Can the suicidal intentions of young people be predicted using the language analysis of Reddit posts?

I. Restatement and Summary

1. Restatement of the question

The question "Can the suicidal intentions of young people be predicted using the language analysis of Reddit posts?" addresses the potential of using language analysis, specifically Natural Language Processing (NLP), to predict the suicidal intentions of young individuals based on their posts on the online platform Reddit.

Feasibility and significance: Language analysis through NLP has demonstrated potential in extracting insights from text data. The prevalence of online platforms like Reddit provides a vast amount of textual content that could be mined for signals related to mental health and well-being. If successful, this approach could offer an early detection and intervention, potentially saving lives.

2. Summary of the model

Table 1 analyses challenges when answering the question and explains how the RNN model copes with the solutions. The modelling exercises suggest that predicting the suicidal intentions of young people is possible using language analysis of Reddit posts.

The RNN model was selected over other models because of its intrinsic model qualities and its highest recall score. The chosen hyperparameters also strike a balance between model complexity, regularization, and generalization and lead to a good result for the classification task.

Challenges	Solutions
Context and Nuances: Language on platforms like Reddit can be heavily nuanced, sarcastic, or ironic. Distinguishing between genuine concerns and casual remarks requires a deep understanding of context, which can be difficult for automated systems.	The RNN model can effectively capture the sequential nature of textual data and interpret the sequence of words in a sentence, capturing the context and nuances that might indicate suicidal intentions and understanding the meaning of words within sentences.
False Positives and Negatives: NLP models can produce both false positives (predicting suicidal intentions where there are none) and false negatives (missing actual instances). Striking a balance between sensitivity and specificity is crucial to avoid unnecessary distress or missed cases.	RNN not only exhibits the highest recall score but also maintains a balanced trade-off between precision and recall, so it can correctly identify individuals at risk of suicide while minimizing false positive predictions.
Data Quality and Bias: The quality of the training data and potential biases in the dataset can impact model performance. If the training data is skewed towards certain linguistic patterns or demographics, the model might not generalize well.	A single hidden layer with 64 neurons using the ReLU activation function of the RNN model provides sufficient capacity for learning while preventing excessive complexity. The sigmoid activation in the output layer suits binary classification tasks. A dropout rate of 0.5 mitigates overfitting by randomly deactivating neurons during training. Training for 100 epochs and early stopping with a batch size of 32 provides adequate learning without overwhelming computational resources.

Table 1: Question challenges and solutions description

II. Analysis and Visualisation

1. Feature Importance

We use the RNN model's predicted outcomes on the test data to display important words (features) of each class. Word Clouds in Figure 1 identify words that RNN considers important for classifying samples. The size of each word in the cloud represents its frequency in the data. Figure 2 represents the correlation value of the top words associated with each class to see which words are more strongly correlated with each class, indicating their importance in the classification task.

Figure 1: Word Clouds of the most important words for non-suicidal and suicidal class

