Towards Suicide Prevention: A Natural Language Processing and Machine Learning Approach Integrated with Chatbot

Conference Paper · March 2024

DOI: 10.1109/AUTOCOM60220.2024.10486154

5 authors, including:

Anju Pillai
Amrita Vishwa Vidyapeetham

Towards Suicide Prevention: A Natural Language Processing and Machine Learning Approach Integrated with Chatbot

Manit Maini
Department of Electrical and Electronics Engineering,
Amrita School of Engineering,
Coimbatore, Amrita Vishwa Vidyapeetham, India.
Email: manitmaini@gmail.com

Pranjal Srivastava
Department of Electrical and Electronics Engineering,
Amrita School of Engineering,
Coimbatore, Amrita Vishwa Vidyapeetham, India.
pranjal.sanskaar@gmail.com

Himanshu Soni
Department of Electrical and Electronics Engineering,
Amrita School of Engineering,
Coimbatore, Amrita Vishwa Vidyapeetham, India.
himanshusoni1942@gmail.com

Hemprasanna
Department of Electrical and Electronics Engineering,
Amrita School of Engineering, Coimbatore
Amrita Vishwa Vidyapeetham, India.
Email: hemprasanna2002@gmail.com

Dr. Anju S Pillai
Department of Electrical and Electronics Engineering,
Amrita School of Engineering, Coimbatore
Amrita Vishwa Vidyapeetham, India.
Email: s_anju@cb.amrita.edu

2024 International Conference on Automation and Computation (AUTOCOM) | 979-8-3503-8272-3/24/$31.00 ©2024 IEEE | DOI: 10.1109/AUTOCOM60220.2024.10486154

Abstract—The paper presents a prediction model for identifying suicide intent in social media messages, especially on Twitter and Reddit. A huge dataset of postings, including symptoms of suicide ideation or behaviour, was analyzed using natural language processing (NLP) techniques, and characteristics were collected to build a machine learning (ML) model. The prediction algorithm was subsequently incorporated into a functioning mental health chatbot, which will give users at risk of suicide relevant resources and help. The initiative intends to improve mental health and suicide prevention efforts while also considering critical ethical factors, such as respecting user privacy and ensuring that the chatbot does not replace professional mental health treatment.

Keywords – Suicide text, Machine learning, Deep learning models, Transformer algorithm, Suicide prevention, Chatbot.

I. INTRODUCTION

The World Health Organization (WHO) reports that approximately 800,000 individuals die from suicide yearly, equating to one person every 40 seconds. Suicide ranks as the tenth leading cause of death in India. Researchers have started applying NLP methods to unstructured physician notes in electronic health records (EHR) data to identify high-risk patients. In recent times, social media platforms have become an ideal source for mining data, allowing for the extension of research in more natural settings due to the free sharing of information.

As a result, NLP-based techniques and machine learning to identify suicide ideation on social media platforms are employed [1-4]. Some approaches use handcrafted characteristics such as TF-IDF, LIWC, N-gram, Part-of-Speech (PoS), and emotions. This study introduces various approaches for detecting suicide ideation from Twitter posts [5-6]. The proposed work presents the following contributions: multiple handcrafted feature sets inspired by prior studies in this domain are created. A novel deep learning (DL) architecture that uses latent features from tweets to recognize suicide attempts is created.

Several ML algorithms use solely handcrafted and only latent features. The performance of the proposed approach surpasses the F1, F2, and True Positive Rate (TPR) baseline for both subtasks. Our findings indicate that latent features (Word2Vec) are more effective in identifying suicide attempts in tweets than handcrafted features. Finally, analyses of different ML and DL algorithms are carried out.

