Depression Detection by Analysing Social Media Post of User

Kavita S. Kharat, Shraddha Jedgule, Pratiksha Bhokate, Mukta Jamage

Pimpri Chinchwad College of Engineering, Nigdi, Pune, India

Abstract: *In recent times the problem of early depression detection is one of the most important challenges in psychology. Depression is a big mental health issue for people world-wide. There is no bound of age, gender and race for depression. For people it is easier to share their thoughts online on social networking sites than sharing thoughts in person. The main purpose of this paper is to detect depression in early stage and avoid help to avoid self-harm or letting close persons know about it. Depression is detected based on users posts on social media. Facebook, Instagram, Twitter are some mostly used social media sites for sharing thoughts. According to the report generated in 2019 estimates, about 280 million people worldwide have depression, including 5 percent of the world's adults and 5.7 percent of adults above the age of 60. Large amount of data is generated daily from social media posts. This data is collected by ubiquitous computing devices like smartphones, tabs, etc. Researchers are using this social media data for research like sentiment analysis, to spot trends, market mood analysis, etc. In this paper, we've got proposed a despair evaluation and suicidal ideation detection system, for predicting the suicidal acts that supported the volume of despair.[11] The gift examines ambitions to make the most device gaining knowledge of strategies for detecting a likely depressed Social Media person in his/her Posts. We leverage large-scale datasets with tweets to accomplish sentiment analysis with the aid of machine learning algorithms and a deep learning model, BERT for sentiment classification. With this model we will evaluate persons mental health.*

Keywords: Machine Learning, Natural Language Processing, BERT Algorithm, Depression, Classification, Social Media Post, Decision Tree, Sentiment Analysis.

I. INTRODUCTION

Depression is a common mental disorder characterized by persistent feelings of sadness, hopelessness, and a lack of interest in daily activities. Early detection and treatment of depression is essential for reducing the impact and severity of the disorder, but conventional methods of detection such as clinical interviews and questionnaires have limitations. With the increasing use of social media, it has become possible to analyse large amounts of data about an individual's behaviour and emotions. The content of social media posts has been shown to contain valuable information about an individual's mental health, including depression.[6]

- The median age of depression onset is 32.5 years old.
- The prevalence of adults with a major depressive episode is highest among individuals between 18 and 25.
- 11% of adults who report two or more races have experienced a major depressive episode in the past year.
- 9% of women have depression.
- 5.5% of men have depression.

II. LITERATURE REVIEW

The analysis of customer behaviour on social networks is rapidly evolving, and instrumental opportunities for this are actively being developed. One effective method involves the use of computational linguistics strategies to study social media posts.

1. The paper proposes a model based on record analysis to detect depression in individuals. Data is collected from popular social media sites such as Twitter and Facebook, and machine learning is used to process the scraped data from SNS users. Natural Language Processing (NLP) is utilized, with classification performed using Support

Vector Machine (SVM) and Naïve Bayes algorithms to detect depression in a more convenient and efficient manner.[1]

2. The study uses Natural Language Processing (NLP) techniques to develop a depression detection algorithm for the Thai language on Facebook. Facebook serves as a platform for people to share their opinions, emotions, and life events.[2]

3. The paper analyses health-related tweets to detect depression and anxiety by using Multinomial Naive Bayes and Support Vector Regression (SVR) algorithms as a classifier on a combination of tweets.[3]

4. The paper presents a method to detect an individual's level of depression by examining and extracting emotions from text using emotion theories, machine learning techniques, and natural language processing strategies on various social media platforms.[4]

5. The paper aims to conduct emotion analysis, specifically focusing on depression, using natural language processing on Twitter feeds. The approach involves labelling individual tweets as neutral or negative based on a curated word list to detect signs of depression. A support vector machine and Naive-Bayes classifier were used for classification prediction. The results were presented using primary category metrics, including F1-score, accuracy, and confusion matrix.[5]

6. Yates et al. utilized a neural network model to demonstrate the potential risks of self-harm and depression, based on data from Reddit and Twitter, and found that this diagnostic approach achieved high accuracy. The authors propose that these methods could be utilized for extensive research into mental health, as well as for clinical treatment.[8]

7. O'Dea et al. conducted research on the potential of Twitter as a tool for detecting mental health conditions, including depression and suicidal tendencies in the general population. Their study revealed that it is possible to identify the degree of concern expressed in suicide-related tweets, using both human coders and an automated machine classifier, without plagiarism.[10]

Depression has a significant impact on an individual's physical and mental health. People with depression may experience changes in appetite, causing unintended weight loss or gain, as well as physical aches and pains such as joint or muscle pain, breast tenderness, and headaches. Despite its high treatment rate, nearly two out of three people with depression do not seek proper treatment.

