

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

- † Problem Definition
- † Purpose and Needs
- † Project Objective
- † Literature Survey
- † Existing Methods
- † Proposed Method
- † Architecture diagram
- † Sequence diagram
- † Module division
 - † Text Preprocessing
 - † Training
 - † Model Building and Evaluation
 - † Application Interface
 - † Module wise diagram

Contents

- † Assumptions
- † Work breakdown and responsibilities
- † Software requirements
- † Hardware requirements
- † Module wise scheduling
- † Gantt Chart
- † Budget
- † Risks and challenges
- † 30% OUTPUT & SCREENSHOTS
- † Conclusion
- † References
- † Co-Po mapping

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.

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.

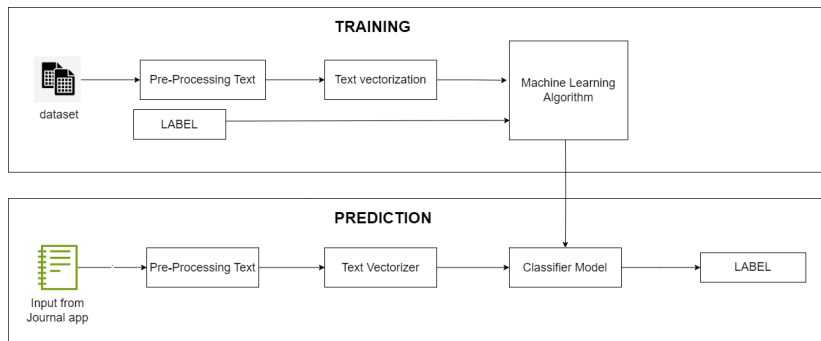
Existing Methods

- † 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.