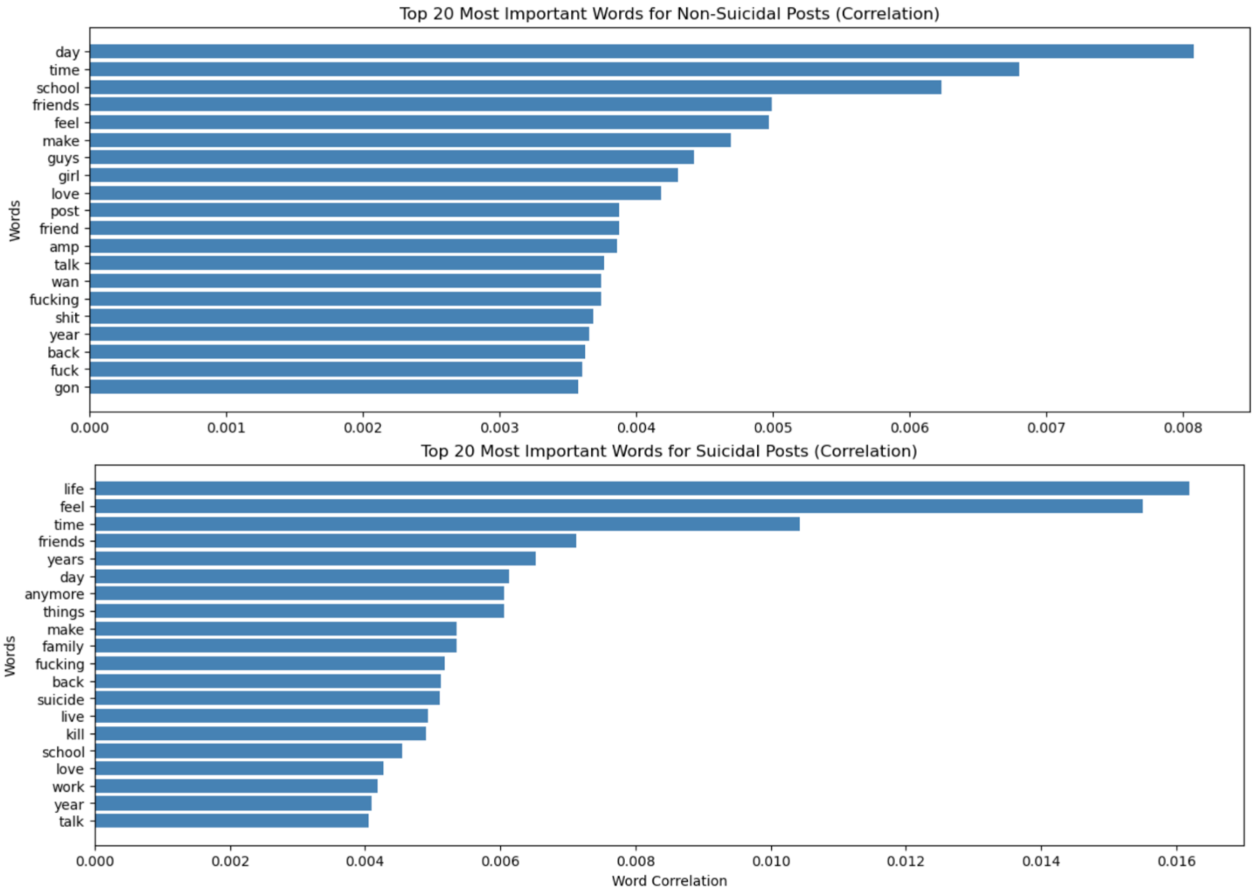

Figure 2: Bar charts of top 20 most important words for each class according to correlation

2. Confusion Matrix

Figure 3 visualizes the true negative, false positive, false negative and true positive predictions to see how well the RNN model is classifying each class. RNN is performing fairly well when classified correctly 91% of instances in “non-suicidal” class and 92% of instances in “suicidal” class. Only 8% of the “suicidal” and 9% of “non-suicidal” were wrongly classified.

