International Journal on Perceptive and Cognitive Computing (IJPCC) Vol 10, Issue 2 (2024)
https://doi.org/10.31436/ijpcc.v10i2.468

Suicide Risk Prediction Using Artificial Intelligence

Elean Sugafta Rafa, Adeeba Mahmooda, Takumi Sase*
Dept. of Computer Science, KICT, International Islamic University Malaysia, 53100 Kuala Lumpur, Malaysia.

*Corresponding author: takumi@iium.edu.my
(Received: 8th May 2024; Accepted: 25th June 2024; Published on-line: 30th July 2024)

Abstract— Over the past decade, social media has been attracting a growing number of people to the online space. Due to the increase in internet usage, a huge number of text data has been produced. Such data can reflect users' mental health status, but it is still challenging to predict suicide risk from data, due to the high complexity of texts. This research aims to predict the suicide risk from Reddit posts using artificial intelligence (AI). The data were collected from the Kaggle dataset, which included postings of suicide subreddits. The data were pre-processed through natural language processing techniques. Logistic regression, naive Bayes, and random forest models were then used for classifying the Reddit users, i.e., to predict if they are in a suicidal or non-suicidal mental state. These models were compared to identify an AI approach that provides the best performance among the three models. Then, the logistic regression model with doc2vec showed the highest precision of 0.92, recall 0.92, and F_1 score of 0.92.

Keywords— Suicide Risk, Artificial Intelligence, Machine Learning, Reddit

I. Introduction

The sad reality is that a lot of individuals in contemporary society are so stressed or hurt that they think of taking their own life. According to the reports, 703 000 individuals attempted suicide annually [1]. Even more alarming are the findings of a recent study conducted on 32 children's hospitals in the United States [2]. The study revealed a steady rise in the incidence of serious self-harm and suicide among children and adolescents between 2008 and 2015. Among teenagers, there is a noticeable tendency to discuss suicide pacts, seek methodological guidance, and post suicidal thoughts in online groups on social media [3].

The Reddit is a social media platform where users can post content, ask questions, and receive answers [4]. With more than 330 million active users worldwide, Reddit is expanding at twice the rate of Twitter. Subreddits are groups on Reddit where content is categorised. One subreddit, called 'SuicideWatch', was founded in 2008 with approximately 214000 users. Members can use this platform to share their suicidal thoughts or leave encouraging remarks.

Since social media data as in Reddit can reflect users' mental health, many studies have attempted to detect suicide risk from the postings using artificial intelligence (AI) [4-6]. For example, a study analysed the data posted on 'SuicideWatch' to identify patterns in individuals, ranging from the stage of suicidal thoughts to the stage of suicide attempts, using a deep learning approach [4]. However, it is still challenging to perform suicide risk prediction due to the high complexity of text data.

The objectives of this research are to construct the models that can classify social media users as suicidal or non-suicidal through training, to evaluate the performance of the models through testing, and to suggest an AI algorithm that can be used for suicide risk prediction. The algorithm will consist of natural language processing (NLP), training, and classification.

This significance of this research is to provide an AI algorithm that can help the prediction of suicide risk based on Reddit. The algorithm comprises an NLP technique and a machine learning model. A good combination of an NLP technique and a machine learning model was determined through comparison of the three models: logistic regression, naive Bayes, and random forest. The doc2vec was used as NLP preprocessing for the logistic regression. For the naive Bayes model, bag-of-words (BOW) and term frequency–inverse document frequency (TF-IDF) were used as NLP preprocessing and compared. The data were collected from the two subreddits: 'SuicideWatch' and 'depression'.

The rest of this paper is organised as Literature Review, Materials and Methods, Results, and Conclusions. The Literature Review will show recent research related to suicide risk prediction using AI. The Materials and Methods section consists of four parts: Data Collection, NLP Preprocessing, Classification Models, and Model Evaluation. The Results section displays the performance of each model, and this research will be summarised in Conclusions.

