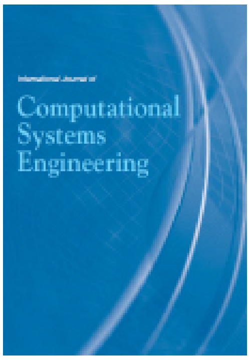
INDERSCIENCE PUBLISHERS
Linking academia, business and industry through research



International Journal of Computational Systems Engineering

ISSN online: 2046-3405 - ISSN print: 2046-3391
https://www.inderscience.com/ijcsyse

Depression prediction and therapy recommendation using machine learning technique

K.G. Saranya, C.H. Babitha Reddy, M. Bhavyasree, M. Rubika, E. Varsha

DOI: 10.1504/IJCSYSE.2024.10062832

Article History:

Received:	02 September 2022
Last revised:	24 March 2023
Accepted:	18 April 2023
Published online:	19 March 2024

Depression prediction and therapy recommendation using machine learning technique

K.G. Saranya*, C.H. Babitha Reddy, M. Bhavyasree, M. Rubika and E. Varsha

Department of CSE,
PSG College of Technology,
Coimbatore-04, India
Email: kgs.cse@psgtech.ac.in
Email: babithareddy2000@gmail.com
Email: bhavyamohan2001@gmail.com
Email: rubika0309@gmail.com
Email: varshaeswaran2001@gmail.com
*Corresponding author

Abstract: The most common misconception around the world would be the definition of 'health'. A person is considered to be healthy as long as they are physically fine, but that is not true. A person's mental health is also equally important while considering a person's health status. This incomprehension towards mental health has taken lives of many people. Despite the government and many NGOs spreading awareness on mental health, there is still a lack of understanding of mental health symptoms, societal stigma, and proper resources and facilities prevent people from seeking help. Among the types of mental disorders, many psychiatrists have agreed that depression and addiction are the most common ones to cause a person's life. There are a couple of existing systems that aids in depression detection and therapy recommendation, but the major issue found in those systems would be inefficiency and high computational cost. In this paper work, a new approach has been proposed to identify depression using Reddit comments.

Keywords: bidirectional encoder representation transformer; BERT; collaborative filtering; cosine similarity.

Reference to this paper should be made as follows: Saranya, K.G., Reddy, C.H.B., Bhavyasree, M., Rubika, M. and Varsha, E. (2024) 'Depression prediction and therapy recommendation using machine learning technique', *Int. J. Computational Systems Engineering*, Vol. 8, Nos. 1/2, pp.120–127.

Biographical notes: K.G. Saranya is working as an Assistant Professor in the Department of CSE at PSG College of Technology since 2010. She has published more than 20 papers in reputed journals in the area of recommender systems.

C.H. Babitha Reddy is an under graduate student of Department of CSE at PSG College of technology.

M. Bhavyasree is an under graduate student of Department of CSE at PSG College of technology.

M. Rubika is an under graduate student of Department of CSE at PSG College of technology.

E. Varsha is an under graduate student of Department of CSE at PSG College of technology.

1 Introduction

Mental health has always been a serious issue and the most common of them all would be depression. Especially since the onset of the pandemic, the depression rates have been on increase in a pretty steady rate.

In general, it is pretty complex to identify a person with mental health related issue without keen observation. The pandemic has made it quite difficult for proper clinical interactions. So what are the other ways to observe a person's mental status, it would be through computational methods. Examining the comments and postings on such sites can reveal how people self-disclose and discuss mental health conditions like depression. This system seeks to diagnose sadness using a dataset of scraped Reddit comments. Comments from both depressed and non-depressed people were scraped. This study looks on machine learning methods for detecting depression in text inputs that have been shared digitally. This paper was done based on a BERT model.

2 Related work

This work helps in analysing the depression and predicts a suitable therapy. On preparation of this work, other notable works were referred based on conventional neural networks NER and RNN. Other notable works are listed below:

Du et al. (2018) has proposed a system titled 'Extracting psychiatric stressors for suicide from social media using deep learning'. This paper suggested a system that uses the combination of deep learning techniques and achieves accuracy of about 83% but it could not able to identify the medical terms like disease name from tweets.