II. BACKGROUND

According to the WHO, about 800,000 people died by suicide in 2016, a tragic event that significantly impacted their families and communities. It is the second highest cause of unnatural death in those aged 15 to 29. With the rise of social media and its open sharing of information, collecting data from these platforms has become an organic approach to expanding research in more authentic settings. The employment of modern NLP techniques and DL architectures to construct an effective big data system for identifying suicidal intent from social media data is vital [7-9]. It is also encouraging to see chatbots being used to detect suicidal intent and give care to individuals in danger. The pre-processing steps for the text data seem comprehensive and necessary, given the unstructured nature of social media data [10-12]. The usage of Word2Vec and GloVe methods to generate text representations is also a solid strategy, and it is interesting to note that the bespoke Word2Vec embedding beat the pre-trained GloVe embedding.

The evaluation of the different models based on accuracy, precision, recall, and F1 score is also a good approach, and it is good to see that the transformer models outperformed the other models for all metrics [13-14]. It is also encouraging to note that the final model chosen for the work is ELECTRA, which has the greatest F1 score, accuracy, and recall score.

Finally, including the suicidal text detection model in a chatbot that uses generative methods to generate conversational responses and retrieval-based ways to offer suitable responses to suicide messages is a great way to put the model to use. It is an excellent use case of NLP and deep learning in improving mental health and saving lives [15-17].

III. System Description

A. Data Collection and Wrangling

Data collection involves gathering diverse data sets, while data wrangling refers to eliminating errors and merging intricate data sets to enhance their accessibility and facilitate analysis. In this work, the suicide and depression detection datasets were obtained from Kaggle [18].

B. Data Cleaning

From the dataset, removal of irrelevant words, empty rows, and various non-needful parameters were removed.

C. Data Exploration

The dataset is divided into three sets with an 8:1:1 ratio: train, test, and validation. Following that, the training dataset is examined.

D. Representation Learning

Turning raw text input into a format that computers can easily interpret is known as representation learning in NLP. Word embeddings are a common way to encode words as dense vectors, allowing machine learning [20] or deep learning classifiers to capture semantic and syntactic information and detect words in comparable settings.

E. Model Building and Evaluation

Multiple models were created and evaluated for their ability to classify text as suicidal or non-suicidal. The problem statement involves predicting a binary outcome, specifically whether the text is indicative of suicide or not.

F. Chatbot Integration

Young people may be reluctant to seek professional help for their mental health issues due to discomfort with confiding in a stranger and sharing personal information. This can lead to early signs of suicidal behavior being missed. Chatbots can play a role in addressing this issue by providing support and detecting suicidal behavior. As online conversational agents, chatbots can build connections with young people through screen-based conversations, which can help overcome the fear and reluctance to seek help. The system description can be found in Fig.1.

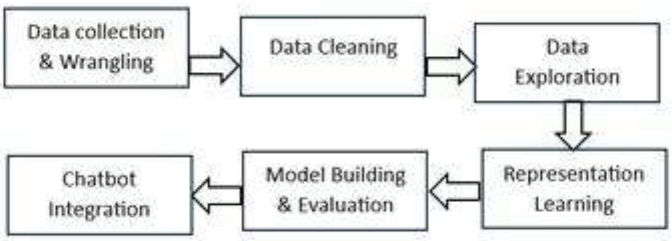

Fig. 1. Block Diagram

IV. Methodology

A. Data Collection

The Suicide and Depression Detection dataset [1] was gathered from Kaggle and comprised postings from the social media network Reddit. The dataset is divided into two columns, as shown in Fig. 2, where "text" represents the post content and "class" represents the post label. To create this dataset, 232,074 posts were taken from two subreddits, SuicideWatch and Teens. Suicide Watch postings were categorized as suicidal, whereas posts from teens were classed as non-suicidal. The classes are spread evenly, as seen in Fig. 3, where each class has 116,037 rows, representing 50 percent of the dataset.

	text	class
0	Ex Wife Threatening SuicideRecently I left my ...	suicide
1	Am I weird I don't get affected by compliments...	non-suicide
2	Finally 2020 is almost over... So I can never ...	non-suicide
3	i need helpjust help me im crying so hard	suicide
4	I'm so lostHello, my name is Adam (16) and I'v...	suicide