II. Literature Review

A methodology for building a suicidal ideation detection was proposed based on social media using deep learning

and machine learning models [5]. They used a model consisting of convolutional neural network (CNN) and bidirectional long short-term memory (LSTM), in addition to the XGBoost model, to classify social posts as suicidal or non-suicidal. The data were collected from the subreddit 'SuicideWatch', and TF-IDF and word2vec were used for text representation. The accuracy, precision, recall, and F_1 score of the CNN-bidirectional LSTM model were 0.95, 0.943, 0.949, and 0.95, respectively. This model outperformed the XGBoost model whose accuracy was 0.915. The study suggests that the proposed method may help identify individuals who require medical treatment [5].

Another research extracted several informative sets of features, to detect suicidal ideation [6]. The suicidal ideation texts were collected from a subreddit 'SuicideWatch'. On the other hand, the texts without suicidal content were collected from other popular subreddits. The study compared six classifiers: support vector machine (SVM), random forest, gradient boost classification tree, XGBoost, multilayer feed forward neural network, and LSTM. Among the six models, the XGBoost using TF-IDF showed the best performance; the accuracy, precision, recall, and F_1 score were 0.9571, 0.9499, 0.9668, and 0.9583, respectively [6].

Twitter was also used for detecting suicide risk [7, 13]. A study collected 14701 suicide-related tweets and compared the two machine learning models: SVM and logistic regression [7]. The results showed that the SVM with TF-IDF and without word removal can be the best performing algorithm. The study suggests that it is possible to distinguish the level of suicidality using machine learning [7]. Another study generated an algorithm for predicting future risk to suicidal ideation [13]. The study constructed neural networks to infer psychological weights. The area under the curve (AUC) of this model was 0.68, and this value was significantly higher than 0.63 that was the AUC of SVMs. The study further used random forest models to predict suicidal ideation, achieving an AUC of 0.88 [13].

A review has been done through 296 studies mainly based on the following countries: USA (47%), Korea (24%), and Canada (18%), regarding AI and suicide prevention [11]. Among the 296 studies, the PRISMA criteria identified 17 studies that were published between the years of 2014 and 2020. Four prospective designs and thirteen retrospective designs were presented in the research using the sample sizes ranging from 182 to almost 19 million. The AI models that were used in these studies are SVM (12%), gradient-boosting algorithms (18%), random forest (35%), logistic regression (53%), and LASSO (12%) [11]. The AUC was ranging between 0.604 to 0.947 in predicting suicide risk. The paper highlighted the potential applications of AI in assessing suicide risk.

A recent study demonstrated the effectiveness of explainable AI for suicide risk prediction [12]. A dataset was selected according to the three criteria and collected [12]. They used random forest, decision tree, logistic regression, SVM, perceptron, and XGBoost. The random forest shows the best performance among the six models; AUC was more than 0.97.

Overall, the choice of a classification model that enables suicide risk prediction can depend on dataset analysed. In this research, we focused on the data posted on Reddit, especially on the two subreddits: 'SuicideWatch' and 'depression'. As shown below, we compared the three models: logistic regression, naive Bayes, and random forest.

III. Materials and Methods

Figure 1 presents the flow of this research. Data Collection, NLP Preprocessing, Classification Models, and Model Evaluation are presented in turn. All the analyses were conducted using Python code.

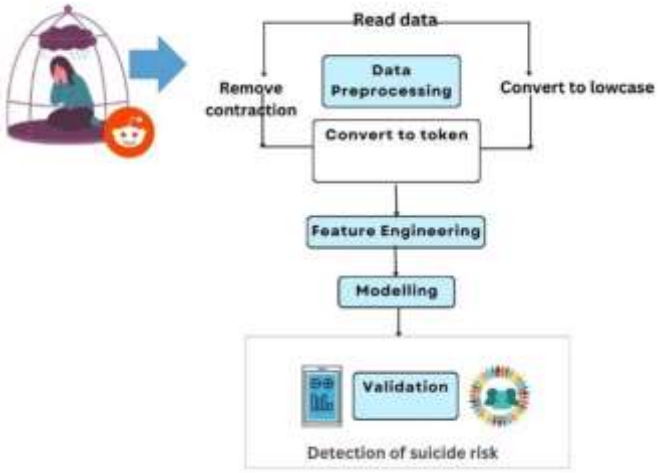

Fig. 1 Methodology Flowchart

A. Data Collection

A dataset of suicide detection datasets was collected from the Kaggle dataset [14], which contains posts of the two Reddit platforms: 'SuicideWatch' and 'depression' subreddits. Postings on 'SuicideWatch' from December 16, 2008, to January 2, 2021, were all gathered; while postings on 'depression' were gathered between January 1, 2009, and January 2, 2021 [14]. We used 116037 posts from each category (suicidal, non-suicidal).

