Automated Suicide Detection in Text Data Using Machine Learning and NLP Techniques

Author 1:

R.priyadarshini

Department of Electronics &Communication

Engineering,SOET-Sri Padmavati Mahila Visvavidyalam Tirupati

Author 2:

P. Sravya

Department of Electronics &Communication

Engineering, SOET-Sri Padmavati Mahila Visvavidyalayam Tirupati

Author 3:

V. Gowri Priya

Department of Electronics &Communication

Engineering,SOET-Sri Padmavati Mahila Visvavidyalayam Tirupati

Author 4:

S. Bhanu

Department of Electronics &Communication

Engineering, SOET-Sri Padmavati Mahila Visvavidyalayam Tirupati

Author 5:

T. Navya Sree

Department of Electronics &Communication

Engineering, SOET-Sri Padmavati Mahila Visvavidyalayam Tirupati

Abstract — Suicide is a serious public health issue, and prompt intervention efforts can be aided by early identification of suicidal intent in online content. This study uses the Suicide_Detection.csv dataset to analyse and categorise posts about suicide using machine learning techniques. Tokenisation and TF-IDF vectorisation are two preprocessing techniques used to extract valuable features from text data. To find high-risk cases, a variety of classification methods are used, such as XGBoost, Random Forest, and Logistic Regression. Exploratory data analysis sheds light on important language trends connected to information about suicide. The study shows how AI-driven methods can be used to identify mental health hazards from textual data. Future developments will involve real-time deployment in social media monitoring systems for proactive suicide prevention and the integration of deep learning techniques such as BERT for enhanced contextual comprehension.

Keywords— preventing suicide, machine learning, Mental health, text classification, Logistic regression, sentiment analysis, suicide detection, natural language processing,TF-IDF vectorisation, Suicide risk assessment,social media analysis.

I. INTRODUCTION

Millions of people die by suicide every year as a result of mental health issues, social pressures, and personal problems, making it a major global public health concern. The challenging task of identifying people who are at risk of suicide necessitates the integration of socioeconomic, psychological, and linguistic factors. Self-reporting, clinical evaluations, and intervention programs are the mainstays of traditional suicide prevention strategies, which may not always be successful in detecting suicide in real time. As digital communication has grown, people frequently use online platforms to convey their feelings and mental states. Understanding suicidal intent can be gained by examining these literary statements. In order to create a system that can help mental health practitioners recognise and act before a crisis arises, this research investigates the potential of machine learning in spotting posts pertaining to suicide.

NLP and ML provide strong methods for examining unstructured text data, which makes them appropriate for spotting trends that point to suicidal ideation. To train predictive models, we use a dataset of social media posts classified as "suicide" and "non-suicide" in this work. Tokenisation, stop-word removal, and feature extraction using TF-IDF vectorisation are among the preparation procedures applied to the dataset. To classify the postings according to linguistic features, three classification models—Logistic Regression, Random Forest, and XGBoost—are used. Trends, word distributions, and sentiment patterns within the collection are visualised through exploratory data analysis (EDA). This project attempts to close the gap between AI-driven analytics and mental health intervention by automating text-based suicide risk assessment, offering a scalable and effective method of suicide detection.

Suicide prevention initiatives will be significantly impacted by the adoption of such a system. Proactive action by mental health providers, crisis hotlines, and legislators can be aided by a prediction model that can identify high-risk individuals. AI-driven analysis, in contrast to conventional techniques, can process enormous volumes of data in real-time, which makes it very helpful for early warning systems and social media monitoring. To guarantee the system's dependability and security, however, ethical issues such as data privacy, false positives, and responsible AI use must be addressed. Deep learning models like BERT can be investigated in future studies to enhance contextual comprehension and improve prediction accuracy. In the end, this study shows how artificial intelligence may improve mental health monitoring and support early intervention techniques to lower suicide rates.

II. RELATED WORK

Because mental health problems are becoming more common in digital communication, there has been a lot of study done on the detection of suicidal thoughts using machine learning and natural language processing. To find patterns that point to suicide intentions, researchers have analysed social media posts using a variety of deep learning and natural language processing approaches. To illustrate its efficacy in examining user posts, Renjith et al. suggested an ensemble deep learning method for identifying suicidal intent on social media sites [1]. The promise of AI-based systems in mental health prediction was also demonstrated by Bhat and Goldman-Mellor, who used neural networks to forecast teenage suicide attempts based on internet activity [2].Numerous research have further confirmed the efficacy of NLP models like transformers and recurrent neural networks (RNNs) [4][5].

