Project Phase II Report On

Suicide and Depression Risk Prediction

Submitted in partial fulfillment of the requirements for the award of the degree of

Bachelor of Technology

in

Computer Science and Engineering

By

Martin Francis Paul (U2003130)
Edwin Joseph K. J. (U2003075)
Joyce George Joseph (U2003113)
Francis J. Kalarickal (U2003081)

Under the guidance of

Dr. Uma Narayanan

Department of Computer Science and Engineering
Rajagiri School of Engineering & Technology (Autonomous)
(Parent University: APJ Abdul Kalam Technological University)
Rajagiri Valley, Kakkanad, Kochi, 682039
April 2024

CERTIFICATE

This is to certify that the project report entitled ***"Suicide and Depression Risk Prediction"*** *is a bonafide record of the work done by* ***Martin Francis Paul (U2003130), Edwin Joseph K. J. (U2003075), Joyce George Joseph (U2003113), Francis J. Kalarickal (U2003081)*** *submitted to the Rajagiri School of Engineering & Technology (RSET) (Autonomous) in partial fulfillment of the requirements for the award of the degree of Bachelor of Technology (B. Tech.) in Computer Science and Engineering during the academic year 2023-2024.*

Dr. Uma Narayanan
Project Guide
Asst Professor
Dept. of CSE
RSET

Dr. Tripti C.
Project Coordinator
Associate Professor
Dept. of CSE
RSET

Dr. Preetha K. G.
Head of the Department
Professor & HOD
Dept. of CSE
RSET

ACKNOWLEDGMENT

We wish to express our sincere gratitude towards **Dr. P. S. Sreejith**, Principal of RSET, and **Dr. Preetha K. G.** , Head of the Department of Computer Science and Engineering for providing us with the opportunity to undertake our project, ”Suicide and Depression Risk Prediction”.

We are highly indebted to our project coordinator, **Dr. Tripti C.** , Asst Professor, Department of Computer Science and Engineering, for her valuable support.

It is indeed our pleasure and a moment of satisfaction for us to express our sincere gratitude to our project guide **Dr. Uma Narayanan,** Asst Professor, Department of Computer Science and Engineering for her patience and all the priceless advice and wisdom she has shared with us.

Last but not the least, we would like to express our sincere gratitude towards all other teachers and friends for their continuous support and constructive ideas.

Martin Francis Paul (U2003130)
Edwin Joseph K. J. (U2003075)
Joyce George Joseph (U2003113)
Francis J. Kalarickal (U2003081)

Abstract

This project presents a groundbreaking application geared towards mitigating the rising rates of stress, depression, and suicides among college students. Utilizing state-of-the-art technology like machine learning (ML) and natural language processing (NLP), the app encourages students to maintain daily journals, which are subjected to a sophisticated NLP module for sentiment analysis and emotion recognition. The novel hybrid text representation technique that incorporates document- and word-level elements, enhances the accuracy of suicide risk predictions while maintaining interpretability. Privacy is paramount, with robust security measures ensuring the ethical handling of sensitive mental health data. The app's discreet notification system alerts counselors to potential risks, fostering timely interventions without compromising user anonymity. A dedicated educational resources section aims to reduce stigma and increase awareness about mental health, contributing to the creation of a supportive college environment. The project aligns with the evolving emphasis on transparency in AI applications, particularly in the sensitive domain of mental health, promising significant contributions to public health interventions and responsible technological advancements.

Contents

List of Abbreviations

SDM - Suicide Detection Model

C-SSRS - Columbia-Suicide Severity Rating Scale

GR - Graded Recall

GP - Graded Precision

SVM - Support Vector Machine

SHAP - SHapley Additive exPlanations

RF - Random Forest

List of Figures

List of Tables

Chapter 1

Introduction

In the era of widespread social media use, individuals often express their thoughts, emotions, and experiences online, providing a unique avenue for understanding mental health and identifying potential suicide risk factors. This project aims to tackle the critical issue of suicide danger identification on social media, recognizing the limitations of traditional clinical prediction methods and the growing need for advanced, interpretable techniques. By leveraging modern natural language processing (NLP)-based approaches and introducing a novel hybrid text representation method, the project seeks to offer a nuanced understanding of language used on social media platforms. This approach incorporates text representations at the word and document levels to capture both subtle nuances and broader contextual information, setting it apart from conventional models.

1.1 Background

The advent of social media platforms has not only facilitated unprecedented connectivity but has also opened new avenues for mental health research and intervention. Traditional clinical prediction methods, however, have struggled to keep pace with the evolving nature of online communication. Modern NLP-based approaches have shown promise, yet limitations persist, particularly in capturing the intricate interplay of low- and high-level features within social media content. Sequence learning methods, a prevalent approach in this domain, often fall short in delivering the required interpretability. The recognition of these challenges underscores the need for an innovative and interpretable solution.

This project recognizes the imperative to bridge the gap between the richness of social media data and the existing predictive methodologies. The hybrid text representation method introduced here aims to redefine the paradigm of suicide risk identification by combining the strengths of word and document-level representations. By doing so, the project seeks to enhance not only the accuracy of risk identification but also the interpretability of the models, a crucial factor in building trust and confidence among stakeholders. The broader societal and industrial relevance of this initiative lies in its potential to reshape mental health interventions, contributing to a more empathetic and supportive digital environment. The outcomes of this project, driven by cutting-edge technology and a deep understanding of mental health dynamics in the digital age, aspire to pave the way for more effective and transparent approaches to addressing mental health challenges.

1.2 Problem Definition

The rising incidence of stress, depression, and suicides among college students underscores the urgent need for comprehensive interventions. A multifaceted approach is essential, starting with increased funding and resources for mental health services on campuses, ensuring easy access to counseling and therapy for students. Additionally, there is a critical requirement to enhance support for mentors and educators in identifying and assisting students with mental health issues. Implementing training programs and providing necessary tools for educators can contribute to a more proactive and supportive environment. The overarching goal is to create a healthier and more empathetic college culture, ultimately reducing suicide rates among students.

1.3 Scope and Motivation

In the realm of mental health, the stakes are nothing short of a battle for life itself. Depression, with its silent descent, and the haunting specter of suicide cast shadows that extend far beyond individual lives; they reverberate through families, communities, and societies In August 2022, the National Crime Records Bureau (NCRB) made data on suicide fatalities in India public. The results were stunning with 1,64,033 suicides reported nationwide in 2021, the total number of suicides increased by 7.2% over the previous year. Moreover mental health is crucial at a time like this, when we take into account the incident at Amal Jyothi College where a student killed herself after receiving a scolding from college officials for using a cell phone in the lab. October 10 is recognized globally as World Mental Health Day each year, with the main goals being to increase public awareness and to galvanize efforts toward improved mental health.

1.4 Objectives

- To explore the critical issue of suicidal tendency in college students particularly focusing on its connection to written language.
- Identify early warning signs of suicidal tendencies in student communication.
- Categorize students into different risk levels.
- Utilize NLP for sentiment analysis, emotion recognition, and feature engineering.
- Train ML models for prediction.
- Ensure privacy, ethics, and user-friendliness.

1.5 Challenges

- Reluctance or idleness of students to write in the virtual journal
- Ensuring that the virtual journal is highly secure and that the entries are confidential.
- Students may be hesitant to record their thoughts and feelings if they fear unauthorized access.
- Data Scarcity

1.6 Assumptions

- Assumption that the data used for risk prediction is accurate, reliable and comprehensive.
- Assumption that there is a contribution of students towards the digital diary.

1.7 Societal / Industrial Relevance

The project's social impact is profound, directly addressing the escalating mental health crisis among college students. By providing an innovative app that employs advanced technologies like NLP and machine learning, the project seeks to reduce the prevalence of stress and depression while identifying and assisting students at risk of suicide. Beyond predictive capabilities, the app fosters a more supportive college environment, reducing the stigma associated with seeking help for mental health issues. Through its focus on early intervention and awareness, the project contributes to a cultural shift towards prioritizing mental well-being in educational settings, ultimately promoting a healthier and more empathetic society.

In the industrial context, the project holds significant relevance as it aligns with the growing demand for technologically-driven solutions in the mental health sector. Leveraging NLP, machine learning models, and ethical considerations, the app exemplifies the integration of cutting-edge technology into mental health support systems. This innovation can extend beyond the educational sector, attracting interest from industries involved in mental health services, healthcare technology, and corporate wellness programs. As companies increasingly recognize the importance of corporate social responsibility, the project showcases a socially impactful solution that has the potential to influence industry practices and contribute to the development of responsible and effective mental health technologies.

1.8 Organization of the Report

The rest of the chapters provides Literature Survey of suicide related papers and to contribute valuable insights and methodologies to the burgeoning field of mental health monitoring through social media analysis. The papers underscore the potential of leveraging technology for timely intervention and support in addressing mental health challenges through the analysis of social media. The Hardware and Software Requirements are mentioned followed by the system architecture where we look into a comprehensive overview that includes its architecture, design, module division, and a visual representation of the planned timeline through a Gantt Chart. In the Results and Discussion Chapter we discuss about the graphical analysis and performance metrics table further highlight ELECTRA's effectiveness in accurately distinguishing between suicide and non-suicide posts compared to other models. In the Final Chapter we conclude by highlighting that the study presents a successful method for suicide risk identification on social media and outlines promising avenues for future research in NLP and mental health.

Chapter 2

Literature Survey

2.1 Hybrid Text Representation for Explainable Suicide Risk Identification on Social Media

2.1.1 Introduction

The paper [1] presents a novel approach to risk of suicide identification on social media through the development of a hybrid text representation method. The suggested approach explains suicide risk identification by combining text representations at the word and document levels., which are then input into a transformer-based encoder with ordinal classification. The study demonstrates the effectiveness of the proposed method, outperforming state-of-the-art baselines with an F1Score of 0.79 on a public suicide dataset. The paper highlights the potential of explainable models in performing at a comparable level to the best non-explainable models while offering advantages when translated for use in clinical and public health practice. The research opens up several promising avenues for future research and application, including the exploration of advanced language models, extension to other mental health-related tasks, development of interpretable models, expansion to other social media platforms, ethical considerations, and longitudinal analysis.

2.1.2 Methodology

The paper discusses several methodologies used in the context of suicide risk identification from social media posts. These methodologies include deep learning, explainable models, social media analysis, suicide detection, and text analysis. Deep Learning: The use of deep learning techniques to automatically extract text properties from unprocessed text is explored in this work. It mentions the use of sequential and graph-based methods for various natural language processing (NLP) tasks. This includes the application

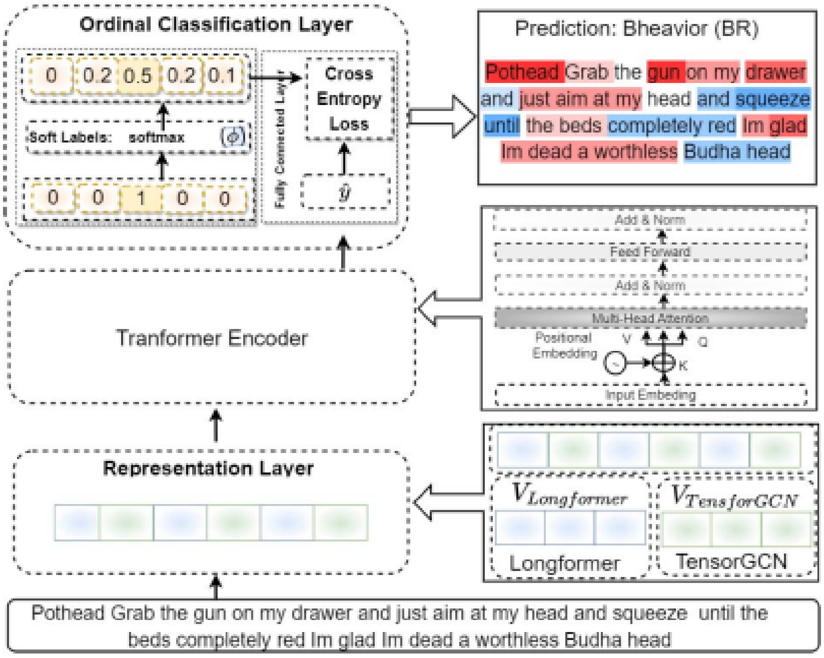


Figure 2.1: Architecture of the proposed method[1]

of transformer models and graph neural networks to capture both low- and high-level features from the text data. Explainable Models: The paper emphasizes the importance of explainable models in the context of suicide risk identification. Its goal is to assess a new approach that generates explainable models that can be used to clinical and public health settings. The proposed hybrid text representation method integrates word and document-level text representations to create explainable models for suicide risk identification. Social Media Analysis: The approaches covered in the study entail using the chronological order of users' social media posts to automatically analyze their mental states. This entails extracting both high-level and low-level attributes from the text data and taking into account the context that is assigned to users, including their social graph and mood spectrum. Suicide Detection:The difficulty of gathering both high- and low-level characteristics for social media suicide risk identification is discussed in the paper. It talks about creating text embeddings using transformer and graph-based models, then fusing them to capture both low- and high-level properties. The objective is to provide a system that can be explained for identifying suicide risk on social media. Text Analysis: The paper delves into the use of word embeddings to capture syntactic and semantic representations from unlabeled text. It also discusses the incorporation of multiple vector representations for each word type and word sense, as well as the use of graph neural networks to extract information about global word dependencies from the text graph formed by the entire corpus. These methodologies are aimed at improving the performance of suicide detection models and providing explainable models that can be translated for use in clinical and public health practice.

2.1.3 Conclusion

The study concludes with the introduction of a novel hybrid text representation technique for social media suicide risk identification. The suggested approach explains suicide risk identification by integrating word and document-level text representations, which are then input into a transformer-based encoder with ordinal classification. The study shows that the suggested strategy is effective, outperforming the state-of-the-art baselines with an F1Score of 0.79 on a dataset of public suicides. The research highlights the potential of explainable models in performing at a comparable level to the best non-explainable models while offering advantages when translated for use in clinical and public health practice. The study opens up several promising avenues for future research and application, including the exploration of advanced language models, extension to other mental health-related tasks, development of interpretable models, expansion to other social media platforms, ethical considerations, and longitudinal analysis. The successful development of an effective and explainable method for suicide risk identification on social media holds great promise for advancing our understanding of mental health signals in online user behavior and facilitating the development of impactful tools for mental health support and intervention.