Figures 2 and 3 highlight words that occur most frequently in texts. A word's frequency in the text increases

https://doi.org/10.31436/ijpcc.v10i2.468

with its size in the visual. The dataset contains same number of suicidal and non-suicidal posts (Fig. 4).

Fig. 2 Mostly used words in suicide posts

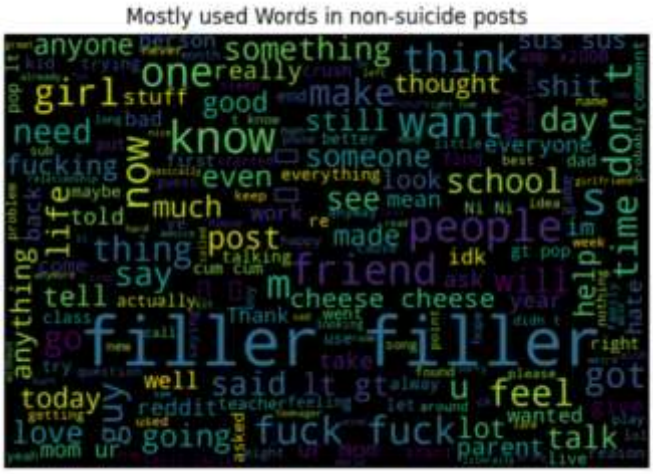

Fig. 3 Mostly used words in non-suicide posts

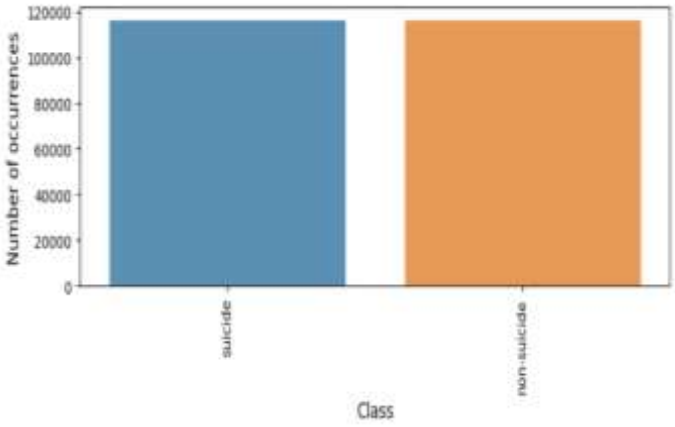

Fig. 4 Suicidal & Non-Suicidal post count

B. NLP Preprocessing

Figure 5 shows data preprocessing. After reading the data, the texts were converted to lowercase. Then, the contractions were removed and converted into tokens. We built a dictionary and prepared the data for vocabulary. These data were labelled for the prerequisite of doc2vec model.

Fig. 5 Data preprocessing steps

1) Doc2vec: The doc2vec was used to generate vectors for documents, as preprocessing for the logistic regression model. The documents were converted to a vector as the following algorithm:

$$Post_i = TaggedDocument(words = [word_1, word_2, \ldots., word_n] \quad (1)$$

$$Tags = [PostID_i] \quad (2)$$

$$Models_{Doc2Vec} = Doc2Vec(Documents = [Post_1, Post_2, \ldots., Post_m]) \quad (3)$$

2) BOW (Bag-of-Words): BOW was created as preprocessing for the naive Bayes model. The BOW is a vector of zeros and ones such as

$$BOW_{Posti} = [0,1,0,1,\ldots,0] \quad (4)$$

3) TF-IDF (Term Frequency–Inverse Document Frequency): The TF-IDF was computed as preprocessing for the naive Bayes model.

$$TFIDF_{Posti} = TF_{Posti} \times IDF_{term} \quad (5)$$

C. Classification Models

The three models: logistic regression, naïve Bayes, and random forest models were used for the classification, and were compared based on the classification performance of each model.

https://doi.org/10.31436/ijpcc.v10i2.468

1) *Logistic Regression*: The logistic regression model was used to classify the data after doc2Vec. This can be represented as

$$Probability\ of\ Suicide = sigmoid(W \cdot Doc2Vec + b) \quad (6)$$

2) *Naive Bayes:* The naive Bayes model was applied to the BOW.

$$Probability\ of\ Suicide = \frac{P(BOW|Suicide) \times P(Suicide)}{P(BOW)} \quad (7)$$

3) *Random Forest:* For the random forest model, none of doc2vec, BOW, and TF-IDF were applied.