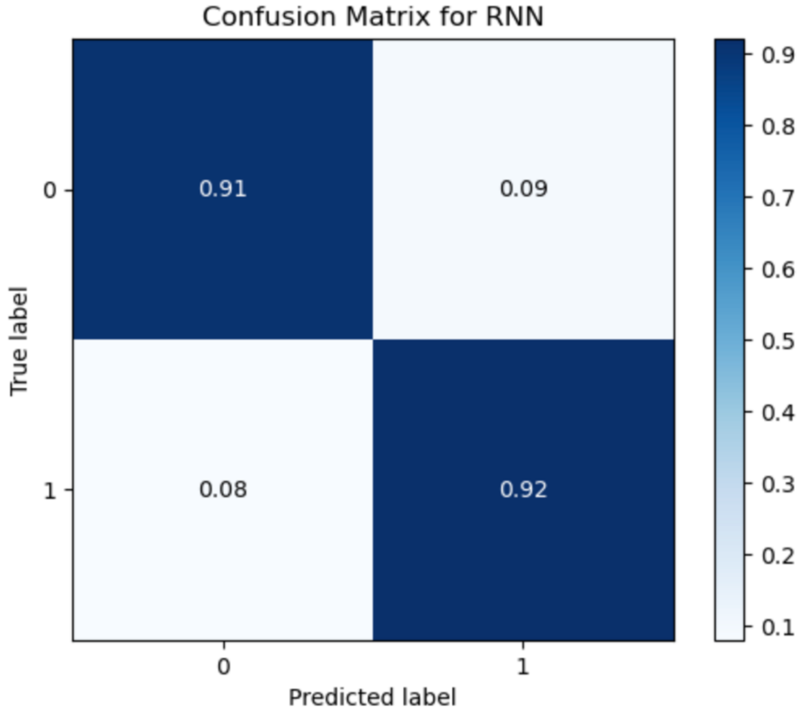

Figure 3: Confusion matrix with the normalized values of the RNN model

3. Learning Curves and Recall Scores Distribution

Figure 4 summarizes the training and testing progress of the RNN model with 5-fold cross-validation and Figure 5 displays the recall scores' distribution.

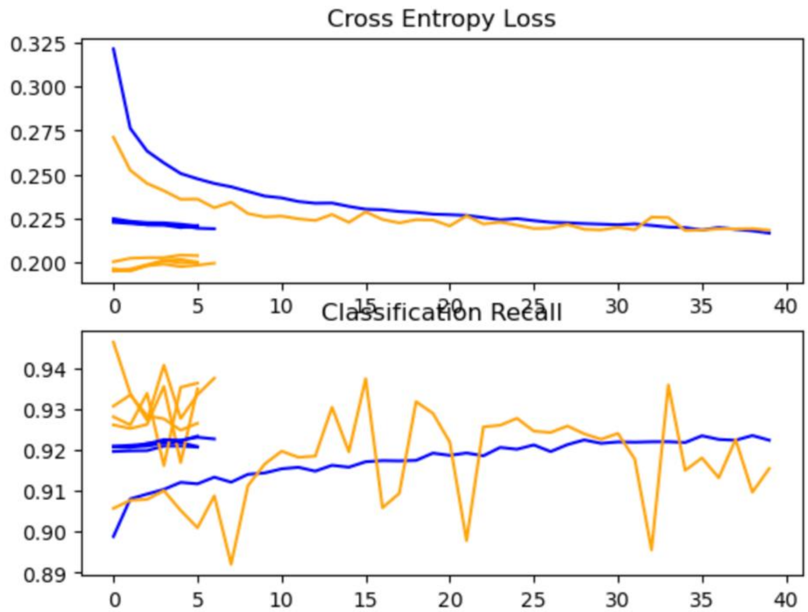

Figure 4: Cross Entropy Loss and Classification Recall of the training dataset (blue lines) and the test dataset (orange lines)