Jamil et al. (2017) has proposed a system titled 'Monitoring tweets for depression to detect at-risk users'. This paper described an automated system which compares twitter level and user level classification to achieve high accuracy but the dataset used in the system contains only 5% depression tweets so the accuracy achieved may not be same for high percentage tweets.

Shen et al. (2017) has proposed a system titled 'Depression detection via harvesting social media: a multimodal dictionary learning solution'. This paper proposed a system detects timely depression using feature modalities to identify depressed and non-depressed users with large number of datasets. The major drawback that the system took a large amount of time for processing as it needs to join the modalities.

Bayer et al. (2016) has proposed a system titled 'A generic coordinate descent framework for learning from implicit feedback'. This paper described a system that was meant to serve as a broad framework for developing recommender system learning algorithms but all resources are used to process one training sample at a time so frequent updates are computationally costly.

Cheng et al. (2016) has proposed a system titled 'Wide and deep learning for recommender systems'. This paper proposed a system combines the features of memorisation and generalisation to achieve better results but it requires large amount of data to process and expensive.

Ramalingam et al. (2019) has proposed a system titled 'Study of depression analysis using machine learning techniques'. In this paper the learning algorithm uses active and passive creation of all grammatical and general properties of text to try to determine the text's possible and probabilistic meanings. In the instance of depression analysis, they also used semantic algorithms to determine the user's emotion. To distinguish between postings with and without depression stigma, SVM classification is utilised. The system attained an average detection accuracy of 82.2% for males and 70.5% for females using this method.

Islam et al. (2018) has proposed a system titled 'Depression detection from social network data using machine learning techniques'. The purpose of this article is to do a depression analysis using Facebook data obtained from a public internet source. The proposed strategy can greatly enhance classification accuracy and error rates. The results reveal that in several studies, decision tree outperforms other ML techniques in detecting depression.

Thorstad et al. (2019) has proposed a system titled 'Predicting future mental illness from social media: a big-data approach'. The purpose of this article is to suggest mental health can be predicted from people's publicly accessible activity on social media, even seemingly outside of mental health contexts but posting on a clinical subreddit is not a gold-standard diagnosis. Thus, it is not possible to be certain that individuals on a clinical subreddit have been diagnosed with that disorder is the drawback that is faced.

Kamal et al. (2020) has proposed a system titled 'Predicting mental illness using social media posts and comments'. This paper described that the effectiveness of the proposed methodology to classify the patient data more effectively as compared to the state of the art classifiers and 68% accuracy was achieved but it is applicable to few classes of disorders.

Hung et al. (2015) has proposed a system titled 'A smartphone-based personalised activity recommender system for patients with depression'. In this paper, it is suggested that the precision of the recommendation improved with time, personal activity recommender system helped users alleviate negative emotions but the system is not capable of communicating with the electronic health record, which would further the detection of clinically notable emotional alteration.

Rohani et al. (2020) have proposed a system titled 'MUBS: a personalised recommender system for behavioural activation in mental health'. In this system, the system was developed in a user-centred design process to support BA therapy for patients with depressive symptoms. Its features were designed to adhere to core BA principles but the recommender model did not include any post-filtering functionality on the recommended activities. Users received recommended activities such as 'cuddle with your pet' while the user does not have one.

Kumar et al. (2019) has proposed a system titled 'Anxious depression prediction in real-time social data'. In this Twitter dataset was used and tokenisation was performed using Treebank WordTokeniser of NLTK toolkit. Three machine learning classifiers are used namely, multinomial naïve Bayes, gradient boosting and random forest. The final prediction is made using an ensemble vote classifier with a majority voting process. For training and testing, the data was split 80/20, utilised a 10-fold cross-validation method.

3 Existing system

In the existing systems, some of the detecting depression systems are discussed and compared to the proposed system to demonstrate that the suggested system is more efficient.