D. Model Evaluation

Seventy percent of the dataset was used for training, and the rest was for testing. The confusion matrix was then computed to evaluate the above models. The accuracy, precision, recall, and F_1 score were calculated as follows.

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (8)$$

$$precision = \frac{TP}{TP + FP} \quad (9)$$

$$recall = \frac{TP}{TP + FN} \quad (10)$$

$$F_1 = 2 \times \frac{precision \times recall}{precision + recall} \quad (11)$$

The Cohen's kappa was used to evaluate agreement between two evaluators [8]. The kappa was computed from the confusion matrix and used to estimate the levels of agreement, as shown in Table 1.

TABLE I
KAPPA VALUE INTERPRETATION TABLE

Kappa values	Interpretation
< 0	Lack of agreement
0 - 0.20	Minimum agreement
0.21 - 0.40	Reasonable agreement
0.41 - 0.60	Moderate agreement
0.61 - 0.80	Substantial agreement
0.81 - 1	Near-perfect agreement

IV. RESULTS

Figure 6 shows the comparison of accuracy of the three models: logistic regression, naive Bayes, and random forest. For the logistic regression model, doc2vec was applied to the dataset as NLP preprocessing. For the naive Bayes model, the two NLP techniques, i.e. BOW and TF-IDF, were compared. Then, the logistic regression model with doc2vec achieved the highest accuracy of 0.93. Accuracy of the naive Bayes model with BOW and naive Bayes model with TF-IDF were 0.89 and 0.90, respectively. The accuracy of the random forest model was 0.83.

Figures 7, 8, and 9 present the receiver operating characteristic (ROC) curves of the logistic regression model, naive Bays model with BOW, and naive Bayes model with TF-IDF, respectively. The closer the ROC curve was to the upper left, the greater the overall accuracy of the model. The AUCs of the logistic regression model, naive Bayes model with BOW, and naive Bayes model with TF-IDF were 0.97, 0.96, and 0.97, respectively. This result suggests that these three models have the potential of becoming a model that can classify the posts as suicidal or non-suicidal.