2.2 Automatic identification of suicide notes with a transformer-based deep learning model

2.2.1 Introduction

The paper[2] introduces a novel deep learning model, TransformerRNN, designed for the automatic identification of suicide notes posted on social media platforms, addressing the escalating global issue of suicide. Leveraging a combination of transformer encoder and Bi-directional Long Short-Term Memory (BiLSTM) structures, the proposed model effectively captures contextual and long-term dependency information. TransformerRNN outperforms comparable machine learning and state-of-the-art deep learning models, demonstrating high precision, recall, and F1-score metrics. The application of this model holds promise for the development of suicide prevention technologies, enabling early detection and intervention in cases of suicidal risk on social media. By providing accurate identification of suicide notes, the model offers valuable insights for mental health professionals and support organizations, facilitating the monitoring and analysis of social media platforms for signs of suicidal ideation.

2.2.2 Methodology

The paper introduces a groundbreaking deep learning model named TransformerRNN, which specifically addresses the challenging task of automatically identifying suicide notes posted online. Unlike traditional approaches, TransformerRNN leverages a unique combination of a transformer encoder and Bi-directional Long Short-Term Memory (BiLSTM) structure. These components collaboratively enable the model to effectively capture contextual nuances and long-term dependencies within the input sequences. To enhance its understanding of the language, the model utilizes pretrained GloVe word representations for embedding the input sequences into word vectors. What sets TransformerRNN apart is the inclusion of a multi-head self-attention layer, allowing the model to simultaneously focus on different sub-spaces and positions in the data. This feature contributes to its superior performance over conventional machine learning models and state-of-the-art deep learning models, as demonstrated by its consistently high precision, recall, and F1-score metrics.

Beyond its technical achievements, TransformerRNN holds significant promise for real-

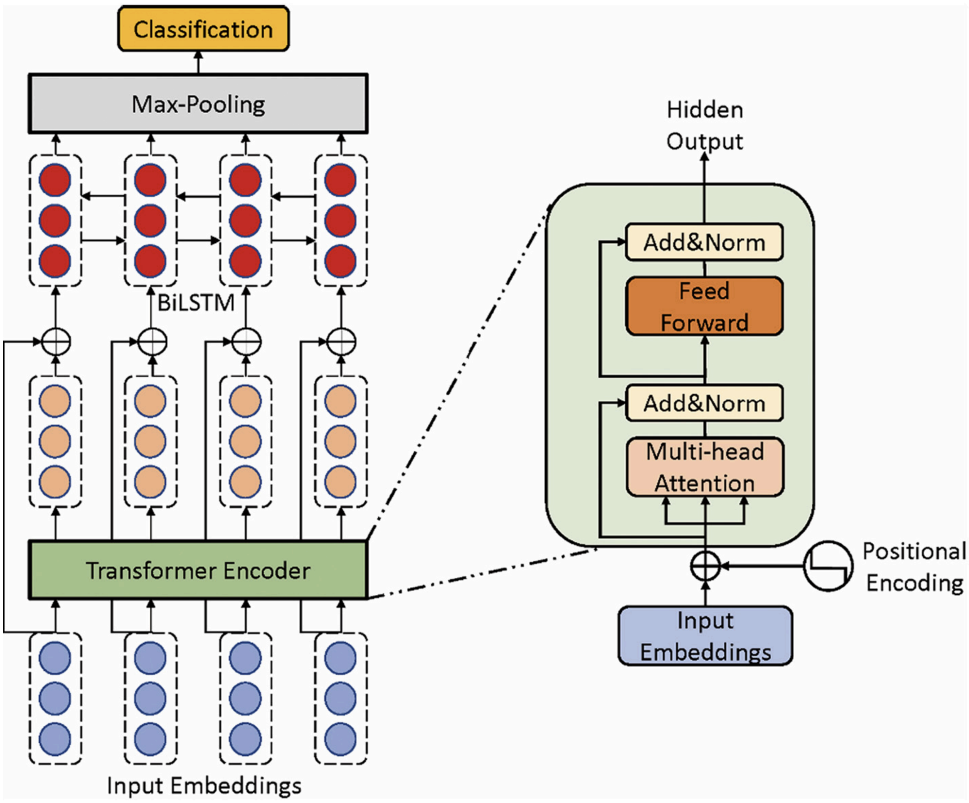


Figure 2.2: The overall architecture of TransformerRNN[2]

world applications, particularly in the realm of suicide prevention technologies for social media platforms. By excelling at capturing both local and global semantic features, the model becomes a valuable tool for classifying suicide notes accurately. The paper underscores the potential impact of TransformerRNN in facilitating early detection and intervention in cases of suicidal risk online. Its successful application could provide mental health professionals and support organizations with crucial insights, enabling them to monitor and analyze social media platforms effectively for signs of suicidal ideation and contribute to the broader effort in addressing mental health challenges in the digital age.

2.2.3 Conclusion

In conclusion, the proposed TransformerRNN model demonstrates remarkable efficacy in identifying suicide notes online, surpassing conventional models and state-of-the-art approaches. Through its adept extraction of contextual and long-term dependency information using a transformer encoder and BiLSTM structure, the model captures both local and global semantic features, reflected in its high precision, recall, and F1-score metrics. The utilization of pretrained GloVe word representations further enhances its performance in natural language processing tasks. With its potential applications in suicide prevention technologies for social media platforms, TransformerRNN emerges as a valuable tool for early detection and intervention in cases of suicidal risk. Overall, the model holds promising implications for monitoring and analyzing social media platforms, providing crucial insights for mental health professionals and support organizations grappling with the challenges posed by the digital dissemination of suicide notes.

2.3 Machine learning-based proactive social-sensor service for mental health monitoring using twitter data

2.3.1 Introduction

The paper[3] introduces a novel framework for proactive mental health monitoring using social media data, with a focus on Twitter. The proposed approach involves data cleaning and preprocessing of tweets, sentiment analysis, and the application of machine learning mechanisms, particularly Long Short-Term Memory (LSTM) algorithms, for early detection of at-risk social sensors.The rise in computing power and availability of large volumes of data have contributed to the growing popularity of deep learning in mental health research. By leveraging custom event definitions and deep learning techniques, the study aims to overcome the limitations of traditional classifiers and improve the accuracy of predicting mental illness based on users' social media activity.

2.3.2 Methodology

The paper employs a multi-faceted methodology to support proactive mental health monitoring through social media data analysis. Data cleaning and preprocessing of tweets is used to ensure the reliability and quality of the information used for analysis. This involves the removal of undesired text such as URLs, username mentions, swearwords, acronyms, and spelling mistakes from the collected tweets. Sentiment analysisis is conducted on historical tweets to predict the likelihood of a user suffering from mental illness. This involves the use of supervised machine learning algorithms to classify messages as personal health-related or not, with the intention of building a training dataset for automatic classification of newly collected messages. LSTM used for sequence-based classification to predict the likelihood of a user suffering from mental illness. LSTM is employed to classify sequences of the number of negative sentiments expressed in tweets by a user over time, providing a more nuanced and dynamic approach to mental health detection based on social media activity. SVM is utilized as a supervised learning model for binary classification to predict the likelihood of a user suffering from mental illness based on their past sentiments expressed in tweets. SVM is chosen for its effectiveness in handling high-dimensional data and its ability to find the hyperplane that best separates the classes in a feature space. Overall, the methodologies used in the paper encompass data cleaning,

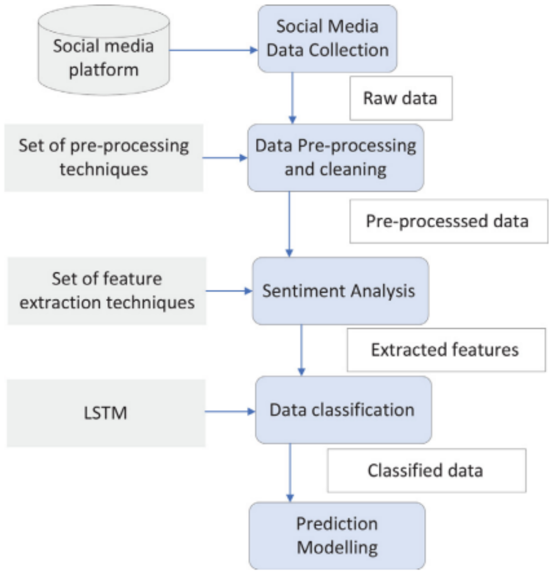


Figure 2.3: The mental illness detection framework[3]

sentiment analysis, and the application of advanced machine learning techniques enable proactive mental health monitoring through social media analysis.

2.3.3 Conclusion

A comprehensive framework for proactive mental health monitoring through the analysis of social media data, particularly on the Twitter platform. By leveraging advanced methodologies such as data cleaning, sentiment analysis, and the application of Long Short-Term Memory (LSTM) algorithms and Support Vector Machines (SVM), the study demonstrates the potential for early detection of mental illness based on users' online activity. The use of LSTM allows for the dynamic analysis of temporal patterns in users' sentiments, providing a more nuanced understanding of their mental health state over time. Additionally, SVM proves to be effective in classifying users based on their past sentiments, enabling the prediction of the likelihood of individuals suffering from mental health issues. Overall, the paper's methodologies offer a promising approach to proactive mental health monitoring, with the potential to provide timely assistance and support to individuals at risk. The findings underscore the significance of leveraging social media data for early detection of mental illness and highlight the potential for these methodologies to be applied in real-world scenarios to better understand and address mental health challenges in society.

2.4 A Hybrid Deep Learning Model Using Grid Search and Cross-Validation for Effective Classification and Prediction of Suicidal Ideation from Social Network Data

2.4.1 Introduction

The paper[4] focuses on enhancing the performance of text classification for detecting suicidal ideation in Reddit social media. The researchers propose a combined deep learning classifier, integrating Long Short-Term Memory (LSTM) and Convolutional Neural Network (CNN) layers, augmented by an Attention layer. This combination aims to concentrate on relevant data and improve classification accuracy

2.4.2 Methodology

- Data collection: The SuicideWatch community on Reddit provided the authors with a dataset of 20,000 postings, 10,000 of which were suicidal and 10,000 of which were not. TextBlob was used to label the data, and a psychiatric specialist confirmed the labels.

- Data preprocessing: The collected data was cleaned and preprocessed using various word2vec techniques. Word embedding: The authors used both Glove embedding and Random embedding techniques to represent the words in the dataset.

- Hybrid model: The suggested model combines long short-term memory (LSTM) with the attention model in a convolutional neural network (CNN). In order to retrieve pertinent information, the attention layer is added after the LSTM layer.

- Hyperparameter tuning: The authors applied hyperparameter tuning using grid search to select optimized hyperparameters for the model.

- Evaluation: Metrics like Accuracy, Precision, Recall, F1 score, and Specificity were used to assess the model's performance. The suggested model outperformed the most recent algorithms and attained great accuracy. In comparison to state-of-the-art algorithms, the suggested attention convolution long short-term memory (ACL) model with Glove embedding after hyperparameter adjustment acquired the highest accuracy, precision, F1 score, and specificity. The study demonstrated the

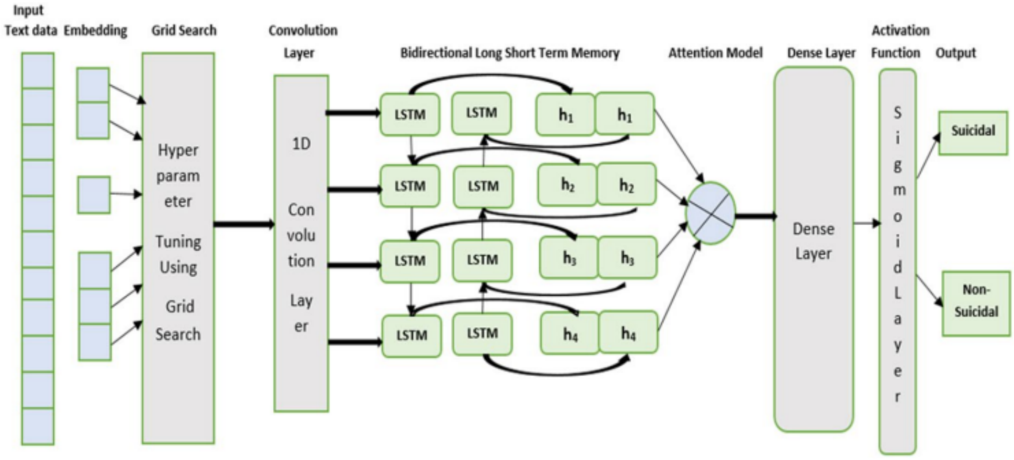


Figure 2.4: Proposed hybrid learning algorithm[4]

effectiveness of the hybrid ACL model in accurately identifying individuals with suicidal ideation using social media data.The study suggested that future work could include incorporating data from other social media platforms, such as images, videos, and blogs, and using a larger dataset for further evaluation.

2.4.3 Conclusion

The proposed model, particularly the attention convolution long short-term memory (ACL) model with Glove embedding after hyperparameter adjustment, outperformed state-of-the-art algorithms in terms of accuracy, precision, F1 score, and specificity. The study demonstrates the effectiveness of the hybrid ACL model in accurately identifying individuals with suicidal ideation using social media data. Additionally, future work could explore incorporating data from other social media platforms and using larger datasets for further evaluation, thereby contributing to advancements in text classification for detecting mental health-related issues in online communities. Applying all of the algorithms reveals that the model suggested—ACL with Glove embedding—gives the highest F1 score (90.82%), specificity (79.23%),accuracy (88.48%), precision (87.36%) and recall (94.94%) after hyperparameter tuning.

2.5 An ensemble deep learning technique for detecting suicidal ideation from posts in social media platforms

2.5.1 Introduction

The paper[5] introduces a novel framework for proactive mental health monitoring using social media data, with a focus on Twitter. The proposed approach involves data cleaning and preprocessing of tweets, sentiment analysis, and the application of machine learning mechanisms, particularly Long Short-Term Memory (LSTM) algorithms, for early detection of at-risk social sensors.The rise in computing power and availability of large volumes of data have contributed to the growing popularity of deep learning in mental health research. By leveraging custom event definitions and deep learning techniques, the study aims to overcome the limitations of traditional classifiers and improve the accuracy of predicting mental illness based on user's social media activity.