The use of social media analytics for early suicide risk detection has been the subject of a large body of study. Burnap et al. shown that language cues and user behaviours can be powerful predictors of suicide risk by using machine learning to separate Twitter data into suicide-related and non-suicide-related content [6]. Similarly, Ji et al. showed how different text categorisation models may accurately detect suicide intent by applying supervised learning approaches to online user-generated content [7]. Additionally, Carson et al. have investigated the integration of AI-driven suicide detection with electronic health records. They have improved the early identification of suicidal behaviour by analysing mental hospitalisation data using machine learning and natural language processing (NLP) [8].

Suicide detection techniques have been considerably improved by recent developments in deep learning and context-aware models. In order to improve the interpretability of AI models, Gaur et al. created a knowledge-aware system that evaluates the severity of suicide risk based on contextual understanding [11]. Additionally, as shown by Vioulès et al. [12], transformer-based models like BERT and GPT have been used for sentiment classification in talks pertaining to suicide. The significance of temporal aspects was emphasised by De Choudhury et al., who proposed that examining trends over time can improve the precision of suicidal ideation detection [13]. All of these research highlight how combining multi-modal data sources, sophisticated NLP algorithms, and psychological insights can greatly enhance the identification and forecasting of suicidal thoughts in digital settings [15].

INFERENCE FROM LITERATURE SURVEY

The literature review emphasises the increasing significance of AI-powered methods for text analysis-based suicidal ideation detection. Numerous research have shown that suicidal and non-suicidal content can be reliably classified using machine learning and deep learning models, such as neural networks, transformers, and ensemble approaches. Finding risk factors for suicide has been made possible by the use of behavioural feature extraction, sentiment analysis, and social media analytics. Additionally, by combining contextual and historical data, AI integration with electronic health records improves prediction accuracy. The detection technique is further improved by recent developments in deep learning, such as transformer-based models and hybrid approaches. According to the combined results, linguistic, behavioural, and contextual factors combined with advanced AI can greatly enhance early suicide risk detection, which may help mental health providers provide prompt assistance and intervention.

III. EXISTING SYSTEM

The present suicide detection techniques mainly rely on manual intervention, in which crisis hotlines and mental health specialists use surveys and interviews to examine patient behaviour. To evaluate a person's risk level, traditional methods use psychological testing and self-report questionnaires. These approaches do have drawbacks, though, such as self-report bias, which occurs when people underreport their discomfort because they are ashamed or afraid of the consequences. Furthermore, because these tests are often performed infrequently, it is challenging to identify changes in a person's mental state in real time. Many high-risk individuals might thus go overlooked, which could cause actions to be delayed. Although social media platforms and mental health organisations have put in place flagging systems based on user reports, these are reactive rather than proactive and do not recognise those who do not specifically seek assistance.

Another current strategy makes use of keyword-based filtering algorithms that are put in place by social media and online platforms. To identify suicide intent, these algorithms keep an eye on user postings, mails, and comments. They mostly use pre-made word lists and basic pattern matching, though, which can result in both false positives and false negatives. For example, a post that uses terms like "depressed" or "suicide" would not always mean that the user intends to kill themselves, while indirect signs of distress might be missed. Furthermore, these keyword-based models are less successful in real-world situations because they ignore linguistic variances, sarcasm, and contextual subtleties. The reliability of conventional filtering systems in identifying those at risk is greatly diminished by this lack of semantic knowledge.

Machine learning-based solutions have recently been developed, in which algorithms diagnose suicidal ideation by analysing textual input. For sentiment analysis and suicide risk prediction, supervised learning models have been used. These algorithms identify suicidal tendencies by extracting textual data including word frequency, emotional tone, and grammatical structure. Traditional machine learning models, on the other hand, frequently call for a great deal of feature engineering, which limits their ability to adjust to changing linguistic expressions. Additionally, the calibre and variety of training data have a significant impact on these models' effectiveness. The generalisability of these models across various populations and cultural contexts is limited by the absence of representation from diverse demographics in many of the datasets that are currently available.