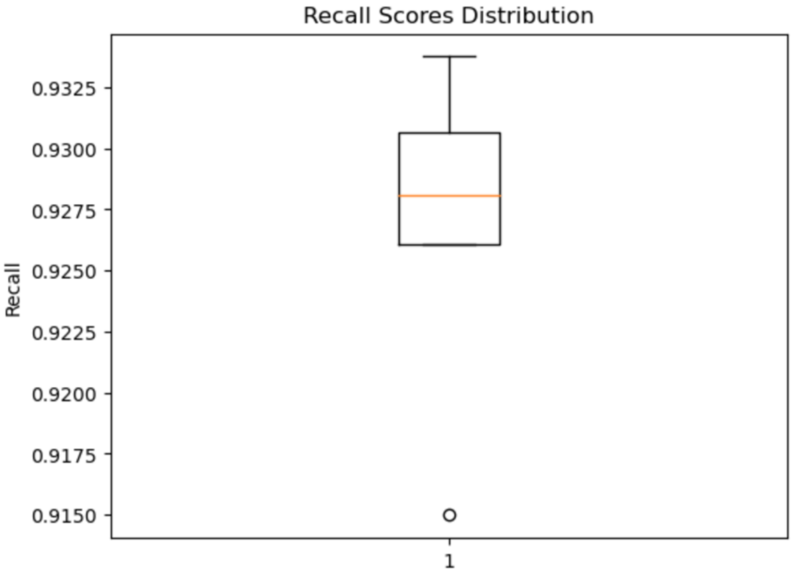

Figure 5: Recall Scores Distribution

Overall, it can be seen from Table 2 that the RNN model achieves high recall scores. The training and test losses (0.196 and 0.213) are relatively low, suggesting that the model is learning effectively and generalizing well to new data. The consistency between the training and test recall scores indicates that the model is not overfitting. The standard deviation with a low value 0.006 and a narrow distance between the upper quartile and lower quartile also indicate that the recall values are mostly clustered within this range and the model's performance is consistent across different runs.

Cross-Validation Recall	92.67%
Training recall	0.9313089203166961 7
Test recall	0.923445569896698
Training loss	0.1961984485387802
Test loss	0.21612417697906494
Recall: mean=0.927 std=0.006, n=5 Min: 0.9150 Max: 0.9350 Lower quartile: 0.9265 Upper quartile: 0.9300	

Table 2: A summary of recall scores above different sets and its distribution

4. Precision-Recall Curve

Figure 6 provides insights into the trade-off between precision and recall. The curve for class 0 reaches 0.883 and the curve class 1 is around 0.878, which underscores the model's ability to achieve a trade-off between precision and recall for both class 0 and class 1 predictions. The micro-average PR curve of 0.881 further indicates a solid overall performance of the model across both classes, so the model is achieving good trade-offs between precision and recall for the entire dataset.