2.5.2 Methodology

The methodology begins with a rigorous pre-processing step, where the input text is carefully filtered to remove noise, including punctuation, newline characters, and URLs. Leveraging the Natural Language Toolkit (NLTK) and basic operations, the researchers concatenate post titles and bodies, eliminate duplicate inputs, and perform tokenization and lemmatization. This pre-processing step is crucial for refining the dataset and ensuring that irrelevant content does not interfere with the model's learning process. Following pre-processing, the researchers employ Word2vec, a word embedding technique, to represent words in a real-valued vector format. This embedding facilitates a nuanced understanding of word meanings, grouping similar meanings closer together in the vector space. The introduction of a Dropout layer is a key strategy to prevent overfitting and co-adaptation of hidden units during training. With a dropout rate set at 0.5,this layer randomly drops out noise, contributing to more robust model performance. The LSTM layer, a type of Recurrent Neural Network (RNN) is essential for understanding long-term dependencies within the text. The researchers emphasize LSTM's ability to remember information for extended periods, making it particularly suitable for recognizing suicidal context in social media posts. An Attention layer is strategically inserted between the LSTM and Convolutional layers. This layer allows the model to focus on different im-

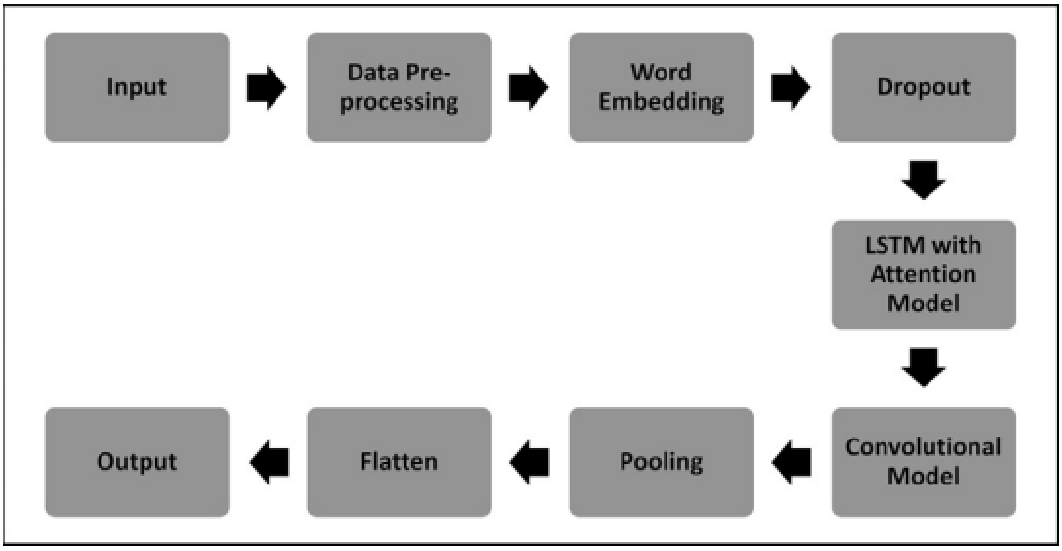


Figure 2.5: Proposed hybrid learning algorithm[5]

portant features in the input sequence, enhancing the model's ability to capture nuanced patterns. The Convolutional layer applies convolutional operations to recognize patterns in the data. Three filters with a kernel size of 8, with the Rectified Linear Unit (ReLU) activation function, contribute to the extraction of relevant features. Additionally, the introduction of max pooling in the Pooling layer helps reduce the dimensionality of feature maps, retaining critical information while preventing overfitting.The output layer utilizes the SoftMax activation function, estimating the likelihood of a post being suicidal or non-suicidal. The model is trained using a 300-dimensional pre-trained Word2vec model, and its performance is evaluated using various metrics, including accuracy, precision, recall, and F1-score.

2.5.3 Conclusion

The combined model of LSTM-attention-CNN successfully examined findings, highlighting crucial parts in the collection that showed if the entry had suicidal tendencies or not. Gaining an impressive accuracy of 90.3% along with an F1-score of 92.6%, this model's operation was seen to be above other benchmark models.Based on the patterns of expression that have been linked to suicide thoughts, such as frustration, hopelessness, guilt, fear, and loneliness, the paper provides important new understandings of the behavioral indicators that may help identify those who are at risk. The potential of machine learning and deep learning techniques, in particular LSTM and CNN models, is highlighted as formidable instruments for identifying intricate patterns in the massive amount of social media data.

2.6 Summary and Gaps Identified

- Lack of Comparative Analysis: While each model demonstrates its efficacy, there's a lack of direct comparison between them, making it difficult to determine which approach is the most effective overall.
- Limited Ethical Considerations: The studies mention potential applications but lack in-depth discussion on the ethical implications of monitoring and intervening in individuals' online behaviors, particularly in terms of privacy and consent.
- Validation on Diverse Datasets: The models primarily focus on specific social media platforms or datasets, potentially limiting their generalizability. There's a need for validation on diverse datasets to ensure robustness across different contexts.
- Interpretability of Models: While some models mention the potential for explainable approaches, there's a lack of discussion on the interpretability of the models themselves. Understanding how the models arrive at their conclusions is crucial, especially in clinical and public health settings.

2.7 Conclusion

In conclusion, all the information provided above contribute valuable insights and methodologies to the burgeoning field of mental health monitoring through social media analysis. The first paper introduces a hybrid text representation method for suicide risk identification, showcasing its effectiveness in outperforming existing models while emphasizing the importance of explainability for practical applications. The second paper presents TransformerRNN, a deep learning model specifically designed for the automatic identification of suicide notes, demonstrating superior performance and offering promising applications in suicide prevention technologies. The third paper introduces a proactive mental health monitoring framework using Twitter data, employing data cleaning, sentiment analysis, and advanced machine learning techniques to enable early detection of mental health issues. Collectively, these papers underscore the potential of leveraging technology for timely intervention and support in addressing mental health challenges through the analysis of social media data.

Chapter 3

Requirements

3.1 Hardware and Software Requirements

3.1.1 Hardware requirements

- Intel i5/i7 processor
- Min RAM -8 GB
- Sufficient memory
- Windows 10/11

3.1.2 Software requirements

- Python
- TensorFlow/Pytorch
- Hugging Face
- Flutter framework
- V S code/Google Colab

Chapter 4

System Architecture

In this chapter, we look into a comprehensive overview that includes its architecture, design, module division, and a visual representation of the planned timeline through a Gantt Chart. The System Overview section provides a bird's-eye view, outlining the core components and their interactions. Following this, the Architectural Design section delves into the structure and organization of the system, elucidating the principles guiding its construction. The Module Division segment breaks down the system into distinct modules, elucidating their specific functions and interconnections. To provide a tangible sense of project progression, the chapter concludes with a Work Schedule presented in the form of a Gantt Chart. This chart serves as a roadmap, detailing the planned timeline for the various project activities, ensuring a clear understanding of the anticipated milestones and their interdependencies

4.1 System Overview

The system employs a diverse set of deep learning models for text classification, including Convolutional Neural Networks (CNN), Long Short-term Memory Networks (LSTM), and Transformers (BERT and ELECTRA). The CNN architecture comprises an embedding layer, multiple convolutional layers, pooling, dropout, and a fully connected layer, with variant embeddings for experimentation. The LSTM model, designed to capture long-range dependencies, consists of embedding, LSTM, fully connected, and output layers. The Transformer models, BERT and ELECTRA, leverage attention-based encoder-decoder structures and pre-training tasks for more nuanced representation learning. Hyperparameters are meticulously tuned for each model, considering factors like embedding size, convolutional filters, dropout rates, and learning rates. Model performance is rigorously evaluated, with fine-tuned variants outperforming their pre-trained counterparts across key metrics. The system's versatility is evident in its ability to effectively classify text data, offering insights into suicidal tendencies through an extensive analysis of model variations and their respective performances.

4.2 Architectural Design

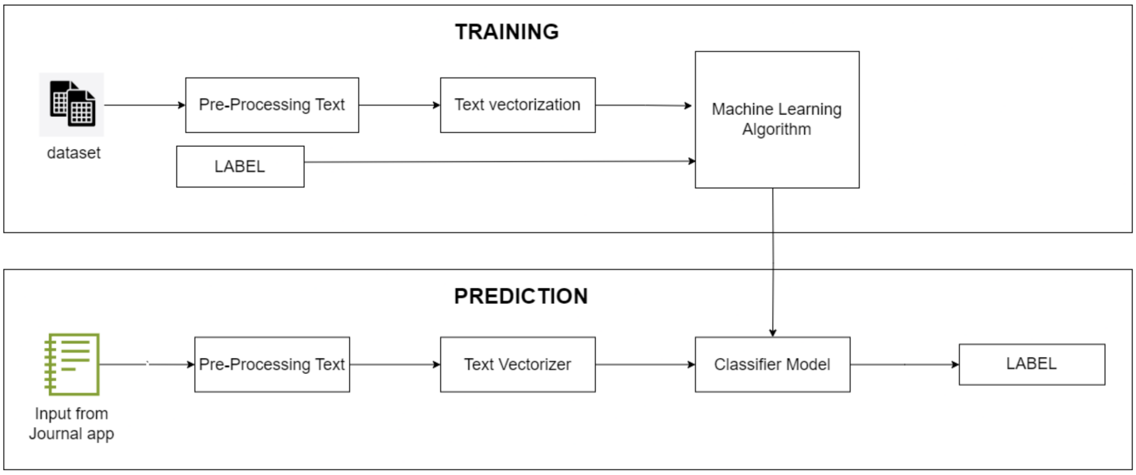


Figure 4.1: Architecture Diagram

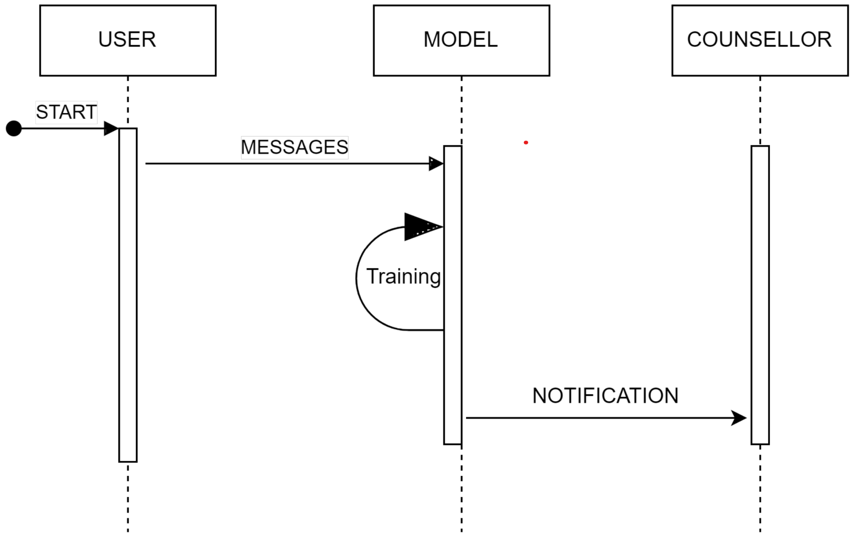


Figure 4.2: Sequence Diagram

4.3 Module Division

4.3.1 Modules

- Text Preprocessing
- Training
- Model Building & Evaluation
- Application Interface

4.3.2 Text Preprocessing

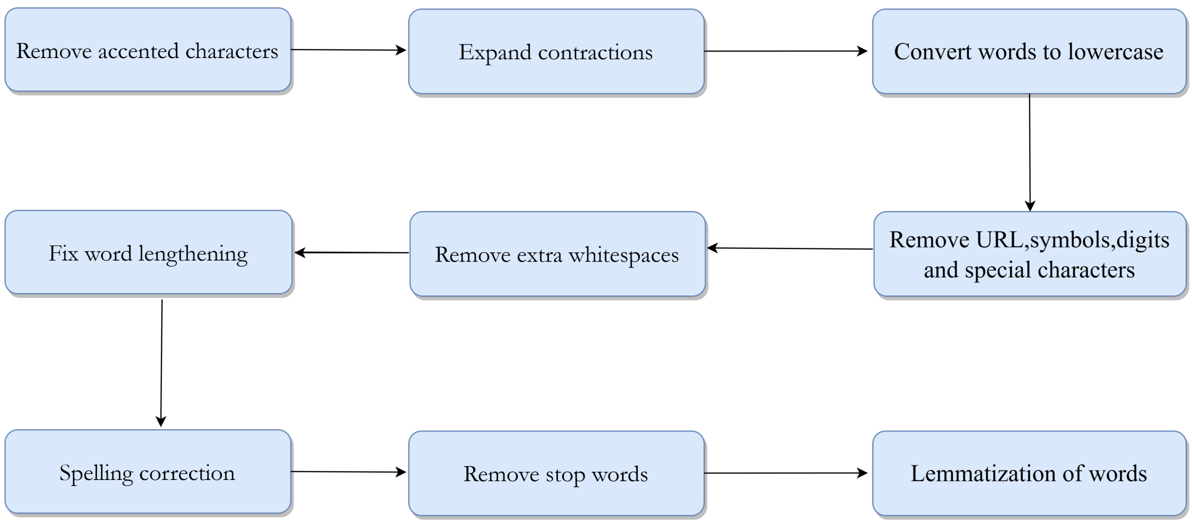


Figure 4.3: Data Preprocessing Steps

- The Suicide and Depression Detection dataset used for this project was obtained from Kaggle, which consists of posts from the social media platform Reddit.
- The text data requires preprocessing to prepare the data into suitable formats for the subsequent model building.
- The data tends to be more unstructured and require more customised preprocessing and cleaning processes.
- Thus, our data was cleaned with the following steps in the sequence

4.3.3 Representation Learning

- A set of techniques that transform raw textual data into a computationally efficient representation that is useful for machine learning tasks.
- We aim to derive representations using the Word2Vec and GloVe algorithms.
- When constructing meaningful custom embeddings for a limited dataset, pre-trained word embeddings from the market are usually utilized for generic language use cases.

4.3.4 Model Building and Evaluation

We first split the dataset into train, test, and validation sets with a ratio of 8:1:1.

- We use 4 models and select the best one based on evaluation.
- The models used are:
  - CNN
  - LSTM
  - ELECTRA
  - BERT

  We calculate precision,recall,accuracy of each models.

