Ethically Driven Synthetic Data for Suicidal Ideation Detection sing NLP

[1]D Avinash , [2]D Lavanya , [3]H Sushumna, [4]K Shankar

[1,2,3] UG Scholars, [4]Assistant Professor
[1,2,3,4] Department of Computer Science and Engineering,
[1,2,3,4] Guru Nanak Institutions Technical Campus, Hyderabad, Telangana, India.

ABSTRACT: Detecting suicidal ideation through social media content is a critical initiative to support mental health intervention strategies. This study presents an explainable framework that leverages advanced natural language processing (NLP) techniques to address the challenges of identifying suicidal intent in user- generated content. A significant innovation in this work is the creation of synthetic datasets informed by psychological and social factors associated with suicidal ideation, designed to supplement limited real-world data while maintaining ethical considerations. The proposed system classifies social media content into two categories: non-suicidal or suicidal. The hybrid approach of combining synthetic and real-world data enhances model performance, achieving superior accuracy and robustness compared to traditional methods. The framework emphasizes explainability by incorporating techniques that identify key linguistic and contextual features driving model predictions, ensuring interpretability for mental health professionals and researchers. This approach underscores the potential of integrating synthetic data and NLP in addressing real-world challenges such as data scarcity, diversity, and ethical concerns. By providing actionable insights and ensuring transparency, the proposed framework contributes to building reliable and scalable solutions for suicide prevention in digital environments.

1 INTRODUCTION:

The rise of social media as a platform for self- expression has made it a valuable source for understanding user emotions and mental health conditions. Suicidal ideation detection using Natural Language Processing (NLP) has emerged as a critical solution to support early intervention and suicide prevention strategies. However, identifying suicidal intent in user-generated content presents several challenges, including data scarcity, ethical concerns, and contextual complexity. Real-world suicidal ideation data is often limited, sensitive, and difficult to annotate due to privacy and ethical considerations. To address these limitations, this project introduces a Socially Aware Synthetic Data Generation approach that incorporates psychological and social factors associated with suicidal behavior to create supplementary datasets. By combining synthetic and real-world data, this study aims to enhance model performance and overcome the shortcomings of traditional methods. The proposed system leverages explainable NLP techniques to ensure model transparency, enabling mental health professionals to understand and trust the decisions made by the detection framework.

1.1 OBJECTIVE:

To design and develop a synthetic data generation approach that supplements real- world suicidal ideation datasets while maintaining ethical considerations. To leverage advanced NLP techniques for classifying user-generated content into Suicidal and Non- Suicidal categories. To enhance model robustness and generalization by integrating synthetic data with real-world data. To incorporate explainability in the proposed framework, enabling the identification of keylinguistic and contextual features driving model predictions. To address challenges of data scarcity, diversity, and ethical concerns through the integration of synthetic data. To provide actionable and interpretable insights for mental health professionals to aid suicide prevention efforts.

1.2 SCOPE:

The scope of this project lies in the development of a robust, explainable, and scalable framework for detecting suicidal ideation from social media content. It involves creating synthetic datasets that are socially and psychologically informed to address the challenges of data scarcity and ethical concerns. The project focuses on classifying content into Suicidal and Non-Suicidal categories while emphasizing the explainability of model

predictions. This framework aims to provide actionable insights to mental health professionals and researchers, supporting suicide prevention strategies in digital environments

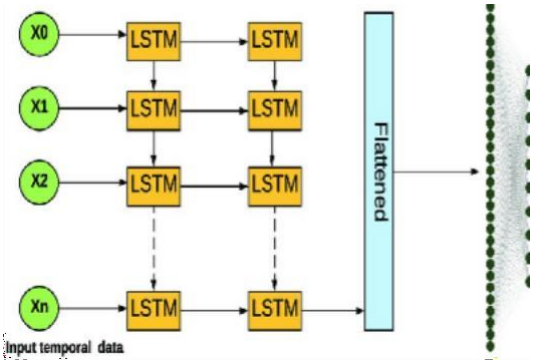
1.3 EXISTING SYSTEM:

ALBERT (A Lite BERT) is a variant of the popular BERT (Bidirectional Encoder Representations from Transformers) model, developed by researchers at Google AI. It was introduced to address some of the key challenges of BERT, such as its large model size, high computational cost, and inefficient memory usage. ALBERT retains the strengths of BERT in understanding natural language but makes it lighter and faster by introducing parameter-reduction techniques. The key focus of ALBERT is to make the model more efficient while maintaining or improving performance on tasks like question answering, sentiment analysis, and natural language inference.