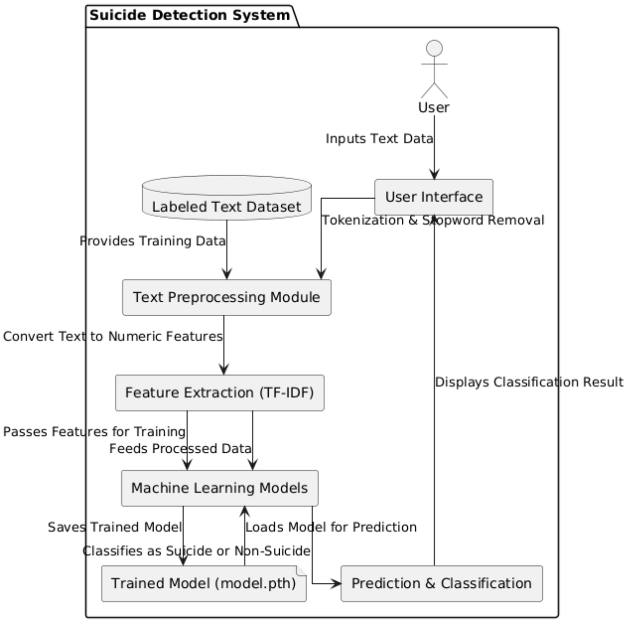
IV. IV. PROPOSED METHODOLOGY

This research suggests a machine learning-based suicide detection system that analyses text data from online forums and social media using Natural Language Processing (NLP) and classification algorithms. The first step in the process is data collection, which involves compiling a sizable dataset of texts pertaining to and unrelated to suicide. Raw text is transformed into numerical representations using text preparation methods as TF-IDF vectorisation, tokenisation, and stopword elimination. Words are weighed according to their importance in the dataset using TF-IDF in order to improve feature extraction. To ensure compliance with machine learning models, label encoding is used to convert categorical labels into numerical values. After that, the dataset is divided into training and testing sets in order to assess how well the models classify posts pertaining to suicide.

Three models are used for categorisation in order to determine whether a certain text conveys suicide intent. While Random Forest uses many decision trees to improve categorisation, Logistic Regression offers a baseline. XGBoost is a gradient boosting algorithm that uses optimised feature selection and boosting techniques to improve predictive performance. Performance indicators like F1-score are used to assess the models after they have been trained on the processed dataset. Confusion matrices are also produced in order to illustrate categorisation errors. Joblib serialisation is used to deploy the finished model for practical use. Sensitive information is anonymised, and findings are solely utilised for research and intervention in order to protect user privacy and ethical considerations.

V. ARCHITECTURE DIAGRAM

The Suicide Detection System's system architecture is made to process textual input and categorise it as either non-suicide-related or suicide-related.


Suicide Detection System
User
Inputs Text Data
User Interface
Labeled Text Dataset
Tokenization & Stopword Removal
Provides Training Data
Text Preprocessing Module
Convert Text to Numeric Features
Displays Classification Result
Feature Extraction (TF-IDF)
Passes Features for Training
Feeds Processed Data
Machine Learning Models
Saves Trained Model
Loads Model for Prediction
Classifies as Suicide or Non-Suicide
Trained Model (model.pth)
Prediction & Classification

The User Interface (UI), Text Preprocessing Module, Feature Extraction Module (TF-IDF), Machine Learning Models, and Prediction & Classification Module are some of the main parts of the architecture. When a user enters text data via the user interface, the procedure begins. After that, this text is sent to the preprocessing module, where crucial operations like lemmatisation, tokenisation, and stopword removal are carried out in order to clean and organise the data. Before being delivered to the machine learning models, the processed text is first transformed into a numerical representation using TF-IDF,which aids in identifying significant patterns in the text.

After analysing the retrieved features, the machine learning models classify the text as either suicide-related or not. A labelled dataset comprising examples of both general non-suicidal text and suicide-related material is used to train these algorithms. After training, the model is loaded and saved for use in subsequent forecasts. After processing the model's output, the Prediction & Classification Module sends the final classification back to the user interface, where the user may see the outcome. The system's overall goal is to automatically identify suicidal words in text, which makes it helpful for mental health monitoring and treatment. For practical applications, this architecture guarantees scalable deployment, precise classification, and effective text processing.