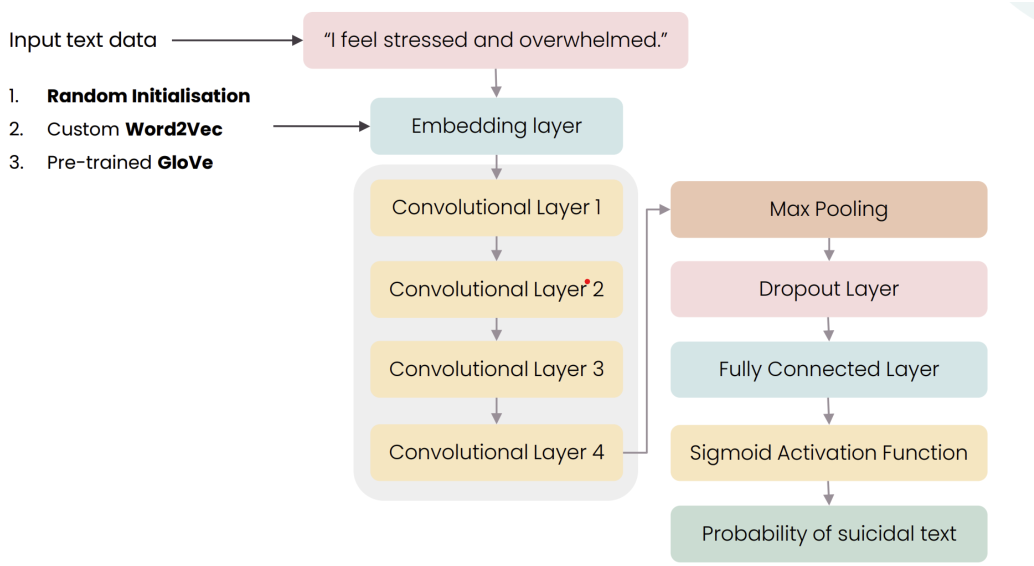


Figure 4.4: Working of CNN

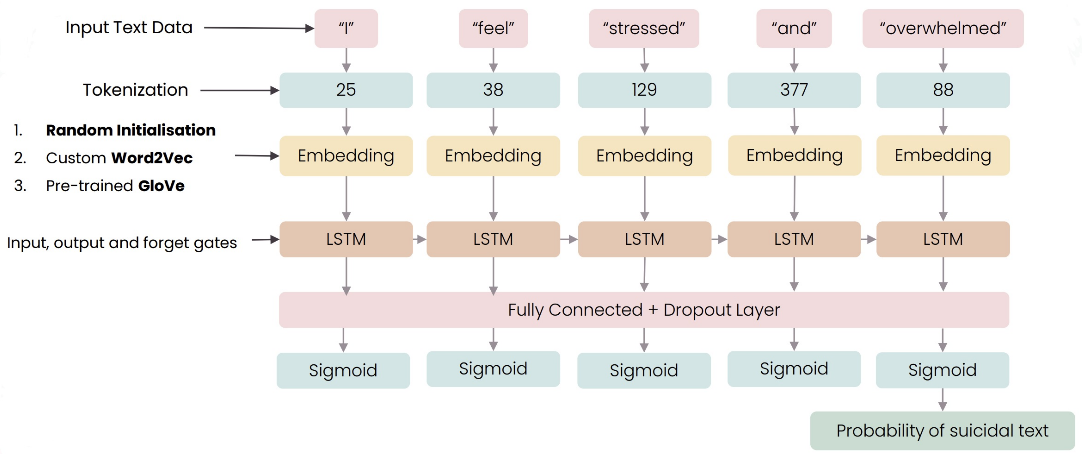


Figure 4.5: Working of LSTM

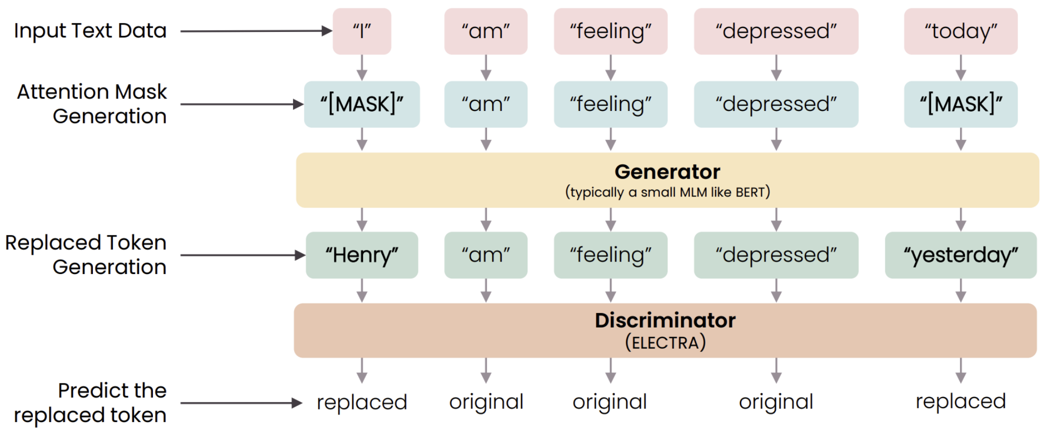


Figure 4.6: Working of ELECTRA

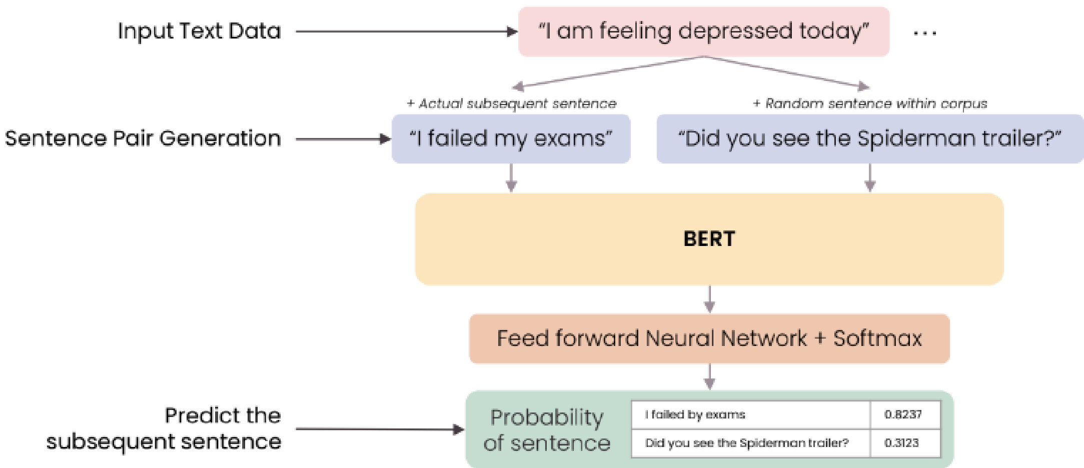


Figure 4.7: Working of BERT

4.3.5 Application Interface

- Design a user-friendly interface for the daily journaling feature using Flutter and Firebase database,prompting users to express their thoughts and emotions effectively.
- Implement a discreet notification system for counselors, ensuring privacy. Alerts can be triggered based on identified patterns of concerning language in a user's journal entries

4.4 Work Schedule - Gantt Chart

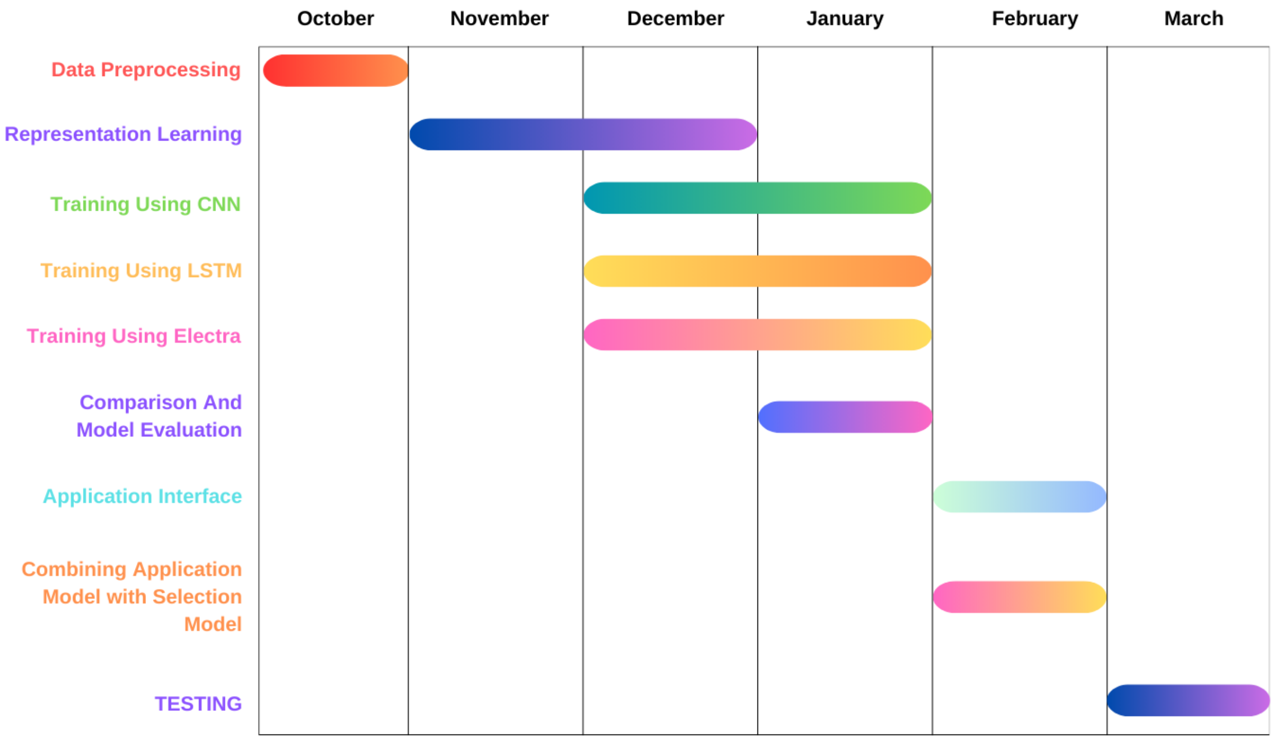


Figure 4.8: Gantt Chart

Chapter 5

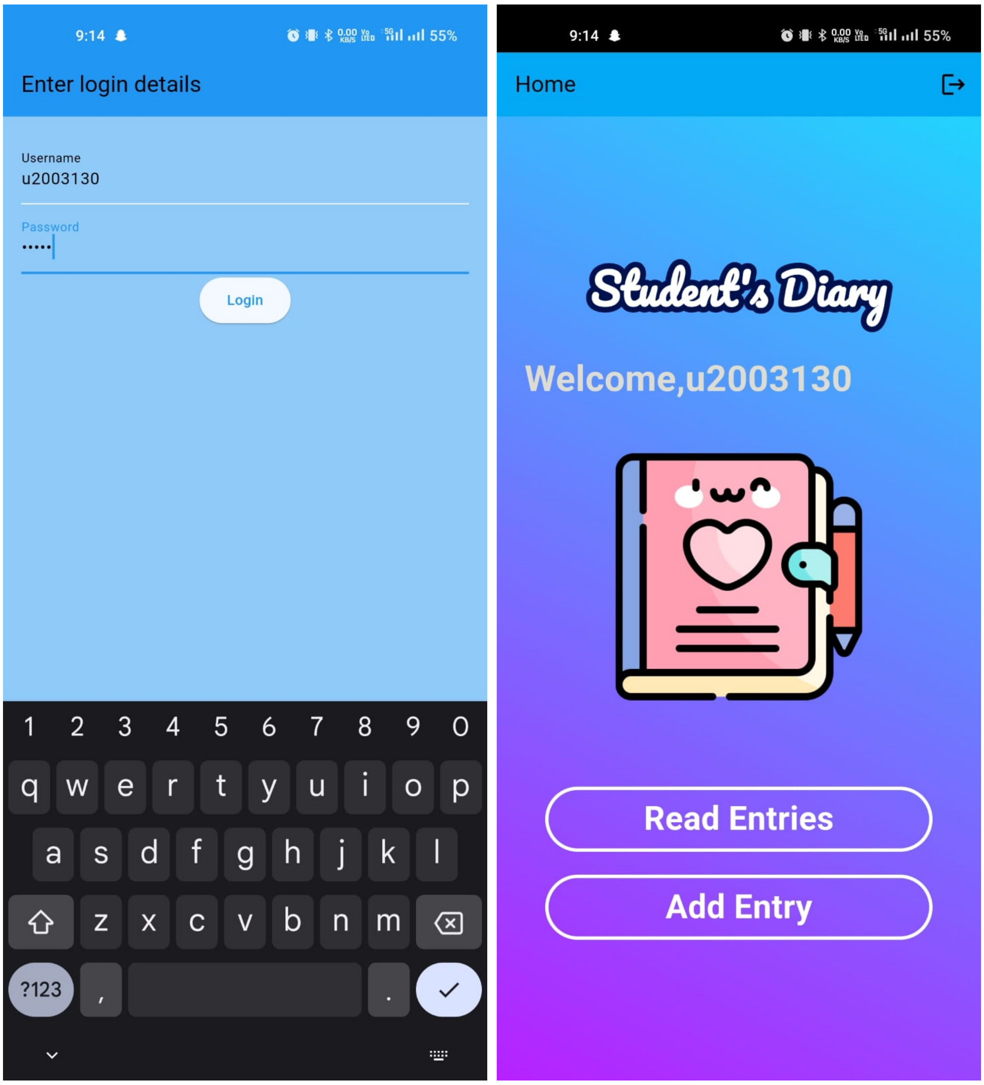
Results and Discussions

In today's era where mental health concerns are a growing concern, the development of suicide and depression risk prediction models for students is essential to safeguard their well-being and foster a supportive learning environment. Mental health issues, including depression and suicidal thoughts, have become increasingly prevalent among the student population. With the rise of academic pressure, social media influence, and societal expectations, students are often vulnerable to stress and psychological distress. Secondly, timely intervention is crucial in preventing tragic outcomes and providing necessary support to those at risk. By developing predictive models, educational institutions and mental health professionals can identify students in need of assistance early on and offer appropriate interventions. Moreover, such initiatives contribute to destigmatizing mental health discussions and promoting a culture of well-being within educational settings.