1.3.1 Existing System Disadvantages:

ALBERT still requires significant time to pre-train on large datasets, although it is lighter than BERT. Parameter sharing and factorized embeddings add complexity to the architecture, which may require more fine-tuning expertise. Reduced Flexibility.

1.5 SYSTEM ARCHITECTURE:



EXPLANATION:

Deployment Diagram is a type of diagram that specifies the physical hardware on which the software system will execute. It also determines how the software is deployed on the underlying hardware. It maps software pieces of a system to the device that are going to execute it.

1.6 PROPOSED SYSTEM

Long Short-Term Memory (LSTM) is a special type of Recurrent Neural Network (RNN) architecture designed to solve problems associated with standard RNNs, such as vanishing and exploding gradients. It was introduced by Sepp Hochreiter and Jürgen Schmidhuber in 1997. LSTMs are particularly effective at capturing long-term dependencies in sequential data, making them widely used in tasks involving time series, natural language processing (NLP), speech recognition, and other sequence- based applications.

Unlike traditional RNNs, LSTMs use gates (input, forget, and output gates) to regulate the flow of information, allowing them to remember relevant information for long periods while forgetting irrelevant details.

1.6.1 PROPOSED SYSTEM ADVANTAGES:

LSTMs are specifically designed to remember information for long periods, solving the vanishing gradient problem seen in traditional RNNs. LSTMs work well for time series, NLP tasks, speech recognition, video processing, and other applications involving sequential data. Avoids Gradient Vanishing/Exploding.

2 DESCRIPTION:

2.1 GENERAL:

This project presents a hybrid framework for detecting suicidal ideation in social media content by combining synthetic and real-world datasets. The generation of synthetic data is informed by key psychological and social factors associated with suicidal behavior, ensuring the datasets are contextually meaningful and ethically appropriate. Advanced Natural Language Processing (NLP) techniques are applied to classify textual data into two primary categories: Suicidal or Non-Suicidal. The hybrid approach addresses the limitations of real-world data, such as scarcity and sensitivity, by enhancing dataset diversity and improving model generalization. A notable feature of the system is its emphasis on explainability, achieved through techniques that highlight key linguistic and contextual features influencing model predictions. This ensures that the decisions made by the model are interpretable and can be trusted by mental health professionals. By providing actionable insights, the proposed framework not only aids in early detection of suicidal ideation but also contributes to developing scalable and reliable solutions for real- world deployment in suicide prevention programs. This study demonstrates the potential of integrating synthetic data generation and NLP to overcome critical challenges, such as limited data availability, ethical concerns, and the need for transparency in AI-based systems. The outcomes of this project pave the way for future research on leveraging AI to address global mental health issues effectively.

2.2 METHODOLOGIES

2.2.1 MODULES NAME:

- Data Collection
- Data Analysis
- Use the NLTK
- Data Preprocessing
- Splitting the data
- Train the model
- Model Evaluation
- Result

Data Collection:

Data collection is the first and most critical step in the project. In this module, relevant data is gathered from social media platforms or publicly available suicide-related datasets. Given the sensitivity of suicidal ideation data, a combination of real-world data and synthetic data is used to ensure ethical considerations are met. Tools such as web scraping techniques (e.g., using Beautiful Soup or APIs) and annotated datasets from research repositories are utilized. Synthetic data is generated using text generation models to supplement the limited real data. The collected data includes user posts, comments, or tweets.

Data Analysis:

Once the data is collected, the next step involves analysing its structure and understanding its nature. Exploratory Data Analysis (EDA) is performed to identify patterns, trends, and anomalies in the dataset. Key tasks include:

- ➢ Identifying the distribution of suicidal vs. non-suicidal data.
- ➢ Analysing word frequency, word clouds, and n-grams to understand common terms or phrases.
- ➢ Analysing linguistic, psychological, and social cues within the textual data.
- ➢ Tools like Pandas, Matplotlib, and Seaborn

are often used for visualizing and interpreting the analysis results.

Data Preprocessing:

Data preprocessing refers to the process of preparing and enriching data for modelling. This includes steps like resizing images to a consistent size, normalizing pixel values, removing noise or unnecessary elements, and converting the data into a format suitable for use with machine learning algorithms.