In this study, a machine learning approach is proposed to detect the level of depression by analysing social media posts from Facebook and Twitter. The idea is that users often express their emotional state through their posts and tweets, which can provide insight into their mental state. To analyse this data, machine learning classification techniques are used.

III. SENTIMENT ANALYSIS

Social media is a virtual community where people can connect and share intimate thoughts. Facebook and Twitter, two of the most popular social networking sites, have a large number of users who share their lifestyle, thoughts, and emotions. The proposed model utilizes Beautiful Soup to collect tweets and manual data collection from Facebook users with their permission. The collected data is processed and trained with signature keywords using NLP, BERT algorithm to determine if the posts are positive, negative, or neutral.

Agarwal et al. proposed a 3-way model for categorizing sentiments in three classes: positive, negative, and neutral. The models tested were unigram, feature-centred, and tree kernel-based. The tree kernel-based model performed the best and concluded that features combining word polarization and pos tags were the most substantial.

Several researchers have attempted to classify public beliefs about movies, news, etc. from Twitter posts. V.M. Kiran et al. utilized IMDB and Blippr databases for movie domain analysis, while Davidov et al. used Twitter user-defined hashtags and a K-Nearest Neighbor approach. Tagging is a common way to sort web content, and scholars have developed various methods and procedures. Xiance et al. offered a flexible tag-LDA model, while Krestel et al. proposed a personalized tag recommendation method.

Pre-trained language models such as ELMo, OpenAI GPT, and BERT have proven valuable in NLP and allow for transfer learning. Zubair et al. introduced a technique enhanced by lexicons and emojis, modifiers, and domain-specific

terms. Traditional methods of designing features have reached a performance bottleneck, while pre-trained language models save time and achieve the same result quickly.

IV. BERT ALGORITHM

BERT, or Bidirectional Encoder Representations from Transformers, is a deep learning model designed to pre-train bidirectional representations from unlabelled text. Bert Algorithm is developed by Google in 2018. It utilizes both left and right context to generate pre-trained representations, which can then be fine-tuned with a single additional output layer to solve various NLP tasks. For this project, a BERT-based model was developed, with a focus on the binary classification of social media comments. The English language uncased version of BERT was selected as it doesn't require information about the case of the words in the comments.

BERT (Bidirectional Encoder Representations from Transformers) is a deep learning model architecture for natural language processing tasks, such as question answering, sentiment analysis, and text classification. The following is a high-level diagram of the BERT architecture:

Input Embedding: The input sequence is tokenized and each token is converted into a dense vector representation (also known as an "embedding") through a lookup table.

Self-Attention Mechanism: BERT uses self-attention mechanisms to weigh the importance of each token in the input sequence and generate a contextual representation for each token.

Transformer Blocks: The self-attended representations are processed through multiple Transformer blocks, which apply multi-head self-attention and feed-forward layers to further refine the representations.

Pooler Layer: The final hidden state of the first token (often [CLS]) is passed through a fully connected layer to produce a fixed-length representation, which is used as the output representation for various NLP tasks.

Output Layer: In fine-tuning, an additional output layer is added on top of the pooler layer to generate predictions for the target task.