Many efforts have been undertaken in recent years to employ machine learning techniques to evaluate social media information in order to detect sadness among users. Facebook and Twitter are the most commonly utilised databases for this purpose. The posts or comments in these datasets may not always reflect depression-like characteristics, despite the fact that they are custom-made. Furthermore, these datasets tend to be older when obtained from sites like Kaggle. When deployed and made to work on real-world data, these machine learning or deep learning models constructed using these datasets may perform poorly.

The classification algorithm is a supervised learning technique that uses training data to determine the category of new observations. Classification is the process of software learning from a dataset or observations and then classifying fresh observations into one of several classes or groupings.

A binary classifier is used when the classification issue has just two possible outcomes. As this work aims at classifying depressed/non-depressed, Binary classifier is used.

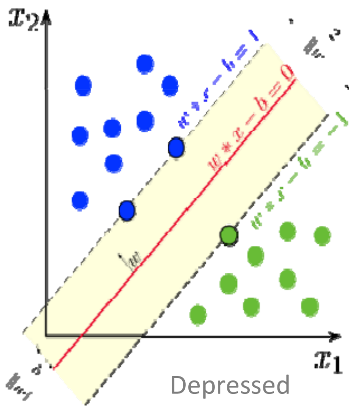
3.1 *Support vector machine*

The support vector machine, or SVM, is a popular Supervised Learning technique that may be used to solve both classification and regression issues. However, it is mostly employed in machine learning for classification challenges. The SVM algorithm's purpose is to find the optimum line or decision boundary for categorisation of n-dimensional space into classes so that additional data points can be readily placed in the correct category in the future. A hyper plane is the name for the optimal choice boundary. The extreme points/vectors that assist create the hyper plane are chosen via SVM (Ravi et al., 2021).

The technique is termed a support vector machine, and the extreme cases are called SVM. The algorithm is known as a support vector machine, and these are the extreme cases.

Since the dataset will be classified under two classes namely depressed and non-depressed, linear SVM classifier is used.

Figure 1 Classification of data with SVM (see online version for colours)



There are two types of SVM:

- Linear SVM: linear SVM is a classifier for linearly separable data, which means that a dataset can be classified into two classes using a single straight line, and the classifier is termed linear SVM.
- Nonlinear SVM: nonlinear SVM is a classifier that is used for nonlinearly separated data. This means that if a dataset cannot be classified using a straight line, it is nonlinear data, and the classifier used is termed nonlinear SVM.

3.2 *Naive Bayes*

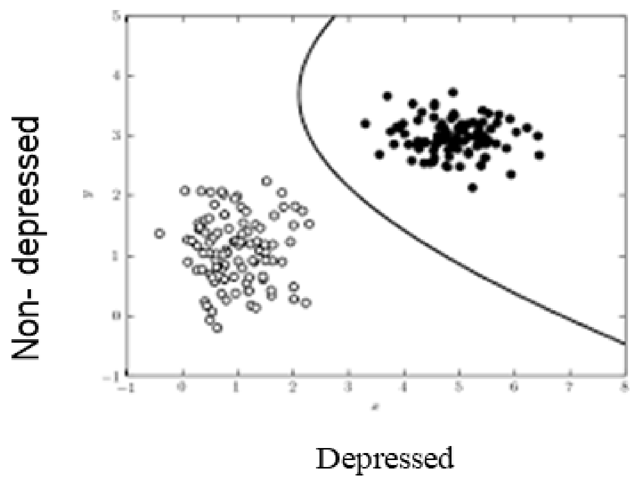
- Naive: it is termed naive because it assumes that the appearance of one feature is unrelated to the appearance of others. If the colour, shape, and flavour of the fruit are used to identify it, a red, spherical, and sweet fruit is identified as an apple. As a result, each aspect helps to identifying that it is an apple without relying on the others.
- Bayes: it is called Bayes since it is based on the Bayes' theorem concept.

$$P(A \mid B) = \frac{P(B \mid A)P(A)}{P(B)}$$

  - $P(A)$ is prior probability: probability of hypothesis before observing the evidence. Here it denotes the dataset which is typically the comments attribute.
  - $P(B)$ is marginal probability: probability of evidence.