VI. METHODOLOGY

1. Gathering and preprocessing datasets:

Gathering a sizable collection of texts pertaining to and unrelated to suicide is the first stage in the process. The Suicide Detection dataset, which includes labelled text samples, was used for this investigation. Data preparation is done to clean and normalise the text before the models are trained. This entails standardising the text by lowercasing it, tokenising it, lemmatising it, and eliminating stopwords. To further enhance the dataset's quality, redundant entries and unnecessary information are eliminated. This stage guarantees that the model is trained on meaningful, high-quality text free of errors and noise. Model training and feature extraction can then begin using the preprocessed dataset.

2. Feature Extraction using TF-IDF:

TF-IDF is used to extract features following preprocessing. By assessing the significance of terms within a given document and lowering the weight of frequently used words, TF-IDF transforms text data into numerical values. This technique aids in gleaning contextual meaning and pertinent language patterns from the text. Every input text is converted into a high-dimensional vector that accurately depicts its content thanks to the feature extraction procedure. Since machine learning models require numerical input rather than raw text, this transformation is essential for training them. After that, the TF-IDF feature set is fed into several categorisation models.

3. Model Selection and Training:

Three machine learning models are used in this research for categorisation. These models were selected due to their effectiveness in handling text classification. To train and assess model performance, the dataset is divided into training and testing sets (80%-20%). The TF-IDF feature set is used to train each model, which teaches it to differentiate between texts that are about suicide and those that are not. To maximise each model's performance, hyperparameter adjustment is done during training. The efficacy of the trained models in practical applications is subsequently confirmed by evaluating them using important metrics including precision, recall, and F1-score.

4.Model Comparison and Evaluation:

Following training, the models' performance is assessed using the test dataset. Standard classification criteria, such as accuracy, precision, recall, and F1-score, are used in the evaluation. Plotting confusion matrices is another way to examine the proportion of accurate and inaccurate predictions. To determine which model is best suited for this classification assignment, the advantages and disadvantages of each model are analysed. Random Forest and XGBoost show gains in categorisation, while Logistic Regression provide a baseline performance. This comparison aids in identifying the model for suicide text detection that strikes the optimal balance between interpretability, efficiency, and predictive accuracy.

5.Deployment and Model Storage:

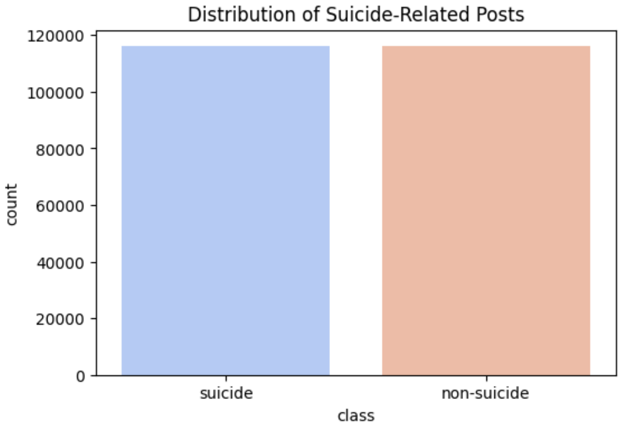
The top-performing model is chosen, and it is then saved using Joblib so that it may be used again for predictions without requiring retraining. After that, the model is included into a web application built with Flask, allowing users to enter text and get real-time categorisation results. To guarantee accurate and seamless predictions, the system architecture consists of the trained model, TF-IDF feature extraction, and a preprocessing pipeline. The deployment method makes it possible to apply the model in real-world settings, like crisis intervention services and mental health platforms, where prompt aid might result from early identification of suicide-related text.

6.Streamlit Deployment:

The application is deployed using Streamlit to make the suicide text detection model accessible and easy to use. An open-source Python framework called Streamlit makes it possible to create interactive, fast web apps with no coding. In the deployment phase, a Streamlit-based user interface is created where users can enter text, and the trained model makes predictions in real time about whether or not the text is relevant to suicide. Before sending the input text to the classifier, the program preprocesses it and extracts TF-IDF features. It then loads the pretrained model, which was saved using Joblib. The results are shown immediately, and the user interface is easy to use. Streamlit Cloud hosts the application.

VII. RESULTS AND DISCUSSION

With balanced precision and recall values for both classes, the Logistic Regression model had the highest overall accuracy of 93%. While its somewhat lower memory for class 1 (92%) implies that some suicidal texts are misclassified, its strong recall for class 0 (94%) shows that it properly classifies the majority of non-suicidal texts. Although this model does well overall, it might need to be improved in order to detect some edge cases.With an accuracy of 90%, the Random Forest model performed marginally worse. It has difficulty balancing sensitivity to suicidal messages (recall = 92%) and non-suicidal texts (recall = 89%), despite having precision and recall values that are near to one another. It's possible that this model is overfitting, which could cause slight prediction irregularities.