It's important to note that BERT is trained on large amounts of unannotated text data, allowing it to learn rich contextual representations that can be fine-tuned for specific NLP tasks.
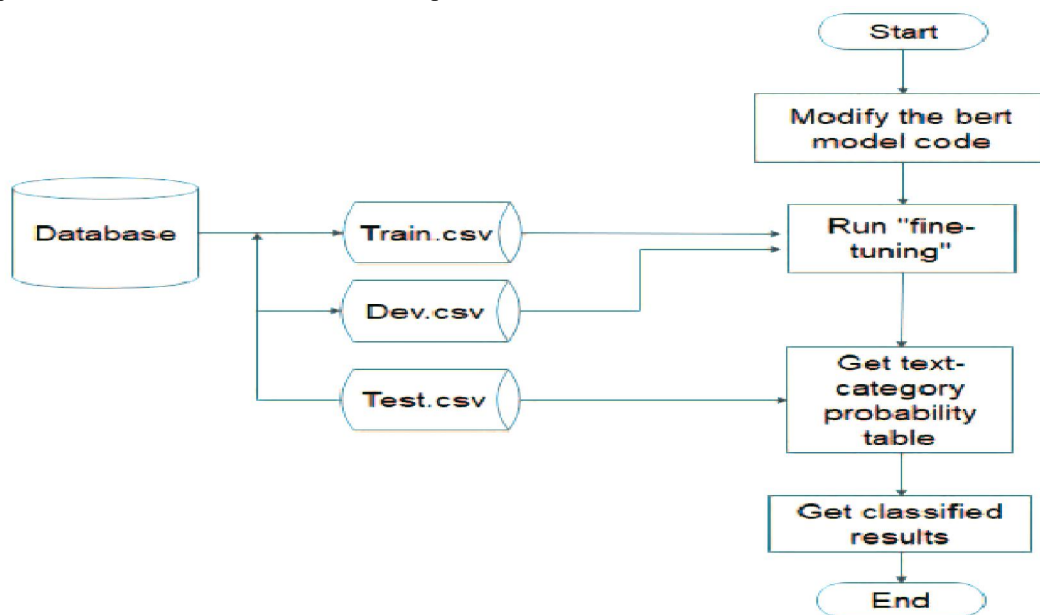

Fig.1 BERT Algorithm

1. Relevant Mathematics Associated with The Project:

System Description:

S= I, O, F, DD, NDD, Failure, Success

Where,

S=System

I= Input

O=Output

F=Failure

S=Success

I is Input of system

Input I = set of Inputs

Where,

I= {Users Social media posts}

F is Function of system

F = set of Function

Where,

F1= {Input Dataset}

F2= {Json to CSV Conversion}

F3={Pre-processing}

F4={Cleaning}

F5= {Train test split}

F6= {Sentiment Dictionary}

F7= {Classifier (BERT Algorithm)}

F8={Tokenization}

O is Output of system

Output O1= {Depression detection}

•Success Conditions: Product working Smoothly. depression detection successfully.

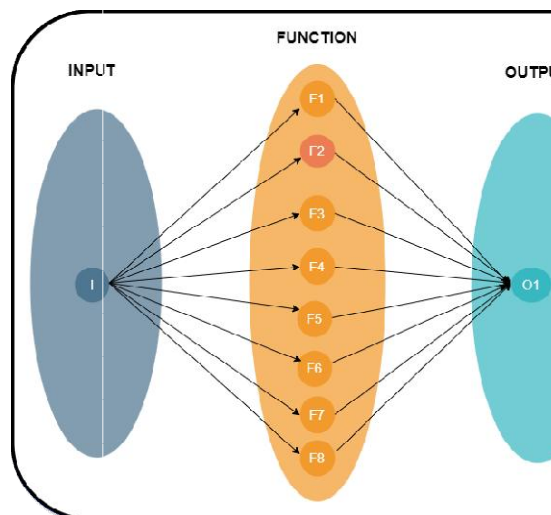•Failure Conditions: if internet connection Unavailable.

Fig.2: Venn Diagram

Where,

I = {Users Social media posts}

F1= {Input Dataset}

F2= {Json to CSV Conversion}
F3={Pre-processing}
F4={Cleaning}
F5= {Train test split}
F6= {Sentiment Dictionary}
F7= {Classifier (BERT Algorithm)}
F8={Tokenization}
Output O1 = {Depression detection}

Compared to Naive Bayes, BERT has several advantages:
1)BERT can capture more complex language patterns and relationships between words, while Naive Bayes assumes independence between words.

2)BERT is pre-trained on a large corpus of text, allowing it to perform well on a wide range of NLP tasks without requiring significant additional training or customization.

3)BERT can handle out-of-vocabulary words, while Naive Bayes struggles with rare or unknown words.

Bert Algorithm give high accuracy as compared to other algorithms which is 87.3%.