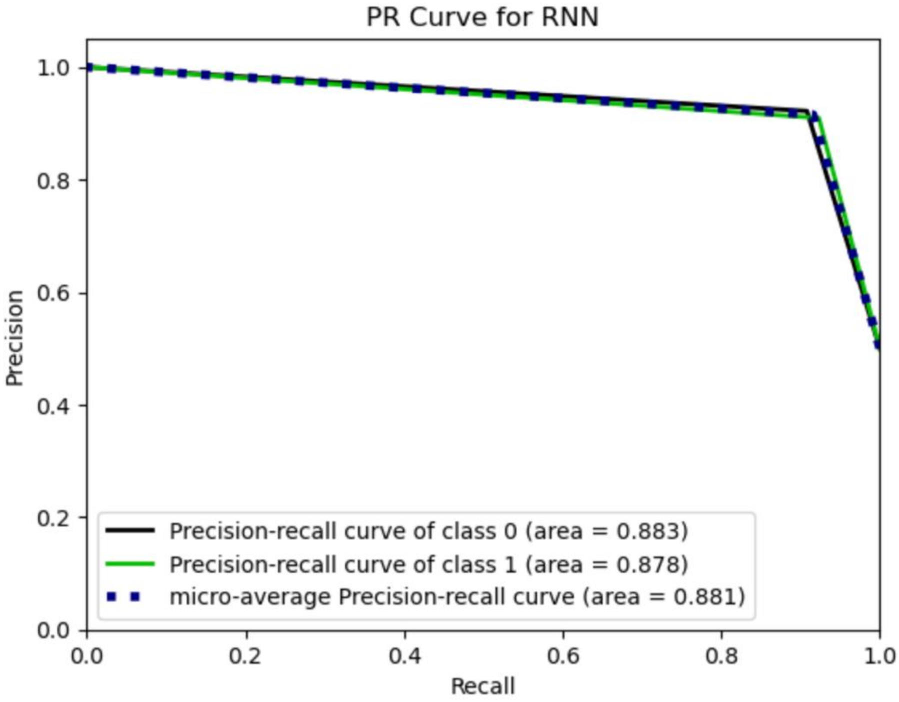

Figure 6: Precision-Recall Curve for the RNN model

5. Sample Predictions

Figure 7 visualizes how well the model's predictions align with the true labels for a few randomly chosen samples from the test dataset.

```
1/1 [==============================] - 0s 13ms/step
Sample Text: making money years make money buy atv guys ideas make mo
ney made money teenager
True Label: 0
Predicted Label: 0
Predicted Score: 0.008550045
------------------------------
Sample Text: happy years ago full illnesses cope anymore draining tir
ing life deal easy
True Label: 1
Predicted Label: 1
Predicted Score: 0.9857928
------------------------------
Sample Text: bid thee farewelli appreciate support ultimately vein li
fe method note written blame decision sleepless nights easy ultimatel
y reality things grateful greats goodbye
True Label: 1
Predicted Label: 1
Predicted Score: 0.9710689
------------------------------
Sample Text: poem wrote suicidal influences open meanings interest po
etry fact lowest forms literature expression events recent past lead
writing poem written free counter productive suicidal individuals del
ete post argument call hear cry wonder attempt wave goodbye put barre
l temple begin wish overdose close close open eyes mesmerized million
lies surface top scream stop open eyes meant goodbyes stand high sky
finally finally goodbye falling flailing failing close close open eye
s lying bed dead join rest cattle hell internal battle paint smile be
gin daily mile time bed beautiful lie rings head
True Label: 1
Predicted Label: 1
Predicted Score: 0.9486991
```

Figure 7: The model's predictions on samples of the test dataset

III. Improvement of Situation

Data preprocessing: Convert all text to lowercase, removes non-alphabetic characters, and removes stopwords and use FastText.	Some stopwords may carry meaning in certain contexts and should not be removed. FastText is capable of understanding context based on n-grams, so removing stopwords might not be necessary and could impact FastText's ability to understand the meaning of words.
Model evaluation metrics: Use recall as the primary evaluation metric.	A high recall might lead to a lower precision. Optimizing solely for recall might lead to overfitting or sacrificing overall model performance.
Model complexity: Use the model architecture having only one hidden layer with 64 units.	Simple models may struggle to capture complex patterns and relationships within the data. They might not have enough capacity to learn intricate features and nuances, leading to not generalizing well to unseen data.
Hyperparameter tuning: Manually choose the hyperparameters without using hyperparameter tuning techniques.	Hyperparameter tuning can help identify better combinations of settings. Neglecting this step might miss opportunities to significantly improve the model's accuracy and generalization ability.

Table 3: Current methods employed by the RNN model and their respective drawbacks

After considering the limitations from Table 3, we propose the following improvements:

1. Data Preprocessing

- We experiment using FastText without removing stopwords to observe how it handles context and word meanings more effectively.
- We compare results between using Lemmatization/Stemming and not using them to see if they affect the performance of the FastText model.

2. Model Evaluation Metrics

We still use Recall() as the main metric during training and validation to optimize recall on the validation set, but then use visualizations like the Precision-Recall curve to comprehensively evaluate the model's performance on the test set.

3. Model Complexity and Hyperparameter Tuning

Table 4 explains the reasons why a single layer model still gets a good result according to a careful data pre-processing and model implementation process. However, achieving a good result with only one hidden layer might be because of luck. Therefore, it is crucial to experiment with deeper neural networks, different activation functions, more hidden layers, or even more complex architectures like LSTM or transformers to find the best one. Additionally, we implement the hyperparameter tuning technique like grid search to see if it can improve the model's performance.

Data Quality	The preprocessing steps contribute to improving data quality. When the data is clean, and contains informative features, a simple model might be sufficient to capture the underlying patterns and relationships in the data.
Feature Representation	FastText represents text features and captures semantic meanings. The dimensionality of the data is 300, corresponding to the size of the word embeddings of FastText model. With the input data has relatively low dimensionality, a simple model might be sufficient to capture the available information.
Task Specificity	The dataset consists of 226,953 rows, with 113,534 labelled as class 1 and 113,419 labelled as class 0. The class boundaries are relatively well-defined, so a simple model can effectively differentiate the classes and using a more complex model might not be necessary.
Hyperparameter Tuning & Overfitting Managing	An appropriate choice of the hyperparameters impacts the model performance. "Early stopping" finds the point where the model starts to overfit and reverts to the weights with the best validation performance. Validation performance monitoring enables early detection of overfitting trends and helps in making informed decisions about stopping training or adjusting hyperparameters.
Luck	Sometimes, the initial random weights and biases of the model can result in unexpectedly good results.