    Here it denotes the prediction output.
  - $P(A|\,B)$ is posterior probability: probability of hypothesis A on the observed event B.
  - $P(B|\,A)$ is likelihood probability: probability of the evidence given that the probability of a hypothesis is true.

Figure 2 Classification of data with naive Bayes



3.3 Working principle

It creates frequency tables from the given dataset and find the probability of given features to generate a likelihood table and then calculate the posterior probability using Bayes theorem.

3.4 Linear regression

One of the most common and simple machine learning methods is linear regression. It is a statistical strategy for predicting the future. Linear regression is used to predict sales, salary, age, product price, and other continuous, real, or numeric factors. The term comes from the fact that the linear regression algorithm illustrates a linear relationship between a dependent (y) variable and one or more independent (x) variables. Because linear regression displays a linear relationship, it may be used to figure out how the value of the dependent variable evolves as a function of the value of the independent variable. In the linear regression model, each link between the variables is represented by a slanted straight line. Mathematically, linear regression is

$$y = a_0 + a_1 x + \varepsilon$$

Y Dependent variable (TargetVariable), here it is the input sentence given by the user.

X Independent variable (predictorVariable), here it is the output being predicted.

a_0 intercept of the line (gives an additional degree of freedom).

a_1 Linear regression coefficient (scale factor to each input value).

ε random error

Linear regression algorithms are divided into two types:

- Simple linear regression: simple linear regression is a linear regression approach that predicts the value of a numerical dependent variable using only one independent variable.
- Multiple linear regressions: multiple linear regressions is a linear regression method that predicts the value of a numerical dependent variable using more than one independent variable.

Multiple linear regressions have been used to classify the data in the existing system.

4 Proposed work

4.1 Collecting dataset

Dataset contains scraped Reddit comments. Comments from both depressed and non-depressed people were scraped. It is a pre-processed dataset.

Figure 3 Classification of data with linear regression (see online version for colours)

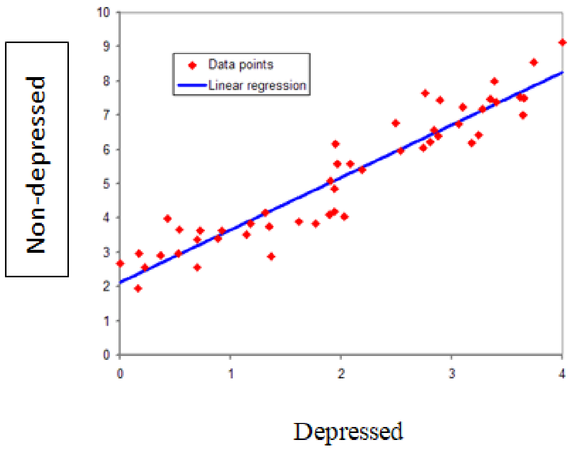
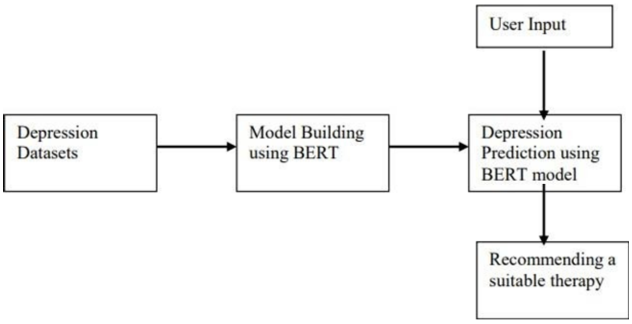


Figure 4 Architecture of the proposed system collecting dataset



4.2 Model building using BERT

The prediction algorithm aims at classifying if a person is really depressed or not. BERT model is used to achieve the same. Negative words in a tweet necessarily do not indicate depression, the model will be trained in such a way that it classifies based on the meaning of the text.