Fig. 2. Dataset Content

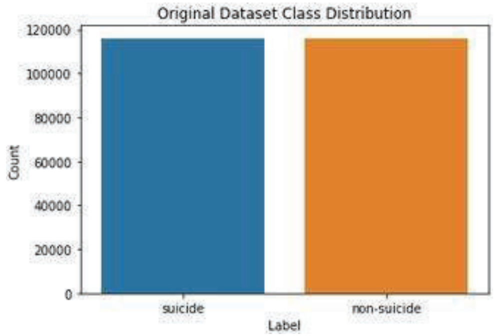

Fig. 3. Original Dataset Distribution

B. Text Preprocessing

Text data must be preprocessed and converted into acceptable forms for further model construction. Because social media data is more unstructured, it needs more tailored preparation and cleaning techniques. As a result, the data was cleaned using the steps shown in Fig. 4.

- Removing characters with diacritical marks: To ensure comparable concepts are treated as the same word, accented letters are removed, and this also serves as a step to reduce the dimensionality of the dataset.

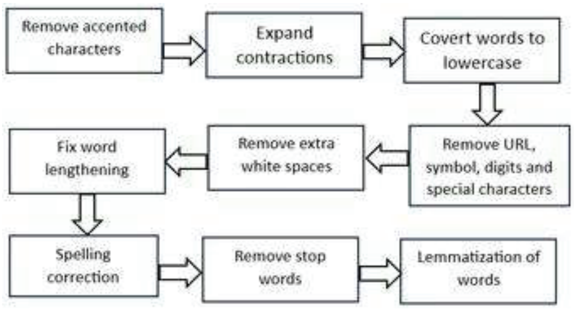

Fig. 4. Text Preprocessing workflow

This is beneficial as it helps to avoid using high-dimensional vectors, which can make the model training process more computationally complex. For instance, "cafe´" and "cafe" have the same meaning and are treated as the same word.

- Expand Contractions: To simplify the representation of text and reduce the complexity of the model training process, contractions are expanded to their original words in text standardizing. Contractions are abbreviated versions of words that leave out certain letters, such as "I've" instead of "I have." Expanding contractions can also help reduce the size of the vocabulary used in the dataset.

- Transform all letters in the text to lowercase: To avoid treating the same word differently due to its case, such as "Boy" and "boy," it is essential to convert all words to lowercase. This is an important step in standardizing the text data and reducing the dataset's dimensionality.

- Eliminate web addresses, symbols, numbers, unique characters, and additional spaces: Irrelevant characters have been eliminated from the text data as they do not carry any significant meaning for the model in this use case. The removal process takes place from left to right in the text data.
- Fix word lengthening: Longer words are often the result of incorrect repetitions of letters, such as "foooood" instead of "food." To avoid misleading information in the prediction model, unnecessary characters should be removed. English words typically have up to two repeated characters, so any additional repetitions should be corrected.
- Spelling correction: The Symspell algorithm is utilized for rectifying spelling errors. It is derived from the Sym- metric Delete Spelling Correction algorithm and performs faster than conventional algorithms. Symspell also has the capability to automatically correct spelling errors in multi-word input strings, which is known as compound automatic spelling correction. This feature is useful in rectifying errors in cases where two words are merged without a space between them. For instance, "havejust" is corrected to "have just" by adding the missing space to indicate the boundary between the two words.
- Remove stop words: Stop words are frequent words that have no meaning or significance, such as "a," "the," and "no." However, the word "not" is an exception and is not removed as a stop word since it conveys a negative sentiment that can be relevant in classifying suicidal intent in posts.