With an accuracy of 91%, the XGBoost model outperformed the Random Forest model but fell short of Logistic Regression. It is a well-rounded choice because it keeps recall and precision in check for both classes. The likelihood of missing non-suicidal texts is decreased by its greater recall for class 0 (93%) which guarantees fewer false negatives. It might, however, miss some suicide messages due to its weaker recall for class 1 (89%). This implies that additional adjustments, like hyperparameter optimisation, can improve its overall functionality.

Metric	Non-Suicidal	Suicidal	Macro Avg	Weighted Avg
Precision	.93	.94	.93	.93
Recall	.94	.92	.93	.93

F1-Score	.93	.93	.93	.93
Accuracy	-	-	.93	.93

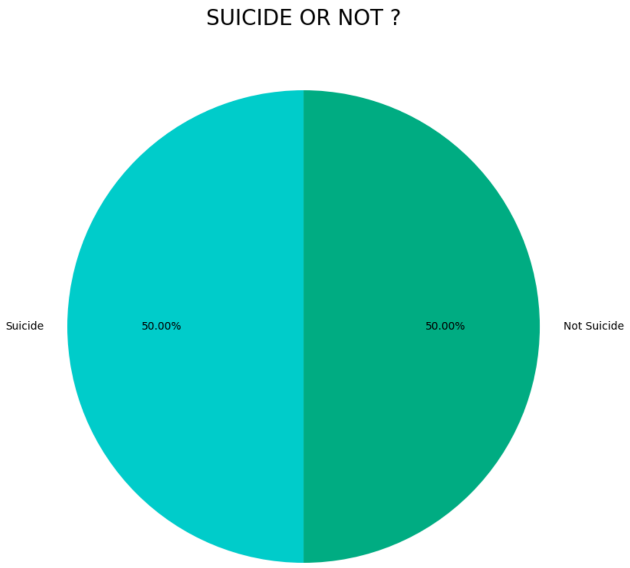
Table 1: Logistic Regression Model Performance

Metric	Non-Suicidal	Suicidal	Macro Avg	Weighted Avg
Precision	.91	.89	.90	.90
Recall	.89	.92	.90	.90
F1-Score	.90	.90	.90	.90
Accuracy	-	-	.90	.90

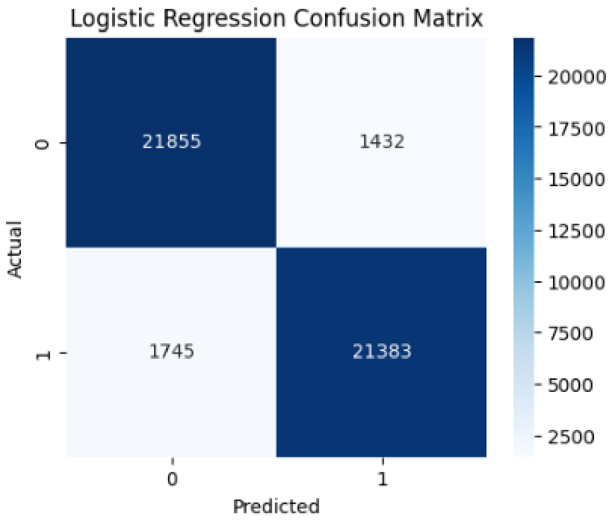
Table 2: Random Forest Model Performance

Metric	Non-Suicidal	Suicidal	Macro Avg	Weighted Avg
Precision	0.89	0.93	0.91	0.91
Recall	0.93	0.89	0.91	0.91
F1-Score	0.91	0.91	0.91	0.91
Accuracy	-	-	0.91	0.91