4.3 Depression prediction using BERT model

The sentence taken as input will be predicted as depressed or non-depressed.

4.4 Recommend a suitable therapy

The recommender system will be developed in such a way that recommendation will be different for different users after predicting their level of depression. Here a collaborative filtering-based model is implemented, where we have a dataset which contains details of the patients and the therapy that they have taken before. Now, for a new user input from the prediction system will be fed into the recommender model where a cosine similarity is calculated to find a suitable match from the existing dataset and based on that a therapy for the new user is recommended.

4.5 User input

Inputs from the user such as age, family background, year of diagnosis and level of depression are taken, from which, the person's depression status will be predicted and a therapy would be recommended.

5 Proposed system architecture

5.1 BERT model

BERT model is a deep learning model that has given state-of-the-art results on a wide variety of natural language processing tasks. It stands for bidirectional encoder representations for transformers. It has been pre-trained on Wikipedia and Books Corpus and requires task-specific fine-tuning.

BERT model is a multi-layer bidirectional transformer encoder. There are two models introduced in the paper.

- BERT base model – 12 layers (transformer blocks), 12 attention heads, and 110 million parameters.
- BERT large model – 24 layers, 16 attention heads and, 340 million parameters.

It is intended to condition both left and right context to pre-train deep bidirectional representations from unlabeled text. As a result, with just one additional output layer, the pre-trained BERT model may be fine-tuned to generate state-of-the-art models for a wide range of NLP tasks.

That sounds way too complex as a starting point. But it does summarise that BERT model does pretty well so let's break it down.

First, it is easy to get that BERT stands for bidirectional encoder representations from transformers. Each word here has a meaning to it and we will encounter that one by one in this article. For now, the key takes away from this line is – BERT is based on the transformer architecture.

Second, BERT model is pre-trained on a large corpus of unlabeled text including the entire Wikipedia (that's 2,500 million words!) and Book Corpus (800 million words).

This *pre-training* step is half the magic behind BERT's success. This is because as we train a model on a large text corpus, our model starts to pick up the deeper and intimate understandings of how the language works. This knowledge is the Swiss army knife that is useful for almost any NLP task.

Third, BERT model is a 'deeply bidirectional' model. Bidirectional means that BERT model learns information from both the left and the right side of a token's context during the training phase.

The Bert model encodes only the sentences which are tokenised. The tokenised statements contain

- [CLS] class variable which represents beginning of the sentences.
- [TOK] tokens indicate the words in the sentence. Every word is represented as tok1, tok2 … and token.
- [SEP] separator is used to mention the end of each sentence when there are 2 or more statements.

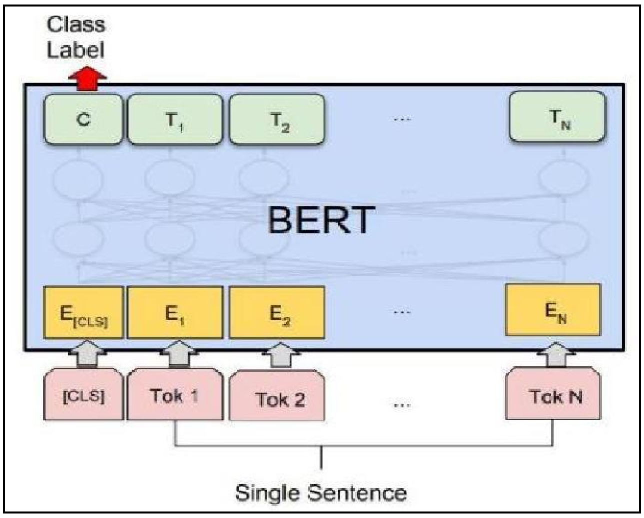
In the encoder, the tokenised words are encoded by considering their positions as E0, E1 and En.

In this architecture, multihead attention mechanism is used. This mechanism clubs each word with every other word in the sentences to get the actual representation of the word according to the sentences. For example, Python is my favourite programming language.