Lemmatization: The process of converting inflected words to their basic form is known as lemmatization. The goal is to combine a word's many forms into a single root form because they all imply the same thing. For example, when given into the model, the terms "shoot," "shoots," and "shooting" all mean the same and are represented by similar terms. Unlike stemming, which removes the suffixes of words to derive their base form, and lemmatization considers the grammatical context of the word to ensure that the root form is correct. For example, the lemma of "better" is "better," but stemming cannot recognize this; hence lemmatization is preferred over stemming.

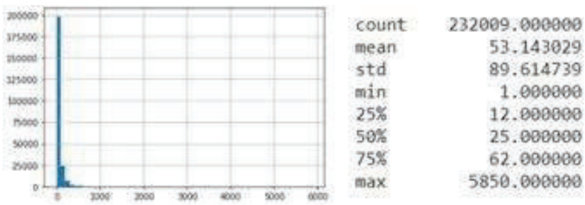

Fig. 5. Histogram of Distribution of Word Count in Posts

C. Data Cleaning

- Remove Unnecessary Words: After preprocessing the posts, an initial data analysis was conducted to determine the frequency of each word. We observed that "filler" appeared 55,442 times, making it an outlier among the 30 most common words. Upon investigation, it was determined that it was a meaningless token likely introduced during data collection and therefore removed from all posts.
- Preprocessing: After performing text preprocessing to eliminate irrelevant words, it was discovered that approximately 60 text boxes were empty and did not contain any words. As a result, these lines were removed from the dataset.
- Remove Outlier Rows with Higher Word Count: The word count of the preprocessed post was obtained, and it was observed that there were a significant number of outliers in the word count distribution, as shown in Fig. 5. The maximum word count of 5,850 was well above the 75th percentile of 62 words. To optimize subsequent model training, a threshold was set to remove processed posts with a word count above the 75th percentile of 62 words. To reduce training time, the dataset was split, and the model was focused on training on shorter texts.
- Final Dataset: The processed text fields were used to create a final cleaned dataset that consists of 174,436 rows, as illustrated in Fig. 6, which shows the class distribution of this cleaned dataset, which has a ratio of approximately 4:6 for suicidal and non-suicidal text, compared to the original dataset's 5:5 ratio (shown in Fig. 3). The decrease in the suicidal-to-non-suicidal ratio may be attributed to the removal of longer suicidal messages during the cleaning process. Although the class distribution is somewhat imbalanced, it is a common issue in classification tasks and does not affect the model's performance.

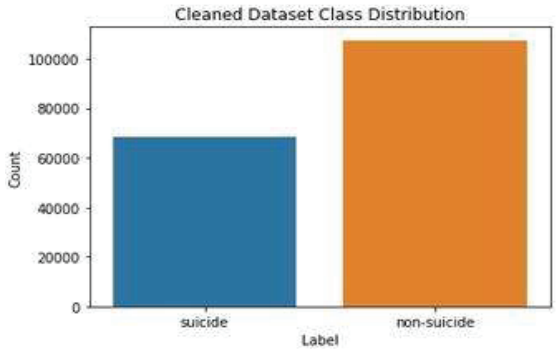

Fig. 6. Cleaned Dataset Distribution

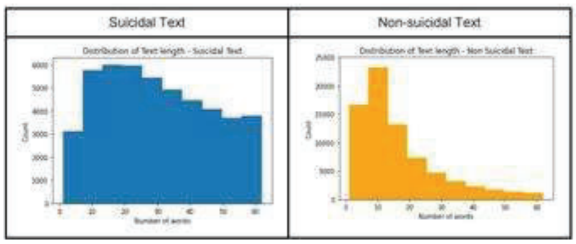

Fig. 7. Histogram of Distribution of Word Count in Posts

D. Data Exploration

In data exploration, the dataset is split into training, test, and validation sets 2 by a ratio of 8:1:1. Then, a data exploration on the train set was run and zoomed in further on each class.