Table 4: Reasons for achieving good results with only a single layer model

IV. Conclusion and Future Work

1. Conclusion

Detecting suicide tendencies poses considerable challenges due to its complexity. This project aimed to identify suicidal content on Reddit. Evaluating distinct and combined feature sets, we assessed suicide signs via diverse text classification methods. Reddit's timeliness, emotional expression, and medical compatibility make it an ideal complement to traditional health systems. The Recursive Neural Network model showcased effectiveness with 92% recall and 92% precision. However, more research is essential in this field, involving the integration of sentiment analysis and the expansion of datasets encompassing diverse online platforms and blogs.

2. Future Work

Table 5 explains how the outcomes of analyzing posts to determine suicidal or non-suicidal intentions can be applied in various real-world domains such as mental health care, cybersecurity, and community well-being.

Domains	Outcomes application
Mental Health Care	Support mental health professionals in assessing the suicide risk of patients (When an unusual online post or interaction is detected, the system can alert healthcare experts to engage with individuals at risk and provide appropriate support)
Cybersecurity	Detect online interactions related to suicidal intentions (Security systems and online content management platforms can implement additional measures to prevent harmful content and provide supportive resources for those at risk)
Community Well-being	Monitor online forums or social networks to identify signs of suicide risk (Organizations and communities can provide support materials, direct contact, or even interventions when necessary to safeguard the mental well-being of their members)
Global Mental Health Analysis	Generate statistical reports to provide insights into the global mental health landscape, identifying trends and emerging mental health
Related Mental Health Research and Development	Provide valuable data for researchers in psychology and social sciences to delve deeper into suicide risk factors and develop effective intervention strategies

Table 5: Application of suicide classification task in different domain

Reference list:

1. Arowosegbe, A & Oyelade, T 2023, 'Application of Natural Language Processing (NLP) in Detecting and Preventing Suicide Ideation: A Systematic Review', International Journal of Environmental Research and Public Health, vol. 20, no. 2, p. 1514–.
2. Bejan, CA, Ripperger, M, Wilimitis, D, Ahmed, R, Kang, J, Robinson, K, … Walsh, CG 2022, 'Improving ascertainment of suicidal ideation and suicide attempt with natural language processing', Scientific Reports, vol. 12, no. 1, pp. 15146–15146.
3. Goyal, A 2021, *Reddit dataset for Multi-task NLP*, kaggle, 21 June, viewed 12 June 2023, https://www.kaggle.com/datasets/amangoyl/reddit-dataset-for-multi-task-nlp?select=Dataset_Suicidal_Sentiment.csv.
4. Jain, P, Ram Srinivas, K & Vichare, A 2022, 'Depression and Suicide Analysis Using Machine Learning and NLP', Journal of Physics. Conference Series, vol. 2161, no. 1, p. 12034–.
5. Komati, N 2021, *Suicide and Depression Detection*, kaggle, 19 May, viewed 12 June 2023, https://www.kaggle.com/datasets/nikhileswarkomati/suicide-watch.
6. Levis, M, Leonard Westgate, C, Gui, J, Watts, BV & Shiner, B 2021, 'Natural language processing of clinical mental health notes may add predictive value to existing suicide risk models', Psychological Medicine, vol. 51, no. 8, pp. 1382–1391.
7. Yadhu, A 2022, *Predicting Suicide and Word Analysis*, kaggle, 2 July, viewed 4 July 2023, https://www.kaggle.com/code/yadhua/predicting-suicide-and-word-analysis.