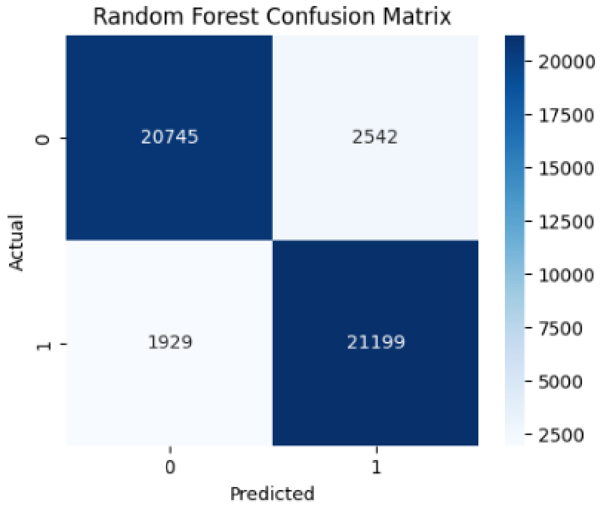
Table 3: XGBoost Model Performance



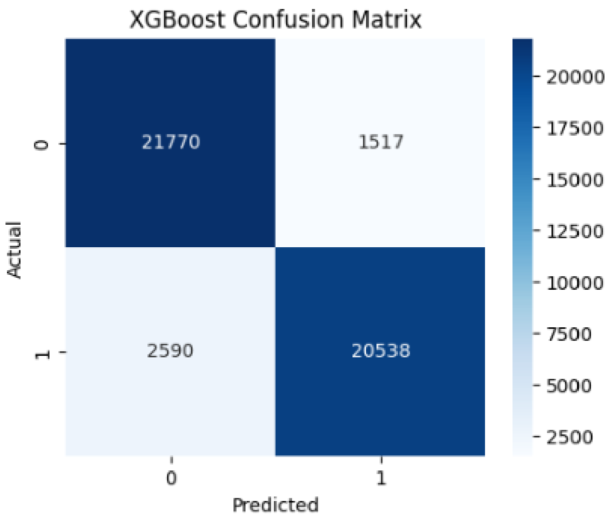
know
want
think
now
life
end
feel
time
people
help
much
anything
day
live
hate
someone
wish
anyone
kill
way
make
one
tried
love
talk
everything
lot
even
go
die
really
need
feeling
anymore
better
nothing
year
thought
friend
thing
going
family
say
try
happy
told
will
take
shit
don



Using XGBoost, Random Forest, and Logistic Regression, the research effectively uses machine learning models to distinguish between texts that are suicidal and those that are not. While each model performs differently in terms of recall and precision, they all show strengths in textual pattern recognition. The findings suggest that various models could have to choose between detecting positive and negative cases, which is important for practical uses. Since failing to detect suicidal intent in text categorisation might have dire consequences, the significance of recollection is emphasised in particular.



The system is accessible to mental health experts and support systems due to its Streamlit deployment, which guarantees user-friendly engagement and real-time analysis. To further increase its efficacy in suicide prevention initiatives, future developments might use sophisticated natural language processing methods, such transformers, to improve contextual knowledge and lower misclassification.



VIII. CONCLUSION

Using models like XGBoost, Random Forest, and Logistic Regression, the project effectively deploys a machine learning-based system for identifying suicidal intent in text. The technology helps mental health practitioners with early intervention by classifying communications as either suicidal or non-suicidal based on linguistic patterns. Tokenisation and feature extraction are two preprocessing methods that improve the models' comprehension of linguistic subtleties and increase prediction reliability. Accessibility is guaranteed by using Streamlit for system deployment, which enables real-time user interaction with the model. This tool, which provides an automated method of detecting those at risk, exemplifies the promise of artificial intelligence in mental health support.

Notwithstanding its efficacy, the project has drawbacks, such as difficulties interpreting confusing language and identifying nuanced emotions. Biases may be introduced by the use of labelled datasets, hence it is necessary to update the model frequently using a variety of real-world data. In order to decrease misclassification and enhance contextual awareness, future research could incorporate deep learning architectures like transformers. Furthermore, adding real-time monitoring capabilities and language support would improve the system's usefulness for a range of demographics. This project can make a substantial contribution to efforts to prevent suicide by improving the model and enhancing its capabilities, offering a scalable, AI-powered solution for the detection of mental health crises.

REFERENCES

[1] Renjith, S., Abraham, A., Jyothi, S. B., Chandran, L., & Thomson, J., "An ensemble deep learning technique for detecting suicidal ideation from posts in social media platforms," arXiv preprint arXiv:2112.10609, 2021.

[2] Bhat, H. S., & Goldman-Mellor, S. J., "Predicting Adolescent Suicide Attempts with Neural Networks," arXiv preprint arXiv:1711.10057, 2017.