The top 100 most common terms were identified and presented in the form of word clouds. The regularly used verbs such as “feel,” “know,” and “want” are found in word clouds for both suicidal and non-suicidal writings. Surprisingly, suicide texts are rife with negative terms such as “kill,” “die,” “hate,” “end,” and “help,” non-suicidal messages, on the other hand, were filled with more neutral and positive terms like “good,” “people,” “friend,” “girl,” and “guy.” This finding implies that suicide texts communicate more negative emotions than non-suicidal texts. The histograms shown in Fig. 7 illustrate the distribution of average text length in posts. The distribution for suicidal texts is more evenly spread out, while non-suicidal texts have a right skew. This indicates that non-suicidal messages are generally shorter than suicidal ones. These findings suggest that individuals who are suicidal have more emotions to convey and tend to write longer posts, which aligns with the initial hypothesis.

A score depicting the polarity was explored, which lies between -1 to 1, with -1 signifying negative emotion, 0 neutral mood, and +1 positive sentiment. The histograms in Fig. 8 show the distributions of polarity scores achieved. Both distributions have a bell-shaped distribution. Suicidal texts have more data points on the left side of the distribution, thus indicating greater negative emotion in the posts. Most non-suicidal literature, on the other hand, is categorized as neutral. This data supports our initial hypothesis that suicide writing has a higher negative emotion.

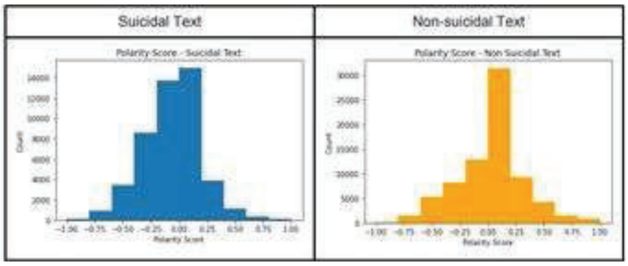

Fig. 8. Polarity Distribution for Suicidal and Non- Suicidal text

A bigram is a pair of successive words in a text. With the use of bigrams, we can reveal insights into the dataset. Examples of suicidal bigrams include “want to die,” “want to kill,” and “commit suicide,” whereas non-suicidal bigrams include “year old” and “good friend.” This implies that people who write suicide messages are frequently experiencing strong negative feelings, emphasizing the critical need for immediate assistance. The bigrams formed from the initial sequence of words are not our focus because they were generated after text preparation.

E. Representation Learning

In natural language processing, representation learning refers to methods for translating raw textual input into a computationally efficient representation suited for machine learning tasks. Because natural language documents are often unstructured, with numerous granularities, tasks, and domains, NLP performance can be problematic. To represent words as a dense vector, word embeddings are widely utilized. They enable the extraction of semantic and syntactic information from words as well as the recognition of related words in various contexts by machine learning classifiers. Because they are more effective than random initialization, Word2Vec, GloVe, and fastText models are frequently employed for deep learning model initialization.

Without pre-trained word embeddings loaded into the embedding layer, our model has to learn the word embeddings from scratch during training, which might be difficult for the reasons listed below. Our training data might be minimal, with a high proportion of unusual words. Due to a lack of knowledge, the embedding learned may not appropriately represent these words. Second, learning the embeddings from the start would greatly lengthen our training procedure because of the enormous number of trainable parameters.

Word2Vec: Word2Vec embeddings can give a dense vector representation of words while still capturing semantic information. Its success arises from grouping vectors of similar words together and estimating the meaning of the term based on its appearance in the text. Word connections with other terms in the corpus are provided by the estimations.

Glove: Another unsupervised learning approach for obtaining vector representation for words is Global Vectors for Word Representation (GloVe)3. It is a count-based model rather than a predictive model, and it combines both local and global context by employing the word co-occurrence matrix as opposed to Word2Vec.