In the Amazon Forest, the python can be found. Here the sentences contain the same word python but it contributes different meaning in both the sentences. Hence multihead attention mechanism considers both python as different and represents accordingly.

Using this mechanism, the encoded tokens are represented in a vectored form as T1, T2 and Tn. C is the actual output that is trained according to the dataset.

Figure 5 Architecture of BERT model (see online version for colours)



5.2 Collaborative filtering

Collaborative filtering tends to locate similar people and recommend what they like in collaborative filtering. We do not use the item's features to recommend it in this form of recommendation system; instead, we group users into clusters of similar categories and recommend each user based on the preferences of their cluster.

Only the users' past preferences on a set of items are required for collaborative filtering. Because it is based on historical data, the underlying premise is that users who have agreed in the past are more likely to agree in the future. User preferences are typically expressed in two categories. An *explicit rating* is a rating given to an item on a sliding scale by a user, such as 5 stars for Titanic. This is the most direct way for users to express how much they enjoy a product. Page views, clicks, purchase records, whether or not to listen to a music piece, and other factors all contribute to *implicit rating*.

5.3 Types of collaborative filtering

There are two types of collaborative filtering'

- User-user-based collaborative filtering
- Item-item-based collaborative filtering.

User-to-user collaborative filtering is a type of recommendation strategy that looks for users who have enjoyed or positively interacted with similar goods.

User-to-user collaborative filtering has been used in the proposed system. If a new user needs a therapy, the user needs to enter their gender, age, year of diagnosis, stage of depression based on which the users are matched using cosine similarity and then the therapy is recommended according to the therapy that the previous user has undergone.

5.4 Cosine similarity

The metric cosine similarity is used to determine how similar two objects are. It estimates the cosine of the angle formed by two vectors projecting in a nonlinear and non-space. The output value is between 0 and 1. 0 indicates no resemblance, whereas 1 indicates that both the elements are identical.

$$\text{Cos}(x, y) = x.y / \|x\| * \|y\|$$

X is the data of the new user

Y is the data of the past user.

5.5 Jaccard similarity

The Jaccard similarity offers us a measure of similarity between two sets by calculating the number of things they have in common and dividing by the total number of unique items between them. It is basically the ratio of how many objects they both share with how many they may potentially share.

$$J(A, B) = |A \cap B| / |A \cup B|$$

J Jaccard distance

A is the data of the new user

B is the data of the past user.

5.6 Euclidean distance score

The length of the line segments connecting two places is defined as the Euclidean distance between them. In this scenario, our Euclidean space is the positive section of the plane, where the axes represent the ranking items and the points represent the scores that a certain individual assigns to both things.

$$d(p1, p2) = dist((x, y), (a, b)) = \sqrt{}\,(x-a)^2 + (y-b)^2$$

d Euclidean distance

p1 is the data of the new user

p2 is the data of the past user.

6 Experimental analysis

The experimental analysis on 'Depression prediction and therapy recommendation' is done with the help of precision, recall, and F1 score.

More specifically, the number of true positives (TP), false positives (FP) and false negatives (FN) is used to compute the recall, precision, and F-score of BERT model.

- *True positive (TP):* a true positive is when the model predicts the positive class correctly.
- *True negative (TN):* a true negative, on the contrary, is a result in which the model correctly predicts the negative class.
- *False positive (FP):* a false positive occurs when the model predicts the positive class inaccurately.
- *False negative (FN):* a false negative is an outcome in which the model predicts the negative class inaccurately.
- *Precision:* precision is one measure of a machine learning model's performance – the correctness of a model's positive prediction. The number of true positives divided by the total number of positive predictions (i.e., the number of true positives plus the number of false positives) is known as precision.

  $$\text{Precision} = TP / TP + FP$$

- *Recall:* the recall is determined by dividing the total number of positive samples by the number of Positive samples correctly classified as positive. The recall determines how well the model can detect positive samples. The greater the recall, the more positive samples are discovered.

  $$\text{Recall} = TP / TP + FN$$

- *F1-score:* the harmonic mean of precision and recall is used to get the F1-score. The mean of precision and recall is the F1-score.

  $$\text{F1-score} = \frac{2 * (\text{Precision} * \text{Recall})}{\text{Precision} + \text{Recall}}$$

- *Accuracy:* accuracy is a statistic that sums up how well a model performs across all classes. It is calculated as the ratio between the numbers of true predictions to the total number of predictions.