V. SYSTEM ARCHITECTURE

Depression is a significant issue in mental and physical health. One approach to addressing this issue is to examine a person's behaviour attributes, which can be obtained from social networking websites like Facebook, Twitter, and Instagram. The use of social media platforms is increasing, particularly among younger generations, as people express their feelings, daily activities, opinions, and other personal information on these sites. Thus, social media can be used as a screening tool to predict depression levels by analysing a person's experiences, socialization, personality, and other behavioural attributes. The goal is to extract information from social media posts and use machine learning classifiers, such as the BERT Classification Algorithm, to predict an individual's depression level based on their posts and social media activity. The process involves retrieving data from the person's social media accounts, including posts, number of followers, mentions, and reposts, assembling the posts into a single file, and performing text pre-processing on the file.
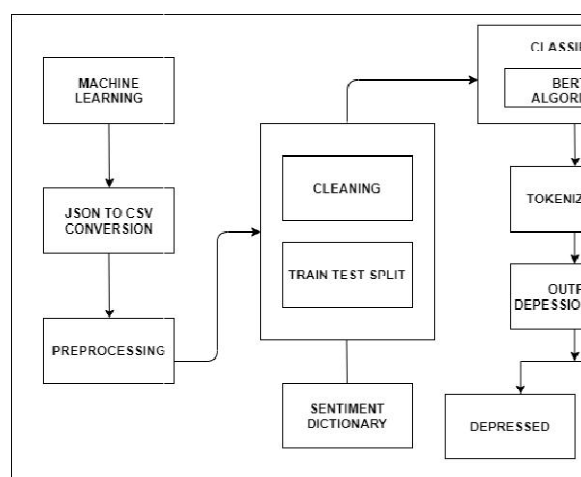

Fig.3: System Architecture

VI. EXISTING SYSTEMS

The current system uses the Naive Bayes algorithm to determine a person's depression level. The information is obtained from Facebook posts using the Facebook Graph API and pre-processed to remove missing or repetitive attributes. Pre-processing techniques such as tokenization, lowercase conversion, stemming, and phrase removal are applied. The proposed system uses a person's Facebook posts to determine if they are depressed, but it also takes into account the comments and chats they have with friends, as these may also reveal the individual's mental state. Based on these analyses, the person can be labelled as either depressed or not depressed.
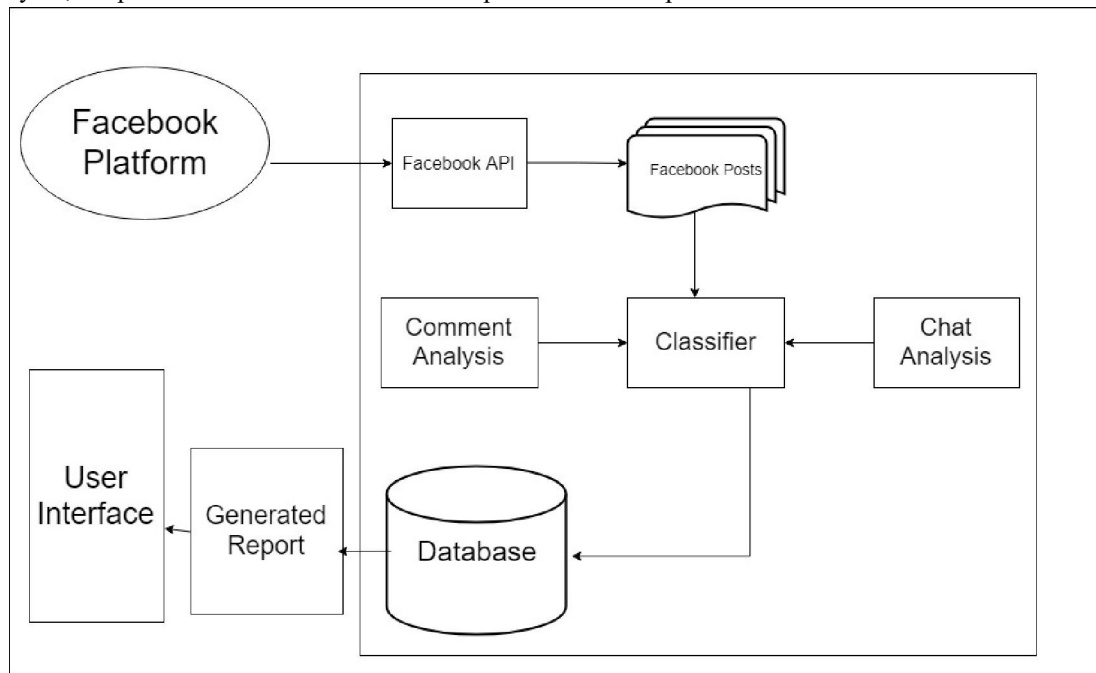