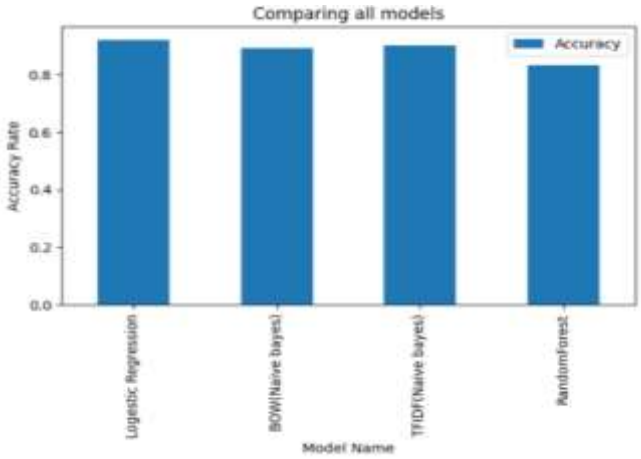

Fig. 6 Accuracy Comparison of Different Models

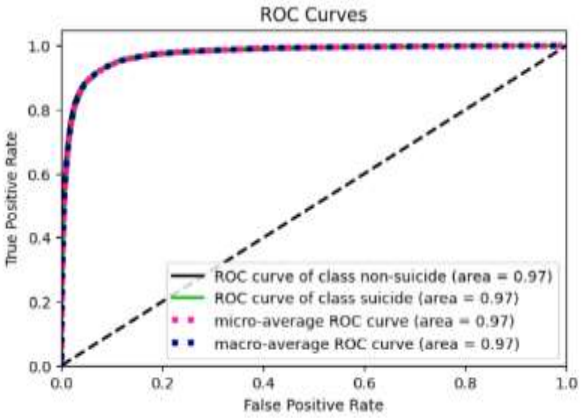

Fig. 7 ROC Curve for Logistic Regression model with Doc2vec

https://doi.org/10.31436/ijpcc.v10i2.468

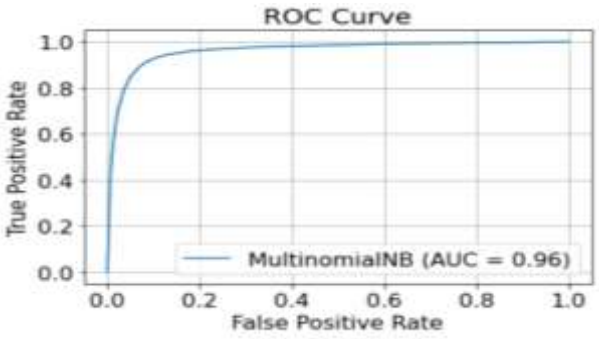

Fig. 8 ROC Curve for Naive Bayes with BOW

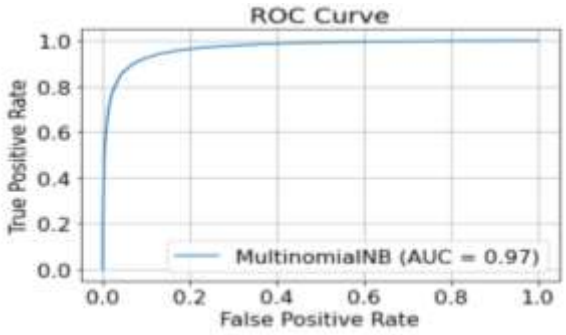

Fig. 9 ROC Curve for Naive Bayes with TF-IDF

Figure 10 shows the comparison of precision, recall, and F_1 score across the three models: logistic regression with doc2vec, naive Bayes with BOW, and naive Bayes with TF-IDF. The precisions of these models were 0.92, 0.89, and 0.89, respectively. The recalls were 0.92, 0.87, and 0.88, and the F_1 scores were 0.92, 0.90, and 0.90. The logistic regression model with doc2vec still had the highest precision, recall and F_1 score.

Figures 11 and 12 display the top 20 suicidal/non-suicidal words used, based on coefficients of the logistic regression model after training. The words 'suicide', 'kill', 'me', and end' were used to categorise suicidal words. On the other hand, the words 'call', 'them', 'matter', 'ago', 'believe' were for non-suicidal words.

To check the validity of the logistic regression model, we computed Cohen's kappa, which was 0.85. According to the Table 1, this means that there is an almost perfect agreement between the classification performed by the model and the classification related to the posts of the subreddits of the Reddit platform.