6.1 Analysis of prediction accuracy of BERT model: results of Bert model

The results of the performance of the BERT model have been tabulated below.

Table 1 Analysis of prediction accuracy of BERT model

Size of the dataset: 5,000 rows	
Metric	*Value*
True positives	2028
True negatives	2218
False positives	358
False negatives	396
Precision	84.99
Recall	83.66
F1-score	84
Accuracy	85

It is inferred from Table 1 that the accuracy is 0.85 which means the BERT model is 85% accurate in making a correct prediction.

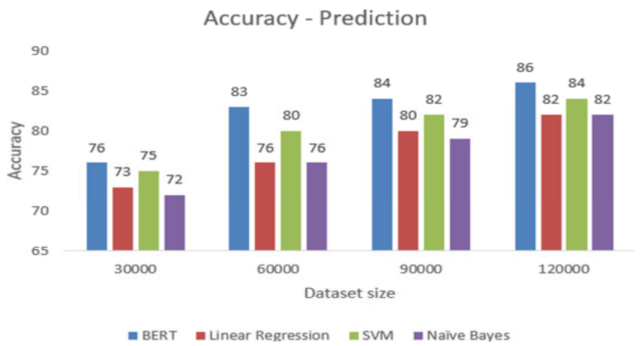
6.2 *Comparison of the BERT model with existing methods*

From Table 2, it is inferred that BERT has 4% of accuracy outperforming the other machine learning models linear regression, SVM and naive Bayes.

Table 2 Comparison of the BERT model with existing methods

Size of the dataset: 1,275,595 rows				
Models	*Measures*			
	Accuracy	*Precision*	*Recall*	*F1 score*
BERT	86	84.99	83.66	84
Linear regression	84	85	81	83
Support vector machine	84	85	81	83
Naïve Bayes	82	77	87	82

Figure 6 Graph of performance measures for proposed and existing systems (prediction) (see online version for colours)



From Figure 6, it is observed that the accuracy is increased with increase in the size of the dataset. From Figure 6, it is observed that the proposed system gives 4% accuracy improvement over existing system

Table 3 Comparison of the cosine similarity with similarity measures

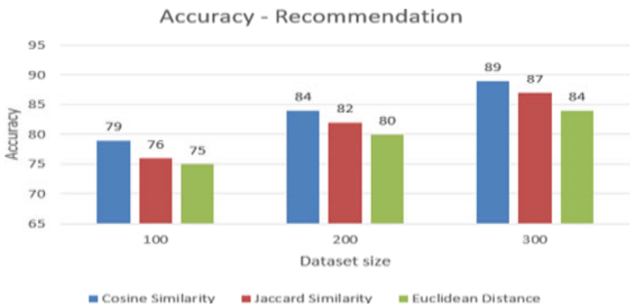
Size of the dataset: 400 rows				
Similarity measure	*Metric*			
	Accuracy	*Precision*	*Recall*	*F1-measure*
Cosine similarity	89	87	85	86
Jaccard similarity	85	86	85	85.5
Euclidean distance	84	86	82	84

6.3 *Recommendation system*

Cosine similarity has been used to find similar users. The performance of cosine similarity is compared with other similarity measures – Jaccard similarity and Euclidean distance.

From Table 3, it is inferred that cosine similarity has 5% of accuracy outperforming the other similarity measures like Jaccard similarity and Euclidean distance.

Figure 7 Graph of performance measures for proposed and existing systems (recommendation) (see online version for colours)



From Figure 7, it is observed that the proposed system gives 5% accuracy improvement over existing system.

These results show that the proposed system has outperformed the existing systems.