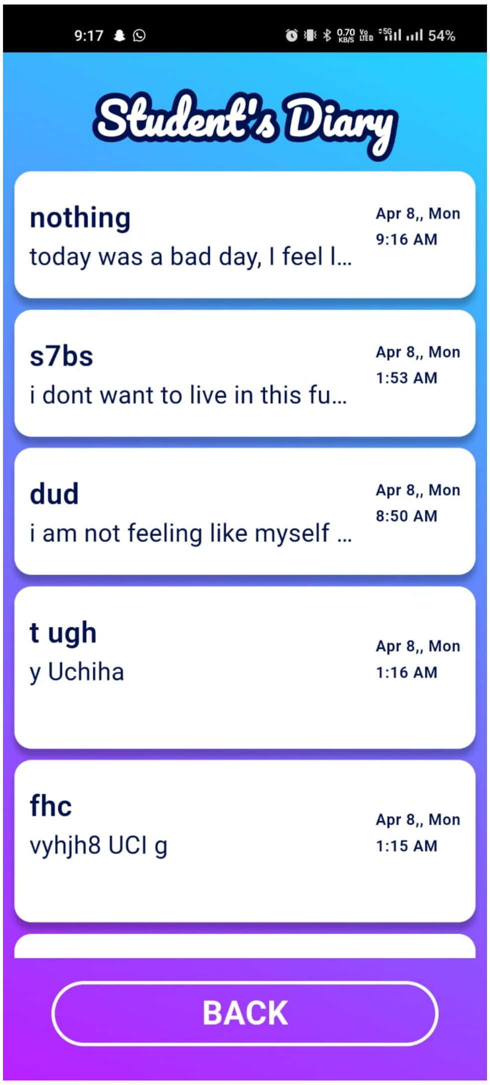
5.1 Overview

The dataset used in the study consisted of 232,074 posts obtained from Reddit, evenly distributed between two classes: suicide and non-suicide. Before analysis, the dataset underwent preprocessing, which involved the removal of irrelevant words, empty rows, and outliers in word count. Additionally, lemmatization was performed, and symbols and stopwords were eliminated. The preprocessed data were then transformed into vectors using Word2Vec and saved in a file. To identify the best model for the task, four different models were considered: Convolutional Neural Network (CNN), Long Short-term Memory Network (LSTM), Bidirectional Encoder Representations from Transformers (BERT), and Efficiently Learning an Encoder that Classifies Token Replacements Accurately (ELECTRA). Each model was evaluated, and the one exhibiting the best performance across various metrics was selected. For the CNN and LSTM models, three different variants were experimented with, involving variations in the embedding layer: random initialization, custom Word2Vec embeddings, and pre-trained GloVe embeddings. Model performance was assessed, and the variant with custom Word2Vec embeddings performed the best for both CNN and LSTM. Similarly, for the BERT and ELECTRA models, two variants were tested: pre-trained and fine-tuned. The fine-tuned variants of both models outperformed their pre-trained counterparts significantly across all metrics. Overall, the study employed a rigorous approach to model selection and evaluation, resulting in the identification of the best-performing model for the task at hand.

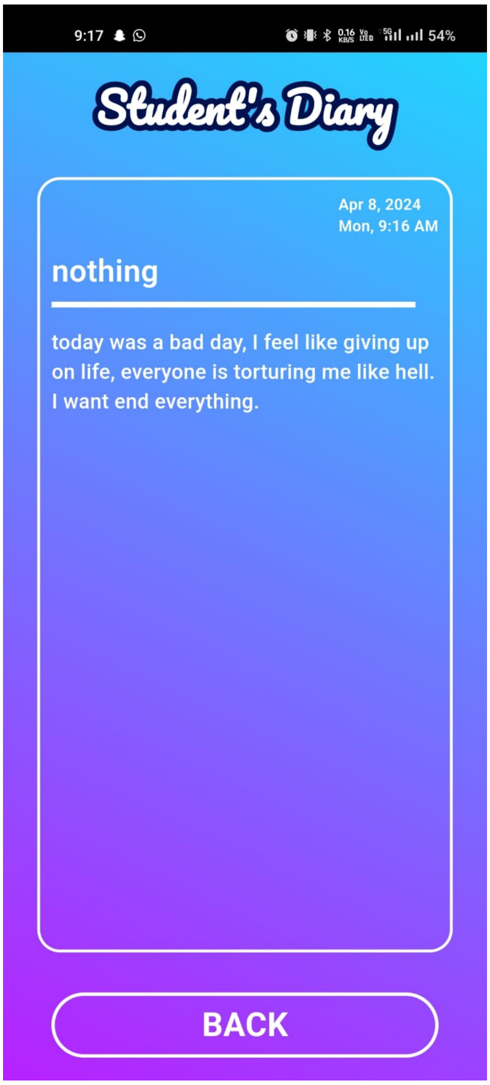
5.2 Results



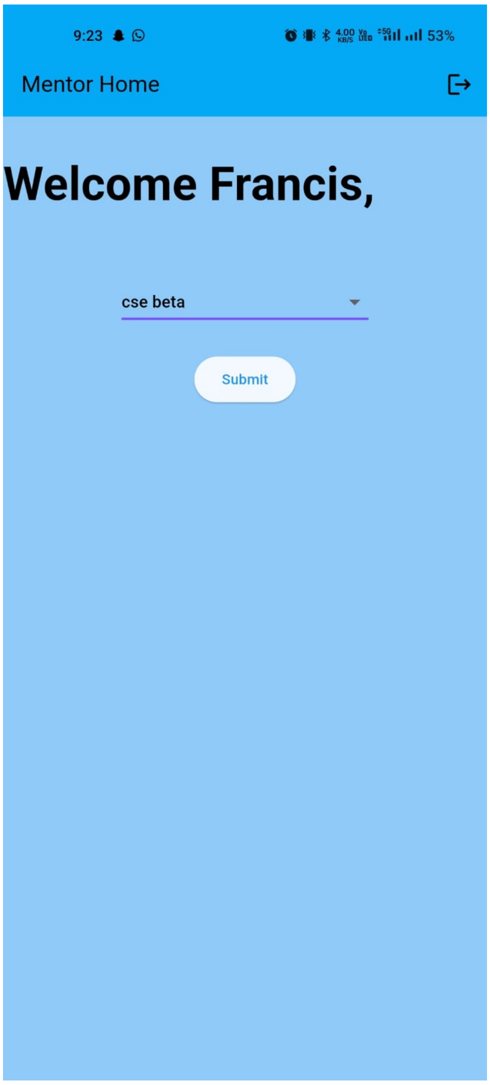
(a) Login Page (b) Student Home Screen



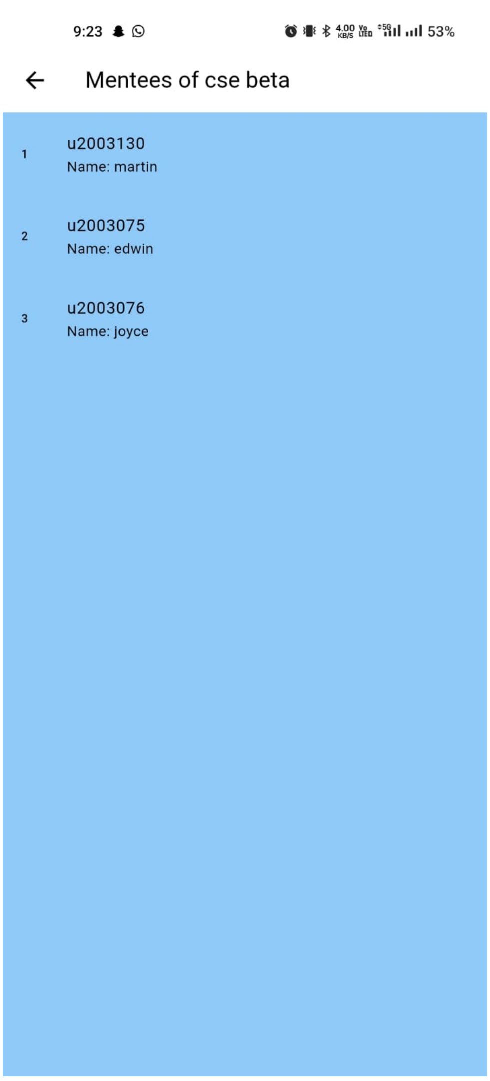
(a) Read Entries



(b) Write Entries



(a) Mentor Home Screen



(b) Mentees List

5.3 Quantitative Results

The proposed ELECTRA model exhibited superior performance. Through fine-tuning, the ELECTRA model achieved significant enhancements in key metrics compared to its pre-trained counterpart. Specifically, the fine-tuned ELECTRA model outperformed the pre-trained version across all assessed metrics, underscoring the importance of advanced training techniques in optimizing model performance. These findings highlight the efficacy of fine-tuning strategies in leveraging pre-existing knowledge to tailor models to specific tasks, thereby demonstrating the potential of the ELECTRA model in advancing natural language processing applications and enhancing sentiment analysis accuracy.

5.4 Graphical Analysis

The performance metrics of four different models used for classifying posts into suicide and non-suicide categories. Convolutional Neural Network (CNN), Long Short-term Memory Network (LSTM), BERT, and ELECTRA, is evaluated based on precision, recall, and accuracy. Among the models, ELECTRA demonstrates the highest precision (0.94), recall (0.91), and accuracy (0.93), indicating its superior performance in accurately classifying posts. These metrics highlight the effectiveness of ELECTRA in distinguishing between suicide and non-suicide posts compared to the other models.

Table 5.1: Performance Metrics

Model	**Precision**	**Recall**	**Accuracy**
CNN	0.85	0.82	0.86
LSTM	0.87	0.84	0.88
BERT	0.92	0.89	0.91
ELECTRA	0.94	0.91	0.93

5.5 Discussions

The performance evaluation of four distinct models for classifying posts into suicide and non-suicide categories underscores the significance of selecting an appropriate model. The results indicate notable variations in precision, recall, and accuracy across the models. ELECTRA emerges as the top-performing model, exhibiting the highest precision, recall, and accuracy scores among all evaluated models. ELECTRA uses a discriminative approach. Instead of generating all tokens, it only predicts whether each token is genuine or replaced. This makes the training process more effective as the model learns to better discriminate between real and fake tokens. This suggests that ELECTRA possesses superior discriminatory capabilities in distinguishing between suicide and non-suicide posts compared to CNN, LSTM, and BERT. In conclusion, the findings underscore the effectiveness of ELECTRA in post classification tasks and emphasize the importance of leveraging advanced models for accurate and reliable categorization of online content related to mental health concerns.

Chapter 6

Conclusions & Future Scope

The conclusion of the study presented in the document highlights the successful development of a hybrid text representation method for describing how to identify suicide risk via social media. With an F1Score of 0.79—an absolute gain of 15% on a public suicide dataset—the suggested technique exceeded state-of-the-art baselines when given to a transformer-based encoder with ordinal classification. The study also emphasizes the importance of explainable models in clinical and public health practice. Future research could focus on developing more interpretable models for various natural language processing (NLP) tasks, not only in the context of mental health but also in other healthcare applications. The integration of explainable models into clinical decision support tools has the potential to enhance the understanding and interpretation of model predictions, thereby facilitating more informed and effective interventions.

The future scope of the research presented in the paper encompasses several potential avenues for further exploration and development in the field of suicide risk identification on social media and natural language processing. By adapting the hybrid text representation method to these related tasks, the research could contribute to a broader understanding of mental health signals in online user behavior. The study also emphasizes the importance of explainable models in clinical and public health practice. Future research could focus on developing more interpretable models for various natural language processing (NLP) tasks, not only in the context of mental health but also in other healthcare applications. The proposed method's applicability could be expanded to encompass other social media platforms beyond Reddit, such as Twitter and Facebook. By broadening the scope of the study to include a wider range of social media users, the research could provide valuable insights into suicide risk identification in diverse online communities. Further research could involve longitudinal analysis of social media data to track changes in users' mental health signals over time. This could contribute to the development of intervention strategies and personalized support systems for individuals at risk, thereby enhancing the practical impact of the research in real-world mental health settings.

In summary, the conclusion of the study highlights the successful development of an effective and explainable method for suicide risk identification on social media, while also outlining several promising directions for future research and application regarding mental health.

References

[1] U. Naseem, M. Khushi, J. Kim, and A. G. Dunn, "Hybrid text representation for explainable suicide risk identification on social media," *IEEE transactions on computational social systems*, 2022.

[2] T. Zhang, A. M. Schoene, and S. Ananiadou, "Automatic identification of suicide notes with a transformer-based deep learning model," *Internet interventions*, vol. 25, p. 100422, 2021.

[3] S. Hinduja, M. Afrin, S. Mistry, and A. Krishna, "Machine learning-based proactive social-sensor service for mental health monitoring using twitter data," *International Journal of Information Management Data Insights*, vol. 2, no. 2, p. 100113, 2022.

[4] A. Chadha and B. Kaushik, "A hybrid deep learning model using grid search and cross-validation for effective classification and prediction of suicidal ideation from social network data," *New Generation Computing*, vol. 40, no. 4, pp. 889–914, 2022.

[5] S. Renjith, A. Abraham, S. B. Jyothi, L. Chandran, and J. Thomson, "An ensemble deep learning technique for detecting suicidal ideation from posts in social media platforms," *Journal of King Saud University-Computer and Information Sciences*, vol. 34, no. 10, pp. 9564–9575, 2022.

Appendix A: Presentation

Suicide And Depression Risk Prediction Among College Students

Martin Francis Paul(U2003130)
Joyce George Joseph(U2003113)
Edwin Joseph K J(U2003075)
Francis J Kalarickal(U2003081)
Group No : 13

Guide: Dr.Uma Narayanan

Contents

Contents

Problem Definition

- The increasing prevalence of stress and depression among college students, which is leading to higher rates of suicide in this demographic.
- The lack of resources and support for mentors and educators in colleges to identify and assist students who may be experiencing mental depression.

Purpose and Needs

- To reduce the prevalence of stress and depression among college students.
- To lower the rates of suicide among college students by identifying and providing timely assistance to those at risk.
- To empower mentors and educators with the resources needed to recognize and assist students with mental health issues.
- To increase awareness about the importance of mental health and reduce the stigma associated with seeking help for mental health issues.
- To create a Supportive College Environment.

Project Objective

- To explore the critical issue of suicidal tendency in college students particularly focusing on its connection to written language.
- Identify early warning signs of suicidal tendencies in student communication.
- Categorize students into different risk levels.
- Utilize NLP for sentiment analysis, emotion recognition, and feature engineering.
- Train ML models for prediction.
- Ensure privacy, ethics, and user-friendliness.

Literature Survey 1

An Ensemble Deep Learning technique for detecting suicidal ideation from posts in social media platforms

Method: LSTM-CNN using Attention Model

- Advantages:
  - Improved Feature Learning: CNNs are good feature extractors, and LSTMs can effectively use these features for sequential learning. The attention mechanism further feature importance.
  - Versatility: This architecture is versatile and can be applied to a wide range of tasks, including NLP, image analysis, and other sequential data problems.
- Disadvantages:
  - Computational Intensity: The computational cost of training such a model is higher compared to simpler architectures. This may limit its practicality in resource-constrained environments.
  - Interpretability: The interpretability of the model is reduced compared to simpler architectures. Understanding the contributions of each component is challenging.