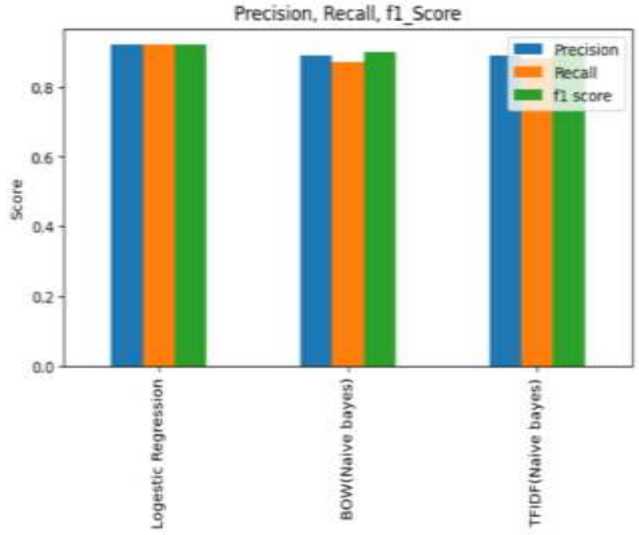

Fig. 10 Precision, Recall and F1-Score comparison

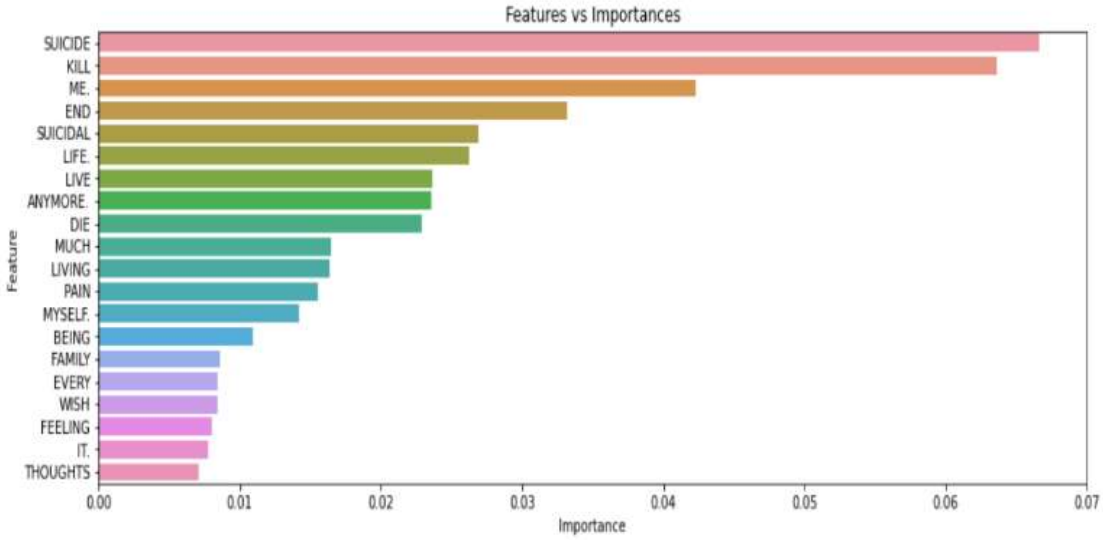

Fig. 11 Suicidal feature importance

https://doi.org/10.31436/ijpcc.v10i2.468

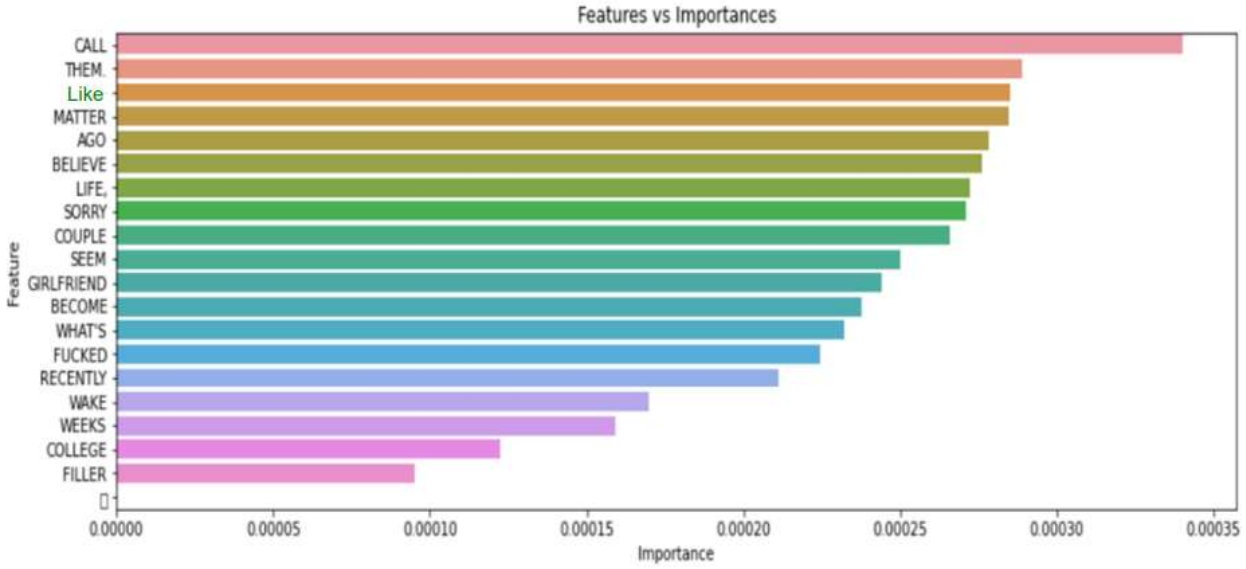

Fig. 12 Non-suicidal feature importance

V. Conclusions

In conclusion, we compared the three models: logistic regression, naive Bayes, and random forest models, to show the possibility of suicide risk prediction from social media data. For the logistic regression model, doc2vec was used as NLP preprocessing. For the naive Bayes model, the two NLP techniques, i.e. BOW and TF-IDF, were compared. Then, the logistic regression model showed the highest precision of 0.92, recall 0.92, and F_1 score of 0.92. Top 20 suicidal/non-suicidal words were identified according to the coefficients of the logistic regression model (Figs. 11 and 12). The suicidal words included 'suicide', 'kill', 'me', and end', while 'call', 'them', 'matter', 'ago', and 'believe' were non-suicidal words. Moreover, the Cohens' kappa was computed, and it was 0.85. This result shows that the model has the potential of classifying posts from people with potential suicidal tendencies.

Further research is necessary to validate the performance of the model. This study suggests that the logistic regression model with doc2vec NLP preprocessing may work for suicide risk prediction.

Acknowledgment

The authors hereby acknowledge the review support offered by the IJPCC reviewers who took their time to study the manuscript and find it acceptable for publishing.

Conflict of Interest

The authors declare that there is no conflict of interest.

References

[1] W.H.O. World Health Organization, "Suicide," World Health Organisation, Aug. 28, 2023. https://www.who.int/news-room/fact-sheets/detail/suicide

[2] H.-C. Shing, S. Nair, A. Zirikly, M. Friedenberg, H. Daumé III, and P. Resnik, "Expert, Crowdsourced, and Machine Assessment of Suicide Risk via Online Postings," ACLWeb, Jun. 01, 2018.

[3] S. Ji, C. P. Yu, S. Fung, S. Pan, and G. Long, "Supervised Learning for Suicidal Ideation Detection in Online User Content," Complexity, vol. 2018, pp. 1–10, Sep. 2018.