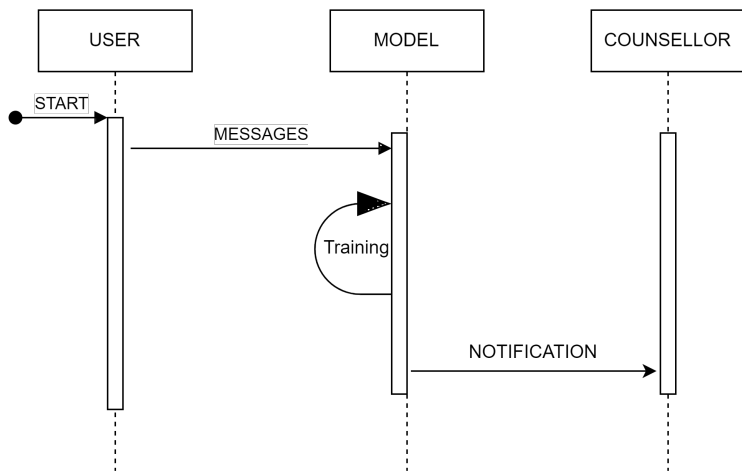
Proposed Method

- † 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

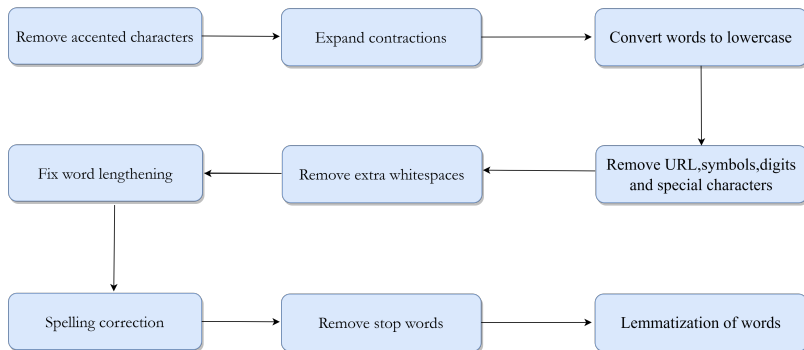


Figure: Data Preprocessing Steps

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

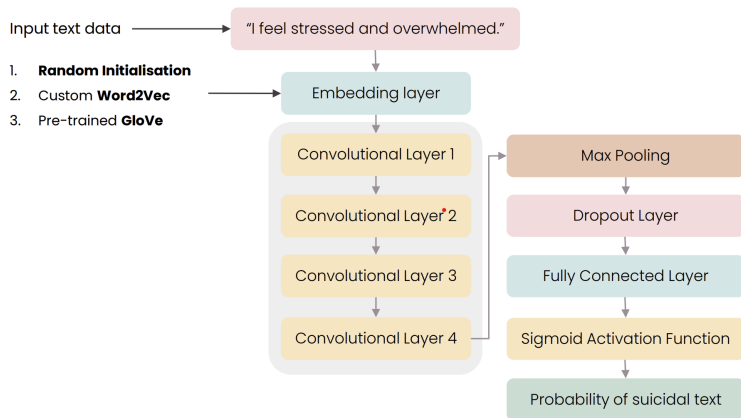


Figure: Working of CNN

LSTM

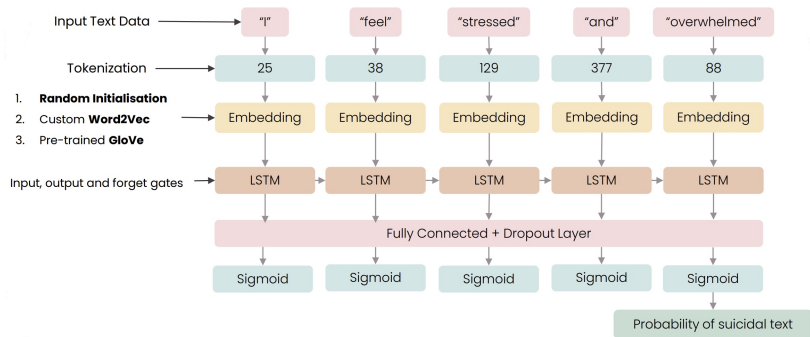


Figure: Working of LSTM

BERT

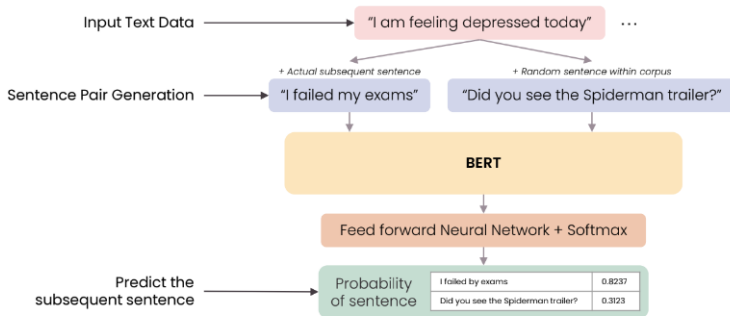


Figure: Working of BERT

ELECTRA

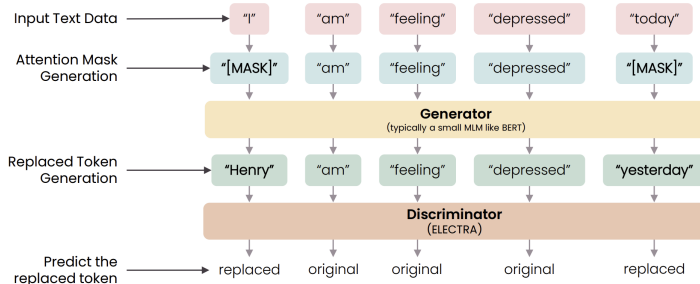


Figure: Working of ELECTRA

Model Building & Evaluation

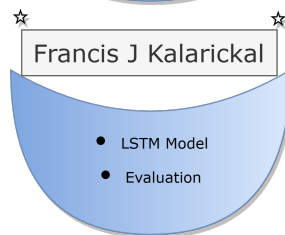
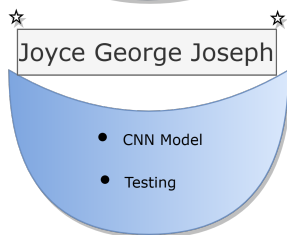
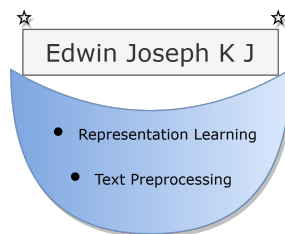
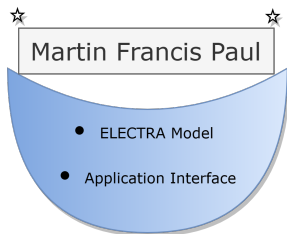
- † We calculate precision, recall, accuracy of each models.
- † This is used to evaluate the F1 score.
- † The best model is selected based on the value of F1 score.