F. Model Building

Several models were constructed in this part and assessed their efficacy in categorizing suicidal text. The proposed model attempted to forecast suicidal intent as a binary variable, suicide or non-suicide. The five models created are machine learning, deep learning, and transformers. They are Logistic Regression (Logit), Convolutional Neural Network (CNN), Long Short-term Memory (LSTM), Bidirectional Encoder Representations from Transformers (BERT), and Efficiently Learning on Encoder that Classifies Token Replacements Accurately (ELECTRA).

1) Logistic regression: The basic supervised approach for classification in NLP is logistic regression. Furthermore, it is closely related to neural networks since it may be considered logistic regression classifiers piled on each other. Logistic regression is trained to serve as a baseline model for comparison.

2) Convolution Neural Network: The order of words affects how a sentence is understood as our study attempts to categorize text data. We used several deep learning techniques to analyze the sequences in our training data because the Logistic Regression model earlier was unable to capture this feature. CNN is an effective technique [19] for categorizing text data since it can achieve good prediction accuracy using fewer computing resources. As a result, we tested the CNN model as one of the deep learning algorithms in the proposed work.

3) Long Short-term Memory Network: CNN models are good for text categorization but struggle with long-range dependencies. RNN models, like LSTM, address this issue but suffer from vanishing gradients. LSTM overcomes vanishing gradients and can handle longer sequences by utilizing memory cells and gating mechanisms to retain and regulate information flow.

4) Transformers: Transformers are encoder-decoder models built on attention mechanisms, allowing for better word representation in NLP tasks. They excel in capturing long-range interactions and outperform models like CNN and LSTM.

Best Models	Accuracy	Recall	Precision	F1 Score
Logistic Regression	0.9111	0.8870	0.8832	0.8851
Convolutional Neural Network (CNN)	0.9285	0.9013	0.9125	0.9069
Long Short-Term Memory Network (LSTM)	0.9260	0.8649	0.9386	0.9003
BERT	0.9757	0.9669	**0.9701**	0.9685
ELECTRA	**0.9792**	**0.9788**	0.9677	**0.9732**

Fig. 9. Selected Model

LSTM models have bidirectional variations to capture long-term dependencies but still face issues with vanishing gradients. Classical models cannot be easily parallelized due to their sequential structure, while transformers can parallelize tasks, reducing processing time. Usually, huge datasets are used for unsupervised pretraining before transformers are customized for prediction tasks. The purpose of this method is to fine-tune frequently used transformers.

5) BERT: In 2018, Google introduced BERT, also known as Bidirectional Encoder Representations from Transformers, which uses a transformer's encoder structure for language modeling.

6) ELECTRA: One of Google's most recent pre-trained transformer models, ELECTRA, which stands for Efficiently Learning an Encoder that Accurately Classifies Token Replacements, was announced in 2020. More datasets and a longer training time are needed for most BERT-augmented transformer models, including RoBERTa and XLNet. However, ELECTRA beats the aforementioned models on the selected benchmark datasets, using only a fifth of the computing resources.

Model Selection: Fig. 9 presents the important outcomes of the model's run. The specifically trained Word2Vec embeddings are used in the best versions for Logistic Regression, CNN, and LSTM, while BERT and ELECTRA are fine-tuned on our dataset.

The transformer models outperformed the other models on all measures. The F1, accuracy, and recall scores of ELECTRA are the highest. Following closely after is BERT, which has the greatest accuracy score. Even though the performance of these two models is comparable, ELECTRA has been shown to perform better and quicker than BERT, with the suggested RTD directly addressing the problem of employing MLM with BERT. As a result, ELECTRA was chosen for the proposed model.

7) ChatBot Integration

With the rising rates of juvenile suicide in India, there is a need for increased support. The stigma surrounding mental health often prevents teenagers from seeking help. Chatbots can play a crucial role in early detection and support by providing an anonymous and non-judgmental platform for dialogue. Chatbots help overcome the barriers of discomfort and mistrust by engaging with young people behind a screen, offering valuable assistance to those in need.