7 Conclusions

This paper has been developed to predict depression and recommend a suitable therapy using BERT model for prediction and collaborative filtering with cosine similarity for recommendation. BERT model was chosen because it is purely bi-directional, it is trained on the BooksCorpus (800M words) and Wikipedia (2,500 M words), learns [SEP], [CLS] and sentence A/B embeddings during pre-training, it was trained for 1M steps with a batch size of 128,000 words, it chooses a task-specific fine-tuning learning rate which performs the best on the development set. BERT model thereby helped to achieve a better accuracy on classifying sentences into depressed or non-depressed. User-user based collaborative filtering has been implemented to recommend an appropriate therapy for different cases of depression. BERT and collaborative filtering combined together has given an accuracy of 87%. In this future, the motive is to scrape comments from more than one sub-reddit as well as including depression comments from different forums to achieve greater accuracy.

With respect to improving the recommender system, the size of the dataset can be increased to achieve a better accuracy and a better efficiency.

References

Bayer, I. et al. (2016) *A Generic Coordinate Descent Framework for Learning from Implicit Feedback*.

Cheng, H.T. et al. (2016) *Wide & Deep Learning for Recommender Systems*, pp.7–10, DOI: 10.1145/2988450.2988454.

Du, J. et al. (2018) 'Extracting psychiatric stressors for suicide from social media using deep learning', *BMC Medical Informatics and Decision Making*, July, Vol. 18, No. 2, p.43, DOI: 10.1186/s12911-018-0632-8.

Hung, G. et al. (2015) 'A smartphone-based personalized activity recommender system for patients with depression', *EAI Endorsed Transactions on Cognitive Communications*, Vol. 2, DOI: 10.4108/eai.14-10-2015.2261655.

Islam, M.R. et al. (2018) 'Depression detection from social network data using machine learning techniques', *Health Information Science and Systems*, Vol. 6, No. 8, DOI: 10.1007/s13755-018-0046-0.

Jamil, Z. et al. (2017) 'Monitoring tweets for depression to detect at-risk users', in *Proceedings of the Fourth Workshop on Computational Linguistics and Clinical Psychology – From Linguistic Signal to Clinical Reality*, Association for Computational Linguistics, Vancouver, BC, pp.32–40.

Kamal, M. et al. (2020) 'Predicting mental illness using social media posts and comments', *International Journal of Advanced Computer Science and Applications*, Vol. 11, pp.607–613, DOI: 10.14569/IJACSA.2020.0111271.

Kumar, A. et al. (2019) 'Anxious depression prediction in real-time social data', *International Journal of Engineering and Advanced Technology (IJEAT)*, August, Vol. 8, No. 6S2.

Ramalingam, D. et al. (2019) 'Study of depression analysis using machine learning techniques', *International Journal of Innovative Technology and Exploring Engineering (IJITEE)*, May, Vol. 8, No. 7C2, ISSN: 2278-3075.

Ravi, L., Devarajan, M., Sangaiah, A.K., Wang, L. and Subramaniyaswamy, V. (2021) 'An intelligent location recommender system utilising multi-agent induced cognitive behavioural model', *Enterprise Information Systems*, Vol. 5, No. 2, pp.1376–1394, Taylor & Francis.

Rohani, D.A. et al. (2020) *MUBS: A Personalized Recommender System for Behavioral Activation in Mental Health*, pp.1–13, DOI: 10.1145/3313831.3376879.

Shen, G. et al. (2017) 'Depression detection via harvesting social media: a multimodal dictionary learning solution', *IJCAI*, pp.3838–3844, ScholarBank at NUS Repository, https://doi.org/10.24963/ijcai.2017/536.

Thorstad, R. et al. (2019) 'Predicting future mental illness from social media: a big-data approach', *Behavior Research Methods*, Vol. 51, No. 4, pp.1586–1600, DOI: 10.3758/s13428-019-01235-z, PMID: 31037606.

Appendix

Experiment was conducted on the depression data available in Kaggle site. The data set contains 5,000 records.