[4] S. C. Shetty, "A Deep Learning Approach for Suicide Risk Assessment using Reddit," norma.ncirl.ie, 2020. https://norma.ncirl.ie/4420/ (accessed Dec. 29, 2023).

[5] T.H. Aldhyani, T. H., Alsubari, S. N., Alshebami, A. S., Alkahtani, H., & Ahmed, Z. A. (2022). "Detecting and analyzing suicidal ideation on social media using deep learning and machine learning models." International journal of environmental research and public health, 19(19), 12635.

[6] S. Ji, Yu, C. P., Fung, S. F., Pan, S., & Long, G. (2018). "Supervised learning for suicidal ideation detection in online user content." Complexity, 2018.

[7] B. O'dea, Wan, S., Batterham, P. J., Calear, A. L., Paris, C., & Christensen, H. (2015). "Detecting suicidality on Twitter." Internet Interventions, 2(2), 183-188.

[8] L.J. Richard and G. G. Koch, "The Measurement of Observer Agreement for Categorical Data," Biometrics, vol. 33, no. 1, pp. 159–174, Mar. 1977.

[9] M. H. Zweig and G. Campbell, "Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine," Clinical Chemistry, vol. 39, no. 4, pp. 561–577, Apr. 1993.

[10] L. Vergni and F. Todisco, "A Random Forest Machine Learning Approach for the Identification and Quantification of Erosive Events," Water, vol. 15, no. 12, p. 2225, Jan. 2023.

[11] A. Lejeune, Le Glaz, A., Perron, P.-A., Sebti, J., Baca-Garcia, E., Walter, M., Lemey, C., & Berrouiguet, S. (2022). "Artificial intelligence and

suicide prevention: A systematic review. " European Psychiatry, 65(1), e19, 1–8.

[12] H. Tang, A. M. Rekavandi, D. Rooprai, G. Dwivedi, F. M. Sanfilippo, F. Boussaid, and M. Bennamoun, "Analysis and evaluation of explainable artificial intelligence on suicide risk assessment," Scientific Reports, 15(1), 53426, 2024.

[13] A. Roy., Nikolitch, K., McGinn, R., Jinah, S., Klement, W., & Kaminsky, Z. A. (2020). "A machine learning approach predicts future risk to suicidal ideation from social media data." npj Digital Medicine, 7(1), 1-10.

[14] D. Dataset, Suicide and Depression Detection dataset https://www.kaggle.com/datasets/nikhileswarkomati/suicide-watch