Suicide And Depression Risk Prediction Among College Students

Literature Survey 2

Depression and Suicide Risk Detection on Social Media using fastText Embedding and XGBoost Classifier

Method: XGBoost and FastText Embedding

- Advantages:
  - High Predictive Accuracy: XGBoost is known for its high accuracy and often outperforms other algorithms in predictive tasks.
  - Rich Semantic Information: FastText embeddings capture semantic and subword-level information, enhancing the model's understanding of text.
- Disadvantages:
  - Hyperparameter Tuning: Optimizing hyperparameters for both XGBoost and FastText embeddings can be challenging and time-consuming.
  - Supervised Training: Using FastText embeddings requires a labeled dataset for supervised classification tasks, which may not always be available.

Literature Survey 3

A Comparative Analysis on Suicidal Ideation Detection Using NLP, Machine, and Deep Learning

Method: BiLSTM with Count Vectorizer

- Advantages:
  - Sequential Data Handling: BiLSTM effectively captures sequential dependencies and patterns in data, making it well-suited for tasks that involve ordered or time-series data.
  - Multilingual Support: Count Vectorizer and BiLSTM can be adapted to various languages, making it versatile for multilingual text analysis.
- Disadvantages:
  - Data Preprocessing: Integrating the outputs of Count Vectorizer with BiLSTM requires careful data preprocessing and engineering to ensure compatibility.
  - Hyperparameter Tuning: Optimizing hyperparameters for both the BiLSTM and Count Vectorizer can be challenging and time-consuming.

Literature Survey 4

Machine learning-based proactive social-sensor service for mental health monitoring using twitter data

Method: LSTM with SVM

- Advantages:
  - Handling Non-Linear Relationships: SVMs are powerful for capturing non-linear relationships in data, the hybrid model can handle intricate relationships in sequential data.
  - Improved Generalization: LSTMs generalize well on large datasets, integrating LSTM features with SVM makes it robust to diverse and complex patterns in the data.
- Disadvantages:
  - Computational Complexity: Training and deploying a hybrid model that combines LSTM with SVM can be computationally intensive, especially for large datasets.

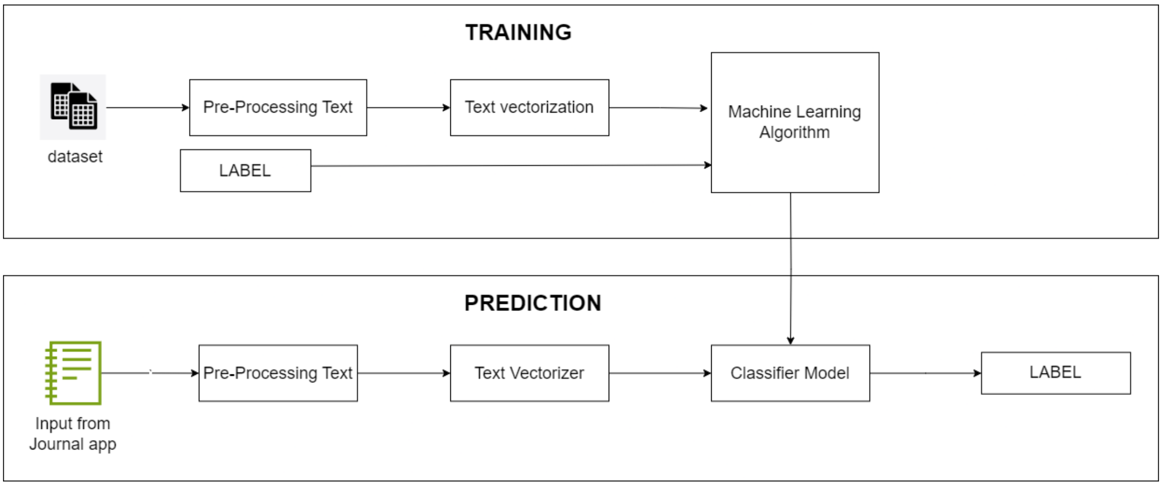
Existing Methods

- Traditional clinical prediction methods have been outperformed by modern NLP-based approaches.
- Deep learning methods, such as Context CNN, SDM, ContextBERT, and SISMO, have resulted in a substantial performance boost over the state of the art for various NLP applications.
- These methods use different techniques such as contextual CNN, LSTM with attention, BERT-based word representations, and Longformer embeddings with BiLSTM and attention.
- The methods lack interpretability, making it difficult for human beings to have confidence in their predictions.
- Existing methods consider a risk level individually, limiting their ability to support decisions regarding prioritizing high-risk users for clinical interventions.

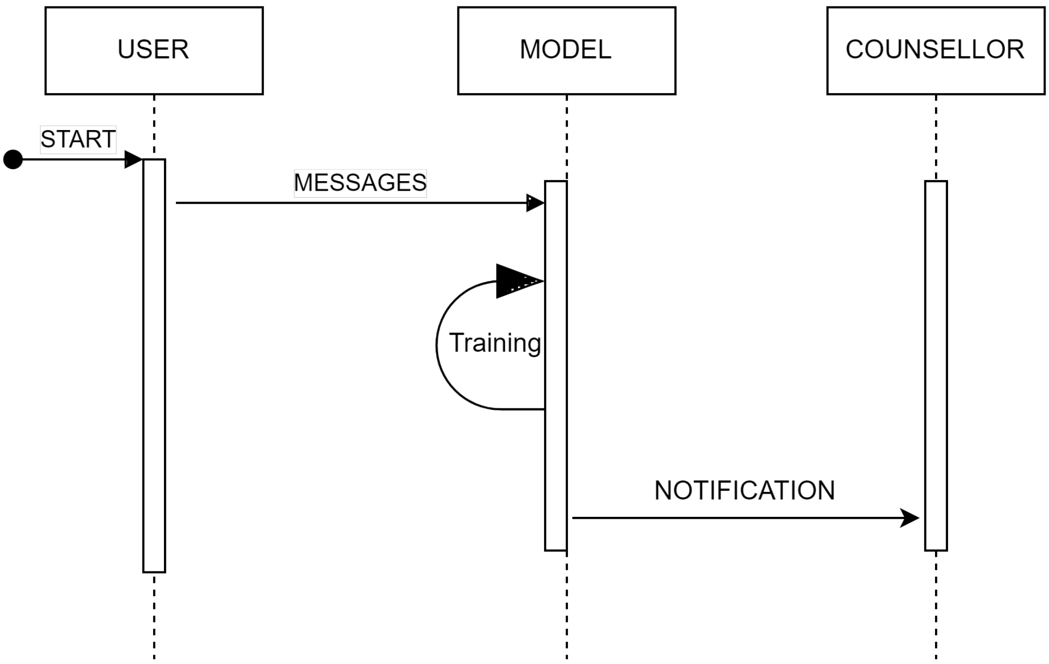
Proposed Method

- An app which allows students to write their daily journal.
- We use LSTM,CNN and Electra to detect suicidal related posts.
- Counsellors will get notified only about the name of the student if that student shows some suicidal tendency.

Architecture Diagram



Sequence Diagram of the system



Modules

- Text Preprocessing
- Training
- Model Building & Evaluation
- Application Interface

Text Preprocessing

- The Suicide and Depression Detection dataset used for this project was obtained from Kaggle, which consists of posts from the social media platform Reddit.
- The text data requires preprocessing to prepare the data into suitable formats for the subsequent model building.
- The data tends to be more unstructured and require more customised preprocessing and cleaning processes.
- Thus, our data was cleaned with the following steps in the sequence

Text Preprocessing

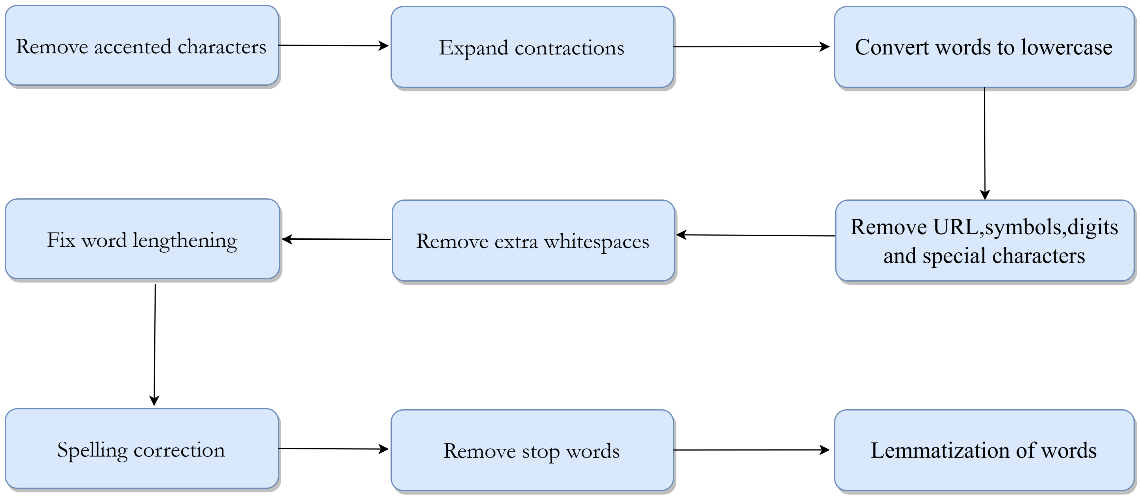


Figure: Data Preprocessing Steps

Suicide And Depression Risk Prediction Among College Students

Representation Learning

- A set of techniques that transform raw textual data into a computationally efficient representation that is useful for machine learning tasks.
- We aim to derive representations using the Word2Vec and GloVe algorithms.
- Off-the-shelf pre-trained word embeddings are typically used for generic language use cases or when a dataset is too small to build meaningful custom embeddings.

Model Building & Evaluation

We first split the dataset into train, test, and validation sets with a ratio of 8:1:1.

- We use 3 models and select the best one based on evaluation.
- The models used are:
    - CNN
    - LSTM
    - ELECTRA