Use the NLTK:

The Natural Language Toolkit (NLTK) is used for various text processing and analysis tasks in this project. NLTK is a powerful Python library specifically designed for Natural Language Processing (NLP) tasks. Some key functions include:

- Tokenization: Breaking text into words or sentences.
- Stopword Removal: Eliminating common words like “and,” “the,” “is” that do not contribute to meaning.
- Lemmatization and Stemming: Reducing words to their base form (e.g., "running" → "run").
- POS Tagging: Assigning Parts of Speech (POS) to words.
- Using NLTK ensures the raw textual data is transformed into a clean and meaningful format, which is essential for training the model.

Data Preprocessing:

Data preprocessing involves transforming raw text data into a suitable format for machine learning models. The key steps include:

- Text Cleaning: Removing punctuation, special characters, numbers, and URLs from the text.
- Lowercasing: Converting all text to lowercase to avoid case sensitivity issues.
- Stopword Removal: Eliminating irrelevant words using NLTK stopword libraries.
- Tokenization: Splitting sentences into individual words or tokens.
- Lemmatization: Converting words to their root form for consistency.
- After preprocessing, the text is vectorized using techniques like TF-IDF or Word Embeddings (e.g., Word2Vec, GloVe) to convert the textual data into numerical features.

Splitting the Data:

Once the data is preprocessed, it is split into training and testing datasets to evaluate model performance effectively. The standard approach is an 80-20 split:

- Training Set (80%): Used to train the machine learning model.
- Testing Set (20%): Used to evaluate the performance of the trained model.

The train-test split ensures the model is trained on a majority of the data while having unseen data to test its generalization capability. Libraries like scikit- learn are used for data splitting.

Train the Model:

In this module, the processed and split data is used to train the machine learning or deep learning model. Since the project involves suicidal ideation detection, NLP model as LSTM, utilized. The steps include:

- Feeding the training data into the model.
- Fine-tuning model hyperparameters (e.g., learning rate, batch size, number of epochs).
- Using synthetic data alongside real data to improve model robustness and generalization.
- The hybrid approach ensures the model learns meaningful patterns from both real and synthetic data

Model Evaluation:

After training the model, its performance is evaluated using appropriate metrics to ensure accuracy and reliability.

Result:

The results of the project showcase the model's performance in detecting suicidal ideation. Key insights include:

- ➢ Identification of key linguistic and contextual features that influence predictions (explainability).
- ➢ Comparisons with traditional approaches to highlight improvements in generalization and performance.

The final results provide actionable insights for mental health professionals, demonstrating the system's ability to detect suicidal ideation effectively while ensuring transparency and reliability.

2.3 TECHNIQUE USED OR ALGORITHM USED

2.3.1 EXISTING TECHNIQUE: ALBERT

➢ ALBERT is a transformer-based language representation model that optimizes BERT's architecture using the following techniques Factorized Embedding Parameterization: Reduces the size of the embedding matrix by factorizing it into two smaller matrices.

➢ Cross-Layer Parameter Sharing: Shares parameters across transformer layers to reduce the overall number of parameters. Sentence-Order Prediction (SOP): Replaces the next-sentence prediction task in BERT with SOP to better capture inter-sentence relationships. These optimizations make ALBERT lighter, require fewer parameters, and make it more computationally efficient without sacrificing accuracy.

2.3.2 PROPOSED TECHNIQUE USED OR ALGORITHM USED:

LSTM (Long Short-Term Memory):

➢ LSTM (Long Short-Term Memory) is a type of recurrent neural network that uses special memory cells and gating

mechanisms to store, update, and regulate information over time. It is designed to overcome the limitations of traditional RNNs by enabling the model to learn and remember long- term dependencies in sequential data. Key components of an LSTM unit include: Forget Gate: Decides what information to discard.

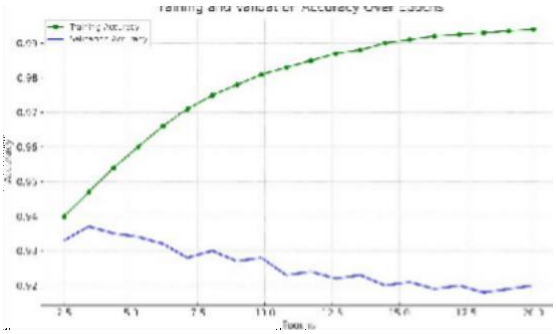
➢ Input Gate: Determines what new information to store in the cell state. Output Gate: Controls the output based on the cell state. These mechanisms make LSTMs capable of handling longer sequences without suffering from the vanishing gradient problem.