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

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.

Work Breakdown and Responsibilities



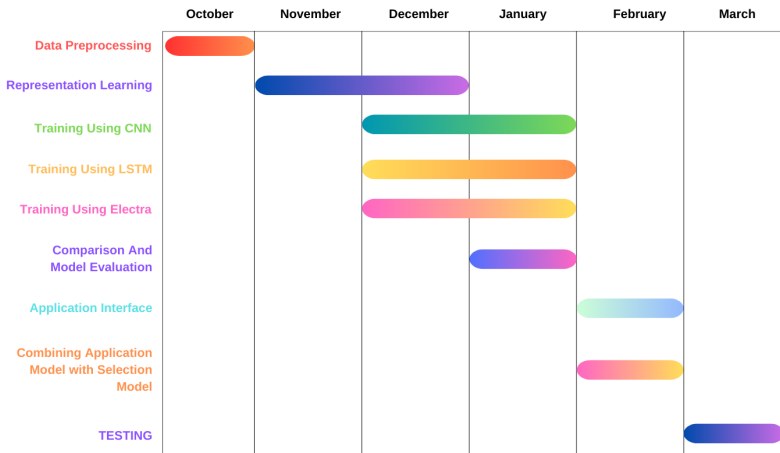
Software Requirements

- † 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



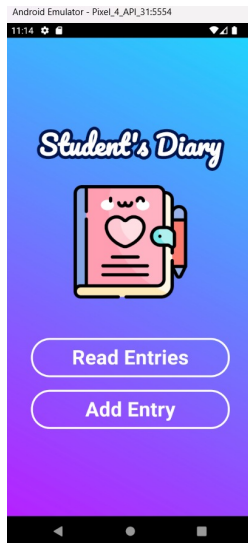
Risks and challenges

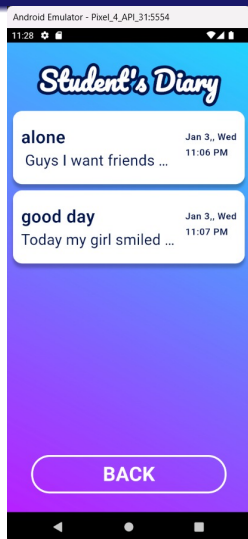
- † Reluctance or idleness of students to write in the virtual journal
- † Ensuring that the virtual journal is highly secure and that the entries are confidential.
- † Students may be hesitant to record their thoughts and feelings if they fear unauthorized access.
- † Data Scarcity

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





Android Emulator - Pixel_4_API_31:5554

11:16

Student's Diary

Entry Title*

Create new Entry

SAVE



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

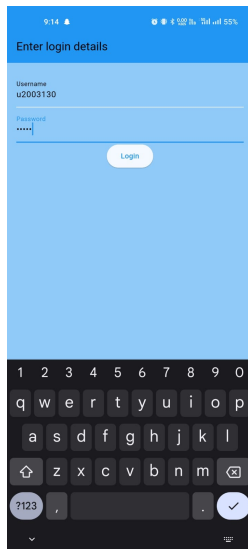


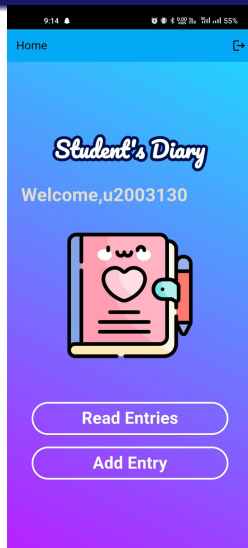
	text	class	cleaned_text
0	Ex Wife Threatening SuicideRecently I left my ...	suicide	sex wife threaten suicide recently leave wife ...
1	Am I weird I don't get affected by compliments...	non-suicide	weird not affect compliment come know girl fee...
2	Finally 2020 is almost over... So I can never ...	non-suicide	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've...	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





9:23 5G 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
C01	2	2	1	1		2	1					3	1		1
C02	3	3	2	3		2	1					3	1		1
C03	3	2			3			1		2		3	1		1
C04	3				2			1		3		3	1		1
C05	3	3	3	3	2	2		2		3		3	1		1

Figure: Co-Po Mapping

Thank You!