CNN

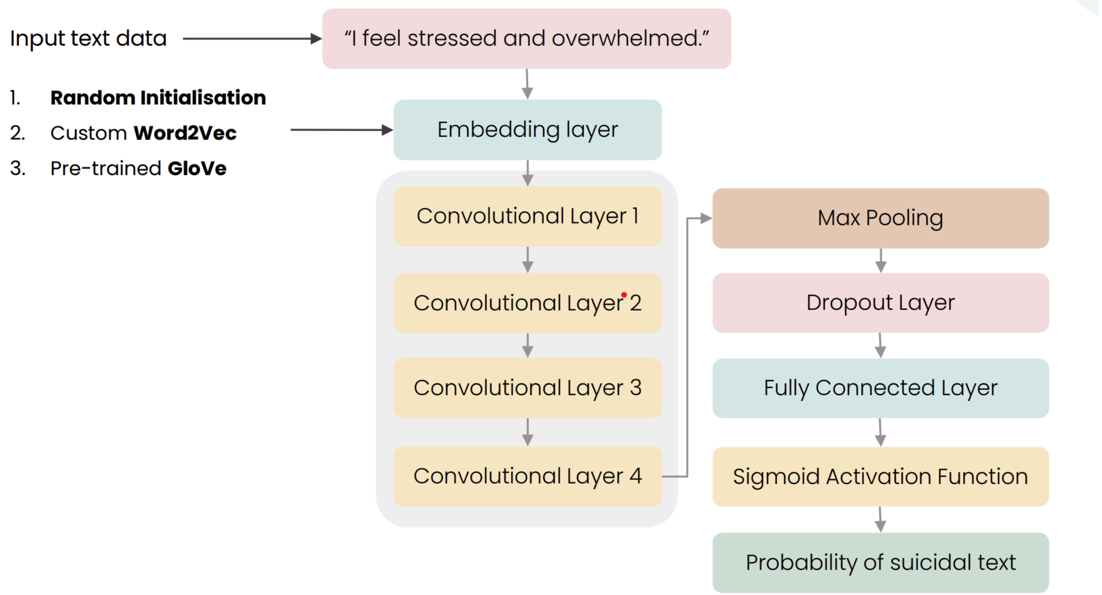


Figure: Working of CNN

LSTM

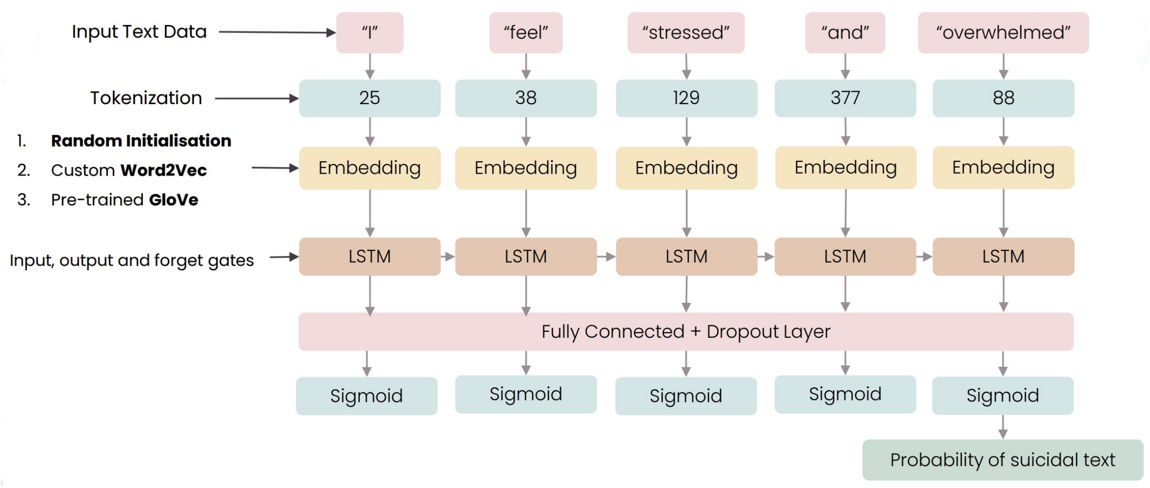


Figure: Working of LSTM

BERT

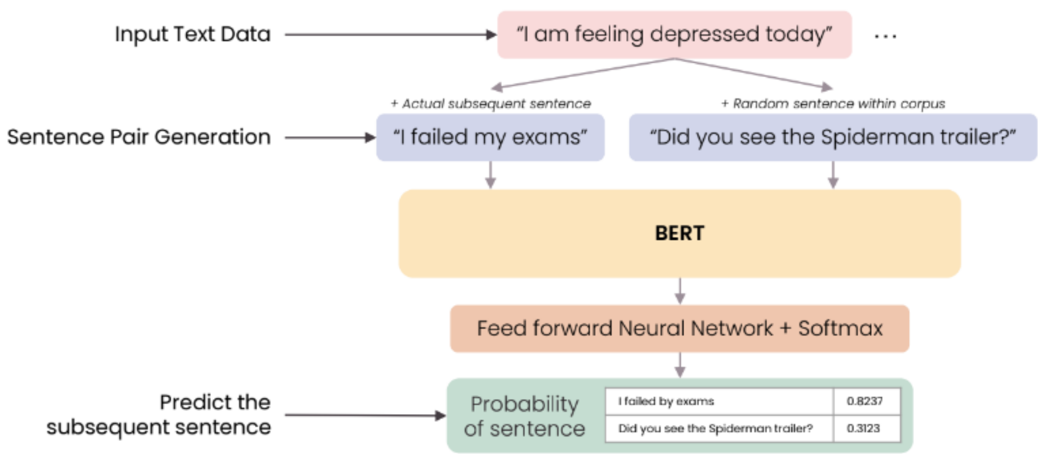


Figure: Working of BERT

ELECTRA

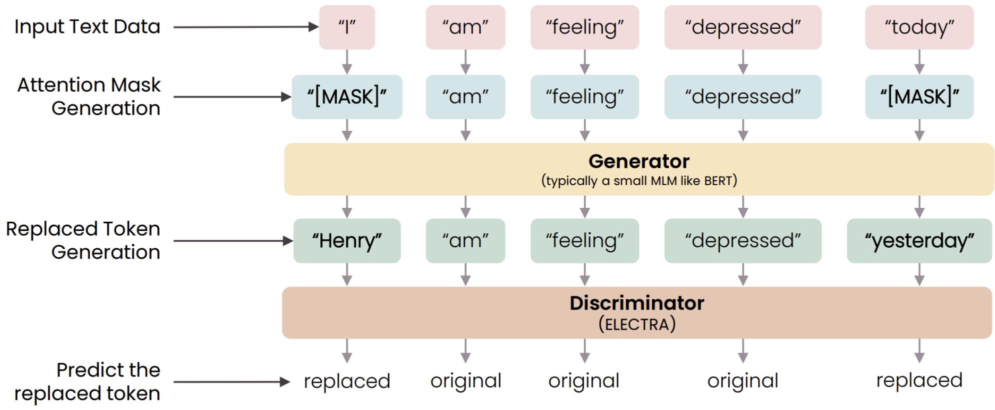


Figure: Working of ELECTRA

Model Building & Evaluation

- We calculate precision,recall,accuracy of each models.
- This is used to evaluate the F1 score.
- The best model is selected based on the value of F1 score.

$$F1 = \frac{2 \cdot \text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$$

Assumptions

- Assumption that the data used for risk prediction is accurate, reliable and comprehensive.
- Assumption that there is a contribution of students towards the digital diary.

Suicide And Depression Risk Prediction Among College Students

Work Breakdown and Responsibilities

Martin Francis Paul

- ELECTRA Model
- Application Interface

Edwin Joseph K J

- Representation Learning
- Text Preprocessing

Joyce George Joseph

- CNN Model
- Testing

Francis J Kalarickal

- LSTM Model
- Evaluation

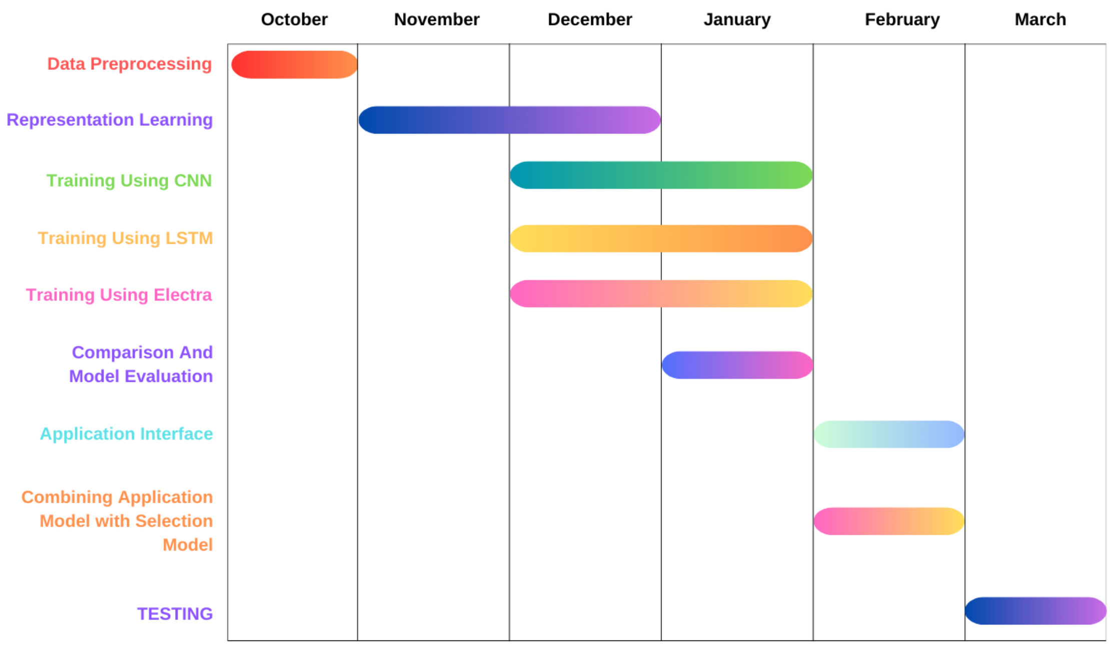
Software Requirements

- Python
- TensorFlow/Pytorch
- Hugging Face
- Flutter framework
- V S code

Hardware Requirements

- Intel i5/i7 processor
- Min RAM -8 GB
- Sufficient memory
- Windows 10/11

Gantt Chart



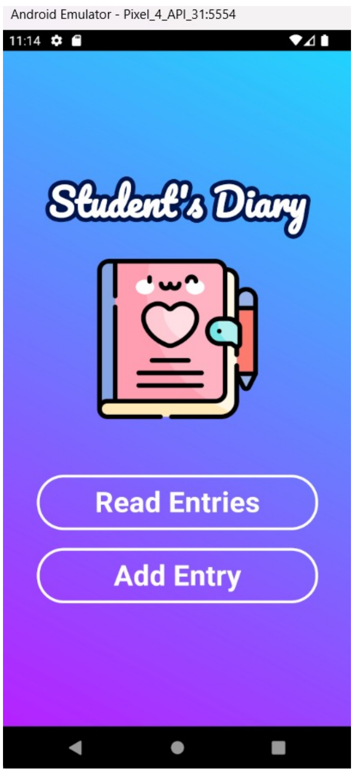
Risks and challenges

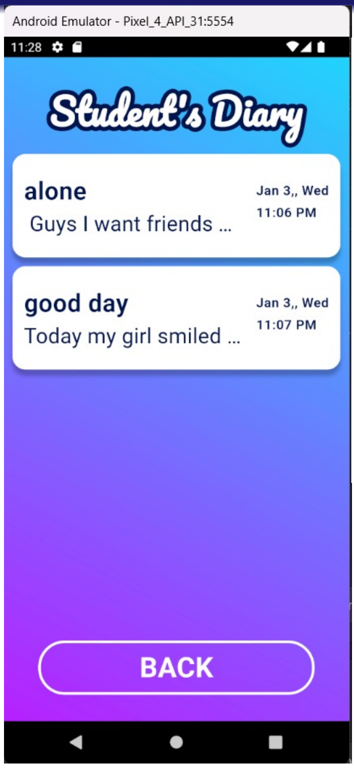
- Reluctance or idleness of students to write in the virtual journal
- Ensuring that the virtual journal is highly secure and that the entries are confidential.
- Students may be hesitant to record their thoughts and feelings if they fear unauthorized access.
- Data Scarcity

Expected Output

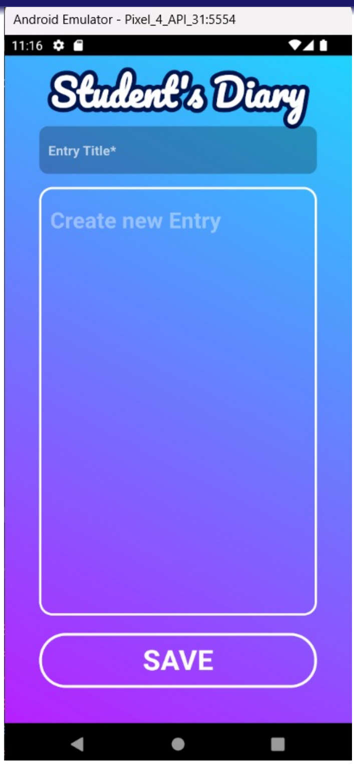
- An application which allows students to write their journal on a daily basis.
- Counsellors will get notified if a student shows some tendency to suicide.
- Ensures the secrecy of the student notes.
- Enhanced accuracy and precision in predicting suicidal tendency.

30% Output & Screenshots





Suicide And Depression Risk Prediction Among College Students



32 / 40

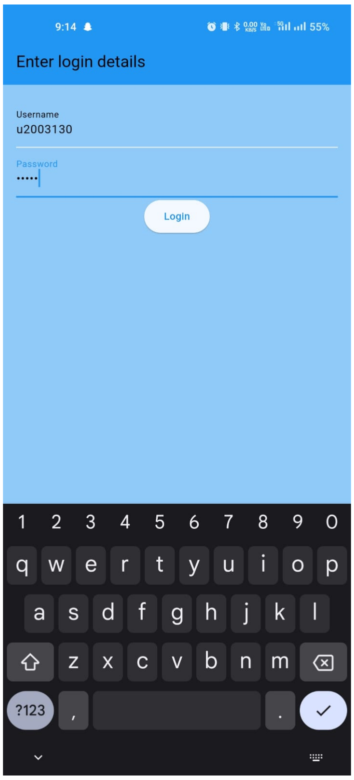
```
# Test functions on a subset of 20 rows
df['cleaned_text'] = df['text'][:20].apply(lambda row: text_preprocessing(row))
df[:20]
```

	text	class	cleaned_text
0	Ex Wife Threatening SuicideRecently I left my ...	suicide	sex wife threaten suicide recently leave wife ...
1	Am I weird I don't get affected by compliments...	non-suicide	weird not affect compliment come know girl fee...
2	Finally 2020 is almost over... So I can never ...	non-suicide	finally hear bad year swear fucking god annoying
3	i need helpjust help me im crying so hard	suicide	need help help cry hard
4	I'm so lostHello, my name is Adam (16) and I'v...	suicide	lose hello adam struggle year afraid past year...
5	Honetly idkI dont know what im even doing here...	suicide	honestly d not know feel like feel unbearably ...
6	[Trigger warning] Excuse for self inflicted bu...	suicide	trigger warn excuse self inflict burn know cri...
7	It ends tonight.I can't do it anymore. \nI quit.	suicide	end tonight not anymore quit
8	Everyone wants to be "edgy" and it's making me...	non-suicide	want edgy make self conscious feel like not st...
9	My life is over at 20 years oldHello all. I am...	suicide	life year old hello year old bald male hairlin...
10	I took the rest of my sleeping pills and my pa...	suicide	take rest sleeping pill painkiller not wait en...
11	Can you imagine getting old? Me neither.Wrinkl...	suicide	imagine get old wrinkle weight gain hair loss ...
12	Do you think getting hit by a train would be p...	suicide	think getting hit train painful gun hard come ...
13	death, continuedI posted here before and saw s...	suicide	death continued post see interesting ask infor...

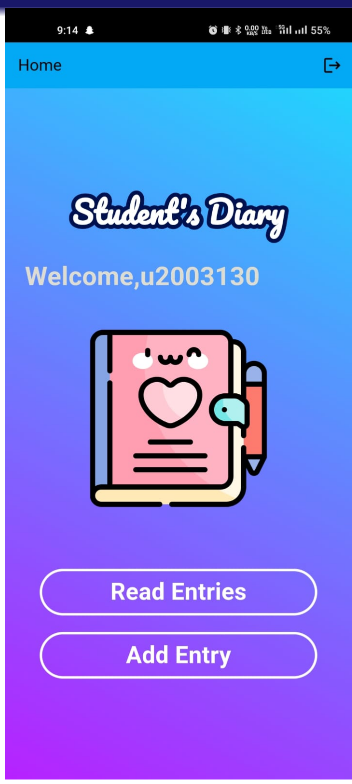
60% Work

- Vectorize using word2vec.
- Model Creation, Training and Testing using CNN.
- Model Creation, Training and Testing using LSTM.
- Prediction Using Electra.

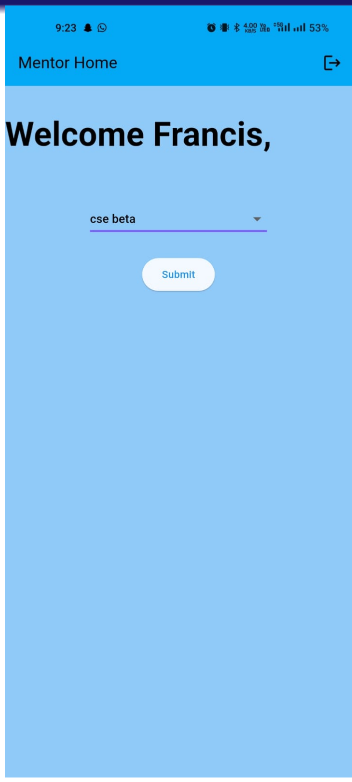
100% Output & Screenshots

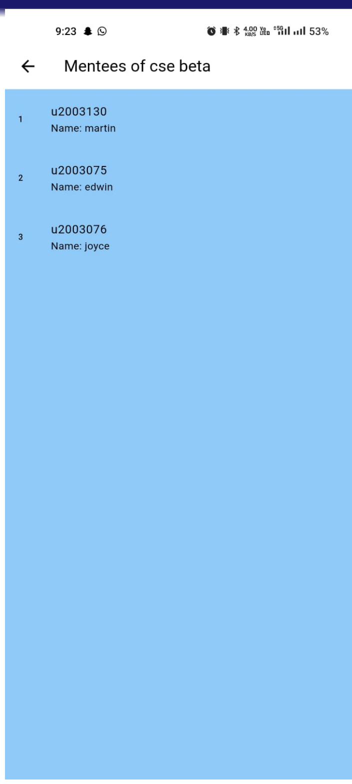


Suicide And Depression Risk Prediction Among College Students



34 / 40


9:23
53%
Mentor Home
Welcome Francis,
cse beta
Submit


9:23
53%
Mentees of cse beta
1
u2003130
Name: martin
2
u2003075
Name: edwin
3
u2003076
Name: joyce

Future Scope

- Professional Collaboration: Extending the Implementation of app in other colleges.
- Integration with Wearable Devices: Integrating the app with wearable devices could enable continuous monitoring of physiological indicators such as heart rate variability, sleep patterns, and activity levels. By analyzing real-time data from wearables, the app could provide early warning signs of escalating distress and prompt users to take proactive steps to manage their mental health.