Implementation: The chatbot is built using transformers and leverages DialoGPT, a pre-trained dialogue response generation model. However, the pre-trained DialoGPT cannot respond appropriately to suicide messages. A retrieval-based component was added to address this with consoling messages and local helpline information for users expressing suicidal intent, as in Fig.10.

Limitations: Chatbots have increased use in various industries but need help providing compassionate and realistic responses. In mental health care, this can lead to inappropriate replies and privacy concerns. While chatbots cannot replace healthcare, they have the potential to contribute to addressing rising suicide rates, especially among young people. Further research is needed before widespread adoption.

Fig. 9. Sample output from the Chatbot

V. Conclusion And Discussion

The frequency of depression among Indian youngsters is still a severe societal concern, and early identification is critical for prompt care. As support, the proposed work identifies suicidal writing in social media postings. A model was constructed that earned the top F1 score of 0.9732 using ELECTRA. Finally, the detection model was integrated into a chatbot to reach people in need. Moving forward, we intend to improve the efficacy of the detection algorithm and chatbot while also reaching out to more people in need.

Building Multilingual Chatbots: Mental health care chatbot research participants highlighted the need for multi-language support to enhance user engagement. Our current chatbot only supports English, but expanding to other languages would reach a wider audience. This can be achieved by training algorithms on conversational data from different languages or using pre-trained models like DiGPTame. Multi-language support will improve accessibility and impact on a global level. However, this improvement is only relevant to commonly spoken languages because gathering significant conversational data for dialects remains difficult.

Integrating chatbot with social media platform: Social media consumption is on the rise, with an estimated 4.41 billion users by 2025. During the COVID-19 pandemic, social media provided support to isolated individuals. Integrating our chatbot with Facebook, which has 2.91 billion monthly active users, could improve its visibility and usage. Facebook flags content related to suicide, but privacy concerns arise when sharing data with third parties. However, the chatbot avoids involving third parties and can provide quicker support. Therefore, integrating the chatbot on Facebook would offer rapid assistance while reducing privacy discussions.

References

[1] Nasser-Eddine Rikli, "An ensemble deep learning technique for detecting suicidal ideation from posts in social media platforms," 2021, Elsevier. B.V. Volume 34, Issue 8, Part A, 2021.

[2] Castillo-Sa´nchez, G., Marques, G., Dorronzoro, E. "Suicide Risk Assessment Using Machine Learning and Social Networks: a Scoping Review," Springer, J Med Syst 44, 205, 2020.

[3] Ning Wang, Fan Luo, Yuvraj Shivtare, Varsha D. Badal, K.P D Subbalakshmi, R. Chandramouli, Ellen Lee, "Learning Models for

Suicide Prediction from Social Media Posts," CLPsych 2021, 29 Apr 2021.

[4] T. Agarwal, A. Dhawan, A. Jain, A. Jain, and S. Gupta, "Analysis and Prediction of Suicide Attempts," 2019 International Conference on Computing, Power and Communication Technologies (GUCON), 2019, pp. 650-665.

[5] H. Sakib, M. Ishak, F. F. Jhumu and M. A. Ali, "Analysis of Suicidal Tweets from Twitter Using Ensemble Machine Learning Methods," 2021 International Conference on Automation, Control, and Mechatronics for Industry 4.0 (ACMI), 2021, pp. 1-7, doi: 10.1109/ACMI53878.2021.9528252.

[6] A. Seal, R. Bajpai, J. Agnihotri, A. Yazidi, E. Herrera-Viedma and O. Krejcar, "DeprNet: A Deep Convolution Neural Network Framework for Detecting Depression Using EEG," in IEEE Transactions on Instrumentation and Measurement, vol. 70, pp. 1-13, 2021, Art no. 2505413, doi: 10.1109/TIM.2021.3053999.