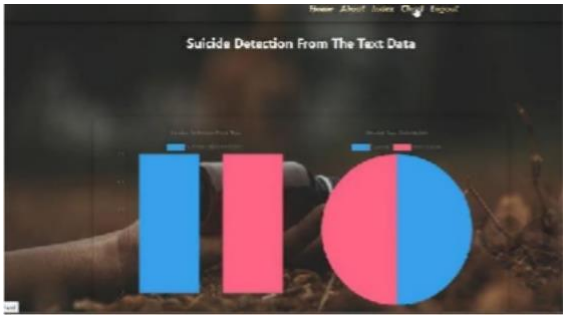
3 RESULT:


Ethically Driven Synthetic Data
for Suicidal Ideation
Detection Using NLP


Login Here
admin


Give The Data For Suicidal Ideation Detection




Suicide Detection From The Text Data

4 FUTURE ENHANCEMENT:

Additionally, deploying lightweight models on edge devices will enable real-time detection while preserving user privacy. A framework for continuous learning will be established to adapt dynamically to evolving online communication patterns, trends, and adversarial inputs. Further, incorporating explainable AI (XAI) techniques will enhance model transparency and interpretability, fostering trust in the system. Cross-lingual and cross-cultural analysis will ensure inclusivity and scalability, while federated learning methods will address privacy concerns by enabling decentralized training. Finally, addressing robustness against adversarial attacks will be prioritized to ensure the reliability of detection systems in real-world scenarios.

5 CONCLUSION:

The accurate identification of suicidal ideation from textual data remains crucial for early intervention and prevention efforts, with Natural Language Processing (NLP) techniques offering promising solutions. However, challenges such as the scarcity and sensitivity of real suicide-related data necessitate innovative approaches like synthetic data generation and data augmentation.

6 REFERENCES:

1] D. De Berardis, G. Martinotti, and M. Di Giannantonio, ‘‘Understanding the complex phenomenon of suicide: From research to clinical practice,’’ Frontiers Psychiatry, vol. 9, p. 61, 2018.

[1] E. R. Kumar and N. Venkatram, ‘‘Predicting and analyzing suicidal risk behavior using rule-based approach in Twitter data,’’ Soft Comput., pp. 1–9, 2023.

[2] A. Raza, F. Rustam, H. U. R. Siddiqui, I. D. L. T. Diez, B. Garcia-Zapirain, E. Lee, and I. Ashraf, ‘‘Predicting genetic disorder and types of disorder using chain classifier approach,’’ Genes, vol. 14, no. 1, p. 71, Dec. 2022.

[3] A. Abdulsalam and A. Alhothali, ‘‘Suicidal ideation detection on social media: A review of machine learning methods,’’ 2022, arXiv:2201.10515.

[4] Z. Li, J. Zhou, Z. An, W. Cheng, and B. Hu, ‘‘Deep hierarchical ensemble model for suicide detection on imbalanced social media data,’’ Entropy, vol. 24, no. 4, p. 442, Mar. 2022.

[5] D. Kodati and R. Tene, ‘‘Identifying suicidal emotions on social media through transformer-based deep learning,’’ Appl. Intell., vol. 53, no. 10, pp. 11885–11917, 2023.

[6] M. M. S. Fareed, A. Raza, N. Zhao, A. Tariq, F. Younas, G. Ahmed, S. Ullah, S. F. Jillani, I. Abbas, and M. Aslam, ‘‘Predicting divorce prospect using ensemble learning: Support vector machine, linear model, and neural

network,'' Comput. Intell. Neurosci., vol. 2022, pp. 1–15, Jul. 2022.

[7] Q. Wei, A. Franklin, T. Cohen, and H. Xu, ''Clinical text annotation— What factors are associated with the cost of time?'' in Proc. AMIA Annu. Symp. Bethesda, MD, USA: American Medical Informatics Association, Dec. 2018, pp. 1552–1560.

[8] R. Babbar and B. Schölkopf, ''Data scarcity, robustness and extreme multilabel classification,'' Mach. Learn., vol. 108, nos. 8–9, pp. 1329–1351,
Sep. 2019.

[9] S. I. Nikolenko, Synthetic Data for Deep Learning, vol. 174. Cham, Switzerland: Springer, 2021.