Conclusion

- The project, dedicated to the early detection of student suicide tendencies using ML and NLP, overcomes a pressing issue with both urgency and innovative solutions.
- Established a safer educational space, utilizing technology as a tool to potentially save lives.

References

- Trotzek, M., Koitka, S. and Friedrich, C.M., 2018. Utilizing neural networks and linguistic metadata for early detection of depression indications in text sequences. *IEEE Transactions on Knowledge and Data Engineering*, 32(3), pp.588-601.
- Chadha, A. and Kaushik, B., 2022. A hybrid deep learning model using grid search and cross-validation for effective classification and prediction of suicidal ideation from social network data. New Generation Computing, 40(4), pp.889-914.
- Abdulsalam, A. and Alhothali, A., 2022. Suicidal ideation detection on social media: A review of machine learning methods. arXiv preprint arXiv:2201.10515
- Galassi, A., Lippi, M., Torroni, P., 2021. Attention in natural language processing.IEEE Trans. Neural Netw. Learn. Syst. 32 (10), 4291–4308.

References

- Saha, S., 2018. A Comprehensive Guide to Convolutional Neural Networks — the ELI5 way. Towards Data Science. (15 December 2018).
- Zhou, C., Sun, C., Liu, Z., Lau, F., 2015. A C-LSTM neural network for text classification. arXiv preprint arXiv:1511.08630
- Coppersmith, G., Leary, R., Crutchley, P. and Fine, A., 2018. Natural language processing of social media as screening for suicide risk. Biomed. Inf. Insights, 10, p. 117822261879286.

Co-Po Mapping

	PO 1	PO 2	PO 3	PO 4	PO 5	PO 6	PO 7	PO 8	PO 9	PO 10	PO 11	PO 12	PSO 1	PSO 2	PSO 3
CO1	2	2	1	1		2	1					3	1		1
CO2	3	3	2	3		2	1					3	1		1
CO3	3	2			3			1		2		3	1		1
CO4	3				2			1		3		3	1		1
CO5	3	3	3	3	2	2		2		3		3	1		1

Figure: Co-Po Mapping

Thank You!

Appendix B: Vision, Mission, Programme Outcomes and Course Outcomes

Vision, Mission, Programme Outcomes and Course Outcomes

Institute Vision

To evolve into a premier technological institution, moulding eminent professionals with creative minds, innovative ideas and sound practical skill, and to shape a future where technology works for the enrichment of mankind.

Institute Mission

To impart state-of-the-art knowledge to individuals in various technological disciplines and to inculcate in them a high degree of social consciousness and human values, thereby enabling them to face the challenges of life with courage and conviction.

Department Vision

To become a centre of excellence in Computer Science and Engineering, moulding professionals catering to the research and professional needs of national and international organizations.

Department Mission

To inspire and nurture students, with up-to-date knowledge in Computer Science and Engineering, ethics, team spirit, leadership abilities, innovation and creativity to come out with solutions meeting societal needs.

Programme Outcomes (PO)

Engineering Graduates will be able to:

1. Engineering Knowledge: Apply the knowledge of mathematics, science, engineering fundamentals, and an engineering specialization to the solution of complex engineering problems.

2. Problem analysis: Identify, formulate, review research literature, and analyze complex engineering problems reaching substantiated conclusions using first principles of mathematics, natural sciences, and engineering sciences.

3. Design/development of solutions: Design solutions for complex engineering problems and design system components or processes that meet the specified needs with appropriate consideration for the public health and safety, and the cultural, societal, and environmental considerations.

4. Conduct investigations of complex problems: Use research-based knowledge including design of experiments, analysis and interpretation of data, and synthesis of the information to provide valid conclusions.

5. Modern Tool Usage: Create, select, and apply appropriate techniques, resources, and modern engineering and IT tools including prediction and modeling to complex engineering activities with an understanding of the limitations.

6. The engineer and society: Apply reasoning informed by the contextual knowledge to assess societal, health, safety, legal, and cultural issues and the consequent responsibilities relevant to the professional engineering practice.

7. Environment and sustainability: Understand the impact of the professional engineering solutions in societal and environmental contexts, and demonstrate the knowledge of, and need for sustainable development.

8. Ethics: Apply ethical principles and commit to professional ethics and responsibilities and norms of the engineering practice.

9. Individual and Team work: Function effectively as an individual, and as a member or leader in teams, and in multidisciplinary settings.

10. Communication: Communicate effectively with the engineering community and with society at large. Be able to comprehend and write effective reports documentation. Make effective presentations, and give and receive clear instructions.

11. Project management and finance: Demonstrate knowledge and understanding of engineering and management principles and apply these to one's own work, as a member and leader in a team. Manage projects in multidisciplinary environments.

12. Life-long learning: Recognize the need for, and have the preparation and ability to engage in independent and lifelong learning in the broadest context of technological change.

Programme Specific Outcomes (PSO)

A graduate of the Computer Science and Engineering Program will demonstrate:

PSO1: Computer Science Specific Skills
The ability to identify, analyze and design solutions for complex engineering problems in multidisciplinary areas by understanding the core principles and concepts of computer science and thereby engage in national grand challenges.

PSO2: Programming and Software Development Skills
The ability to acquire programming efficiency by designing algorithms and applying standard practices in software project development to deliver quality software products meeting the demands of the industry.

PSO3: Professional Skills
The ability to apply the fundamentals of computer science in competitive research and to develop innovative products to meet the societal needs thereby evolving as an eminent researcher and entrepreneur.

Course Outcomes (CO)

Course Outcome 1: Model and solve real world problems by applying knowledge across domains (Cognitive knowledge level: Apply).

Course Outcome 2: Develop products, processes or technologies for sustainable and socially relevant applications (Cognitive knowledge level: Apply).

Course Outcome 3: Function effectively as an individual and as a leader in diverse teams and to comprehend and execute designated tasks (Cognitive knowledge level: Apply).

Course Outcome 4: Plan and execute tasks utilizing available resources within timelines, following ethical and professional norms (Cognitive knowledge level: Apply).

Course Outcome 5: Identify technology/research gaps and propose innovative/creative solutions (Cognitive knowledge level: Analyze).

Course Outcome 6: Organize and communicate technical and scientific findings effectively in written and oral forms (Cognitive knowledge level: Apply).

Appendix C: CO-PO-PSO Mapping

CO-PO AND CO-PSO MAPPING

	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12	PSO1	PSO2	PSO3
CO1	2	2	2	1	2	2	2	1	1	1	1	2	3		
CO2	2	2	2		1	3	3	1	1		1	1		2	
CO3									3	2	2	1			3
CO4					2			3	2	2	3	2			3
CO5	2	3	3	1	2							1	3		
CO6					2			2	2	3	1	1			3

3/2/1: high/medium/low

JUSTIFICATIONS FOR CO-PO MAPPING & CO-PSO MAPPING

MAPPING	LOW/MEDIUM/ HIGH	JUSTIFICATION
100003/ CS722U.1-PO1	M	Knowledge in the area of technology for project development using various tools results in better modeling.
100003/ CS722U.1-PO2	M	Knowledge acquired in the selected area of project development can be used to identify, formulate, review

		research literature, and analyze complex engineering problems reaching substantiated conclusions.
100003/ CS722U.1-PO3	M	Can use the acquired knowledge in designing solutions to complex problems.
100003/ CS722U.1-PO4	M	Can use the acquired knowledge in designing solutions to complex problems.
100003/ CS722U.1-PO5	H	Students are able to interpret, improve and redefine technical aspects for design of experiments, analysis and interpretation of data, and synthesis of the information to provide valid conclusions.
100003/ CS722U.1-PO6	M	Students are able to interpret, improve and redefine technical aspects by applying contextual knowledge to assess societal, health and consequential responsibilities relevant to professional engineering practices.
100003/ CS722U.1-PO7	M	Project development based on societal and environmental context solution identification is the need for sustainable development.
100003/ CS722U.1-PO8	L	Project development should be based on professional ethics and responsibilities.
100003/ CS722U.1-PO9	L	Project development using a systematic approach based on well defined principles will result in teamwork.

100003/ CS722U.1-PO10	M	Project brings technological changes in society.
100003/ CS722U.1-PO11	H	Acquiring knowledge for project development gathers skills in design, analysis, development and implementation of algorithms.
100003/ CS722U.1-PO12	H	Knowledge for project development contributes engineering skills in computing & information gatherings.
100003/ CS722U.2-PO1	H	Knowledge acquired for project development will also include systematic planning, developing, testing and implementation in computer science solutions in various domains.
100003/ CS722U.2-PO2	H	Project design and development using a systematic approach brings knowledge in mathematics and engineering fundamentals.
100003/ CS722U.2-PO3	H	Identifying, formulating and analyzing the project results in a systematic approach.
100003/ CS722U.2-PO5	H	Systematic approach is the tip for solving complex problems in various domains.
100003/ CS722U.2-PO6	H	Systematic approach in the technical and design aspects provide valid conclusions.

100003/ CS722U.2-PO7	H	Systematic approach in the technical and design aspects demonstrate the knowledge of sustainable development.
100003/ CS722U.2-PO8	M	Identification and justification of technical aspects of project development demonstrates the need for sustainable development.
100003/ CS722U.2-PO9	H	Apply professional ethics and responsibilities in engineering practice of development.
100003/ CS722U.2-PO11	H	Systematic approach also includes effective reporting and documentation which gives clear instructions.
100003/ CS722U.2-PO12	M	Project development using a systematic approach based on well defined principles will result in better teamwork.
100003/ CS722U.3-PO9	H	Project development as a team brings the ability to engage in independent and lifelong learning.
100003/ CS722U.3-PO10	H	Identification, formulation and justification in technical aspects will be based on acquiring skills in design and development of algorithms.
100003/ CS722U.3-PO11	H	Identification, formulation and justification in technical aspects provides the betterment of life in various domains.
100003/ CS722U.3-PO12	H	Students are able to interpret, improve and redefine technical aspects with mathematics, science and

		engineering fundamentals for the solutions of complex problems.
100003/ CS722U.4-PO5	H	Students are able to interpret, improve and redefine technical aspects with identification formulation and analysis of complex problems.
100003/ CS722U.4-PO8	H	Students are able to interpret, improve and redefine technical aspects to meet the specified needs with appropriate consideration for the public health and safety, and the cultural, societal, and environmental considerations.
100003/ CS722U.4-PO9	H	Students are able to interpret, improve and redefine technical aspects for design of experiments, analysis and interpretation of data, and synthesis of the information to provide valid conclusions.
100003/ CS722U.4-PO10	H	Create, select, and apply appropriate techniques, resources, and modern engineering and IT tools for better products.
100003/ CS722U.4-PO11	M	Students are able to interpret, improve and redefine technical aspects by applying contextual knowledge to assess societal, health and consequential responsibilities relevant to professional engineering practices.
100003/ CS722U.4-PO12	H	Students are able to interpret, improve and redefine technical aspects for demonstrating the knowledge of, and need for sustainable development.

100003/ CS722U.5-PO1	H	Students are able to interpret, improve and redefine technical aspects, apply ethical principles and commit to professional ethics and responsibilities and norms of the engineering practice.
100003/ CS722U.5-PO2	M	Students are able to interpret, improve and redefine technical aspects, communicate effectively on complex engineering activities with the engineering community and with society at large, such as, being able to comprehend and write effective reports and design documentation, make effective presentations, and give and receive clear instructions.
100003/ CS722U.5-PO3	H	Students are able to interpret, improve and redefine technical aspects to demonstrate knowledge and understanding of the engineering and management principle in multidisciplinary environments.
100003/ CS722U.5-PO4	H	Students are able to interpret, improve and redefine technical aspects, recognize the need for, and have the preparation and ability to engage in independent and life-long learning in the broadest context of technological change.
100003/ CS722U.5-PO5	M	Students are able to interpret, improve and redefine technical aspects in acquiring skills to design, analyze and develop algorithms and implement those using high-level programming languages.
100003/ CS722U.5-PO12	M	Students are able to interpret, improve and redefine technical aspects and contribute their engineering skills in

		computing and information engineering domains like network design and administration, database design and knowledge engineering.
100003/ CS722U.6-PO5	M	Students are able to interpret, improve and redefine technical aspects and develop strong skills in systematic planning, developing, testing, implementing and providing IT solutions for different domains which helps in the betterment of life.
100003/ CS722U.6-PO8	H	Students will be able to associate with a team as an effective team player for the development of technical projects by applying the knowledge of mathematics, science, engineering fundamentals, and an engineering specialization to the solution of complex engineering problems.
100003/ CS722U.6-PO9	H	Students will be able to associate with a team as an effective team player to Identify, formulate, review research literature, and analyze complex engineering problems
100003/ CS722U.6-PO10	M	Students will be able to associate with a team as an effective team player for designing solutions to complex engineering problems and design system components.
100003/ CS722U.6-PO11	M	Students will be able to associate with a team as an effective team player, use research-based knowledge and research methods including design of experiments, analysis and interpretation of data.

100003/ CS722U.6-PO12	H	Students will be able to associate with a team as an effective team player, applying ethical principles and commit to professional ethics and responsibilities and norms of the engineering practice.
100003/ CS722U.1-PSO1	H	Students are able to develop Computer Science Specific Skills by modeling and solving problems.
100003/ CS722U.2-PSO2	M	Developing products, processes or technologies for sustainable and socially relevant applications can promote Programming and Software Development Skills.
100003/ CS722U.3-PSO3	H	Working in a team can result in the effective development of Professional Skills.
100003/ CS722U.4-PSO3	H	Planning and scheduling can result in the effective development of Professional Skills.
100003/ CS722U.5-PSO1	H	Students are able to develop Computer Science Specific Skills by creating innovative solutions to problems.
100003/ CS722U.6-PSO3	H	Organizing and communicating technical and scientific findings can help in the effective development of Professional Skills.