Fig 4. Existing System

VII. CONCLUSION

In conclusion, the results of this study demonstrate the effectiveness of the BERT algorithm for detecting depression in social media posts. The use of BERT has the potential to improve the early detection of depression and facilitate the development of more effective interventions. Further research is needed to validate the results and to determine the optimal pre-processing and training procedures for depression detection using BERT.

The proposed system has the potential to save lives by detecting depression in individuals and providing them with motivational posts based on their level of depression. In today's fast-paced world, where people have limited time to meet and share their thoughts and emotions, this system can play a crucial role in preventing human loss. It can also inform the person's family or friends about their situation, allowing them to support the individual in overcoming their depression. By providing early detection and support, the proposed system can help to improve mental health and well-being.

REFERENCES

[1] N. A. Asad, M. A. Mahmud Pranto, S. Afreen and M. M. Islam, "Depression Detection by Analyzing Social Media Posts of User," 2019 IEEE International Conference on Signal Processing, Information, Communication & Systems (SPICSCON), Dhaka, Bangladesh, 2019, pp. 13- 17, DOI: 10.1109/SPICSCON48833.2019.9065101.

[2] K. Katchapakirin, K. Wongpatikaseree, P. Yomaboot and Y. Kaewpitakkun, "Facebook Social Media for Depression Detection in the Thai Community," 2018 15th International Joint Conference on Computer Science and Software Engineering (JCSSE), Nakhonpathom, 2018, pp. 1-6, DOI: 10.1109/JCSSE.2018.8457362.

[3] P. Arora and P. Arora, "Mining Twitter Data for Depression Detection," 2019 International Conference on Signal Processing and Communication (ICSC), NOIDA, India, 2019, pp. 186-189, DOI: 10.1109/ICSC45622.2019.8938353.

[4] A. U. Hassan, J. Hussain, M. Hussain, M. Sadiq and S. Lee, "Sentiment analysis of social networking sites (SNS) data using machine learning approach for the measurement of depression," 2017 International Conference on Information and Communication Technology Convergence (ICTC), Jeju, 2017, pp. 138- 140, DOI: 10.1109/ICTC.2017.8190959.

[5] M. Deshpande and V. Rao, "Depression detection using emotion artificial intelligence," 2017 International Conference on Intelligent Sustainable Systems (ICISS), Palladam, 2017, pp. 858-862, DOI: 10.1109/ISS1.2017.8389299.

[6] S. Jain, S. P. Narayan, R. K. Dewang, U. Bhartiya, N. Meena and V. Kumar, "A Machine Learning based Depression Analysis and Suicidal Ideation Detection System using Questionnaires and Twitter," 2019 IEEE Students Conference on Engineering and Systems (SCES), Allahabad, India, 2019, pp. 1-6, DOI: 10.1109/SCES46477.2019.8977211.

[7] S. Grover, A. Dutt, and A. Avasthi, "An overview of indian research in depression," Indian journal of psychiatry, vol. 52, no. Suppl1, p. S178, 2010.

[8] Yates, A., Cohan, A., and Goharian, N.: Depression and self-harm risk assessment in online forums. arXiv preprint arXiv:1709.01848 (2017).

[9] I. H. Witten, E. Frank, and M. a Hall, Data Mining: Practical Machine Learning Tools and Techniques (Google eBook). 2011.

[10] O'Dea B, et al. Detecting suicidality on Twitter. Internet Interv. 2015;2(2):183–188. DOI: 10.1016/j.invent.2015.03.005.

[11] Karan Pagare, GiteshBhole, Pooja Mankar, Rucha Mahajan, Abhay Gaidhani, "DEPRESSION DETECTION BY ANALYZING SOCIAL MEDIA POST OF USER", International Journal of Creative Research Thoughts (IJCRT), ISSN:2320-2882, Volume.9, Issue 12, pp.d245-d250, December 2021.