[7] H. Patel and N. Soni, "Machine Learning Based Approach For Prediction Of Suicide Related Activity," 2021 2nd International Conference on Smart Electronics and Communication (ICOSEC), 2021, pp. 967-972, doi: 10.1109/ICOSEC51865.2021.9591836.

[8] R. Kancharapu, A. SriNagesh and M. BhanuSridhar, "Prediction of Human Suicidal Tendency based on Social Media using Recurrent Neural Networks through LSTM," 2022 International Conference on Computing, Communication and Power Technology (IC3P), 2022, pp. 123-128, doi: 10.1109/IC3P52835.2022.00033

[9] Syed Tanzeel Rabani, Qamar Rayees Khan, Akib Mohi Ud Din Khanday, "Multi-Class Suicide Risk Prediction on Twitter Using Machine Learning Techniques," Conference: 2020 2nd International Conference on Advances in Computing, Communication Control and Networking (ICACCCN).

[10] N. Suguna, G Indirani, "Predicting Suicidal Behaviour Analysis Using Deep Learning Techniques," May 2022, IJIRT, Volume 8 Issue 12, ISSN: 2349-6002

[11] Yaakov Ophir, Refael Tikochinski, Christa S. C. Asterhan, Itay Sisso & Roi Reichart, "Deep neural networks detect suicide risk from textual Facebook posts", (2020).

[12] Maryam Mohammed Aldarwish, Hafiz Farooq Ahmed, "Predicting Depression Levels Using Social Media Posts," 2017 IEEE 13th International Symposium on Autonomous Decentralized Systems.

[13] Glen Coppersmith, Ryan Leary, Patrick Crutchley and Alex Fine, "Natural Language Processing of Social Media as Screening for Suicide Risk," Qntfy, February 28, 2018, Boston, MA, USA.

[14] K S Naveenkumar, R Vinayakumar, and K P Soman, "Amrita-CEN-SentiDB 1: Improved Twitter Dataset for Sentimental Analysis and Application of Deep Learning", 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT), 2019.DOI: 10.1109/ICCCNT45670.2019.8944758

[15] Chakravarthi, B.R., Priyadharshini, R., Muralidaran, V. et al. DravidianCodeMix: sentiment analysis and offensive language identification dataset for Dravidian languages in code-mixed text. Lang Resources & Evaluation 56, 765–806 (2022). https://doi.org/10.1007/s10579-022-09583-7

[16] Sachin Kumar, S., Anand Kumar, M., Soman, K.P., Poornachandran, P. (2020). Dynamic Mode-Based Feature with Random Mapping for Sentiment Analysis, Intelligent Systems, Technologies and Applications. Advances in Intelligent Systems and Computing, vol 910. Springer, Singapore.

[17] Dhanya, N.M., Harish, U.C. (2021). Detection of Rumors in Tweets Using Machine Learning Techniques. In: Komanapalli, V.L.N., Sivaku- maran, N., Hampannavar, S. (eds) Advances in Automation, Signal Processing, Instrumentation, and Control. i-CASIC 2020. Lecture Notes in Electrical Engineering, vol 700. Springer, Singapore.

[18] Suicide and Depression Detection,2008, Available from: https://www.kaggle.com/datasets/nikhileswarkomati/suicide-watch

[19] A Padmavathi and D. Sarker, "RecipeMate: A Food Media Recommendation System Based on Regional Raw Ingredients," 2023 14th International Conference on Computing Communication and Networking Technologies (ICCCNT), Delhi, India, 2023, pp. 1-6, doi: 10.1109/ICCCNT56998.2023.10307728

[20] A. Padmavathi, R. M. Suresh, A Fuzzy Swarm Optimized Approach for Piece Selection in Bit Torrent like Peer to Peer Network, World Academy of Science, Engineering and Technology, International Journal of Computer and Information Engineering, Vol:9, No:1, Feb 2015