arXiv preprint arXiv:2011.05864.

[30]. Reimers, N. and Gurevych, I., 2019. Sentence-bert: Sentence embeddings using siamesebert-networks. arXiv preprint arXiv:1908.10084.

[31]. Dai, A.M., Olah, C. and Le, Q.V., 2015. Document embedding with paragraph vectors. arXiv preprint arXiv:1507.07998.

[32]. Wang, S., Tang, J., Aggarwal, C. and Liu, H., 2016, October. Linked document embedding for classification. In Proceedings of the 25th ACM international on conference on information and knowledge management (pp. 115-124).

[33]. Onishi, T. and Shiina, H., 2020, September. Distributed representation computation using CBOW model and skip–gram model. In 2020 9th International Congress on Advanced Applied Informatics (IIAI-AAI) (pp. 845-846). IEEE.

[34]. Adewumi, T., Liwicki, F. and Liwicki, M., 2022. Word2Vec: Optimal hyperparameters and their impact on natural language processing downstream tasks. Open Computer Science, 12(1), pp.134-141.

[35]. Pennington, J., Socher, R. and Manning, C.D., 2014, October. Glove: Global vectors for word representation. In Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP) (pp. 1532-1543).

[36]. Giri, S., Banerjee, S., Bag, K. and Maiti, D., 2022, February. Comparative Study of Content-Based Phishing Email Detection Using Global Vector (GloVe) and Bidirectional Encoder Representation from Transformer (BERT) Word Embedding Models. In 2022 First International Conference on Electrical, Electronics, Information and Communication Technologies (ICEEICT) (pp. 01-06). IEEE.

[37]. Cardoso, Ana Bárbara, Bruno Martins, and Jacinto Estima. "A Novel Deep Learning Approach Using Contextual Embeddings for Toponym Resolution." ISPRS International Journal of Geo-Information 11.1 (2022): 28.

[38]. Cer, D., Yang, Y., Kong, S.Y., Hua, N., Limtiaco, N., John, R.S., Constant, N., Guajardo-Cespedes, M., Yuan, S., Tar, C. and Sung, Y.H., 2018. Universal sentence encoder. arXiv preprint arXiv:1803.11175.

[39]. https://tfhub.dev/google/universal-sentence-encoder-multilingual/3