[3] Gupta, S., Das, D., Chatterjee, M., & Naskar, S., "Machine Learning-Based Social Media Analysis for Suicide Risk Assessment," Emerging Technologies in Data Mining and Information Security, pp. 385–393, 2021. (link.springer.com)

[4] Abdulsalam, A., & Alhothali, A., "Suicidal Ideation Detection on Social Media: A Review of Machine Learning Methods," arXiv preprint arXiv:2201.10515, 2022.

[5] Tadesse, M. M., Lin, H., Xu, B., & Yang, L., "Detection of suicide ideation in social media forums using deep learning," Algorithms, vol. 13, no. 1, p. 7, 2020.

[6] Burnap, P., Colombo, W., & Scourfield, J., "Machine classification and analysis of suicide-related communication on Twitter," Proceedings of the 26th ACM Conference on Hypertext & Social Media, pp. 75–84, 2015.

[7] Ji, S., Yu, C. P., Fung, S. F., Pan, S., & Long, G., "Supervised learning for suicidal ideation detection in online user content," Complexity, vol. 2018, Article ID 6157249, 2018.

[8] Carson, N. J., Mullin, B., Sanchez, M. J., Lu, F., Yang, K., Menezes, M., & Le Cook, B., "Identification of suicidal behaviour among psychiatrically hospitalized adolescents using natural language processing and machine learning of electronic health records," PloS ONE, vol. 14, no. 2, e0211116, 2019.

[9] AlSagri, H. S., & Ykhlef, M., "Machine learning-based approach for depression detection in Twitter using content and activity features," arXiv preprint arXiv:2003.04763, 2020.

[10] Vioulès, M. J., Moulahi, B., Azé, J., & Bringay, S., "Detection of suicide-related posts in Twitter data streams," IBM Journal of Research and Development, vol. 62, no. 1, pp. 7:1–7:12, 2018.

[11] Gaur, M., Alambo, A., Sain, J. P., Kursuncu, U., Thirunarayan, K., Kavuluru, R., Sheth, A., & Welton, R., "Knowledge-aware assessment of severity of suicide risk for early intervention," Proceedings of the 27th ACM International Conference on Information and Knowledge Management, pp. 943–951, 2018.

[12] Coppersmith, G., Leary, R., Crutchley, P., & Fine, A., "Natural language processing of social media as screening for suicide risk," Biomedical Informatics Insights, vol. 10, pp. 1–11, 2018.

[13] De Choudhury, M., Gamon, M., Counts, S., & Horvitz, E., "Predicting depression via social media," Proceedings of the Seventh International AAAI Conference on Weblogs and Social Media, pp. 128–137, 2013.

[14] Shing, H. C., Nair, S., Zirikly, A., Friedenberg, M., Daumé III, H., & Resnik, P., "Expert, crowdsourced, and machine assessment of suicide risk via online postings," Proceedings of the Fifth Workshop on Computational Linguistics and Clinical Psychology: From Keyboard to Clinic, pp. 25–36, 2018.

[15] Sawhney, R., Manchanda, P., Mathur, P., Shah, R. R., & Singh, R., "Exploring and learning suicidal ideation connotations on social media with deep learning,"

Proceedings of the 9th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis, pp. 167–175, 2018.

[16] Cao, B., Shen, J., & Feng, Y., "Suicide risk assessment on social media: A multi-level dual-context language model," Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, pp. 190–200, 2020.

[17] Roy, A., Soni, S., & Singh, P. K., "Deep learning based approach for detection of suicidal ideation in Twitter data," Procedia Computer Science, vol. 167, pp. 2318–2327, 2020.

[18] Yazdavar, A. H., Mahdavinejad, M. S., Bajaj, G., Sheth, A., & Thirunarayan, K., "Analyzing social media data to predict the suicidal ideation," Proceedings of the 2018 IEEE/WIC/ACM International Conference on Web Intelligence, pp. 524–527, 2018.

[19] Desmet, B., & Hoste, V., "Online suicide prevention through optimised text classification," Information Sciences, vol. 439–440, pp. 61–78, 2018.

[20] Gkotsis, G., Oellrich, A., Velupillai, S., Liakata, M., Hubbard, T. J. P., Dobson, R. J. B., & Dutta, R., "Characterisation of mental health conditions in social media using Informed Deep Learning," Scientific Reports, vol. 7, Article 45141, 2017.