Analyzing Social Media texts for Suicidal Risk Identification using Natural Language Processing

Shanskar Rai
Department of Computer Science & Engineering
Sikkim Manipal Institute of Technology
Majitar, India
shanskar_202111503@smit.smu.edu.in

Sumiran Bhattarai
Department of Computer Science & Engineering
Sikkim Manipal Institute of Technology
Majitar, India
sumiran_202111502@smit.smu.edu.in

Premjyoti Dhar
Department of Computer Science & Engineering
Sikkim Manipal Institute of Technology
Majitar, India
premjyoti_202111504@smit.smu.edu.in

Biraj Upadhyaya
Department of Computer Science & Engineering
Sikkim Manipal Institute of Technology
Majitar, India
upadhayaya.biraj@smit.smu.edu.in

Kalpana Sharma
Department of Computer Science & Engineering
Sikkim Manipal Institute of Technology
Majitar, India
kalpana.s@smit.smu.edu.in

Sandeep Gurung
Department of Computer Science & Engineering
Sikkim Manipal Institute of Technology
Majitar, India
sandeep.gu@smit.smu.edu.in

2023 9th International Conference on Signal Processing and Communication (ICSC) | 979-8-3503-8320-1/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICSC60394.2023.10441398

***Abstract*—Suicide is a pressing public health concern, with nearly 700,000 people taking their own lives annually, equating to one life lost every 45 seconds as per the World Health Organization. Recognizing that individuals contemplating suicide may exhibit identifiable behavioral patterns prior to their attempts, this work aims to employ Natural Language Processing (NLP) algorithms for prediction of risk of committing suicide by analyzing these patterns in social media posts. The system presented in this study uses machine learning to predict and identify suicidal ideation in online social content. To detect suicidal ideas via digital communication, the framework uses machine learning approaches. In this work, the machine learning algorithms SVM, logistic regression and CNN provide accuracy of 94.02%, 93.7% and 92.09% respectively.**

Keywords—suicide, public health concern, natural language processing, suicide risk prediction, social media posts, mental health surveillance, predictive algorithms, machine learning.

I. INTRODUCTION

The area of Artificial Intelligence known as Natural Language Processing (NLP) involves the interaction between human languages and computers. Its primary aim is to equip machines with the ability to comprehend, decipher, and produce human language. One of the most important tasks in NLP is Natural Language Understanding (NLU), which involves extracting meaning from text or speech. This is done through techniques such as parsing, semantic analysis, and sentiment analysis. Parsing is the process of analyzing the grammatical structure of sentence to determine its meaning. Semantic analysis involves identifying the meaning of words and phrases in context. Sentiment analysis involves determining the emotional tone of a piece of text.

In the context of predicting suicide risk, NLP can be used to analyze text-based data is extracted from social media to identify patterns or indicators of suicide risk [2]. Social media platforms are a valuable source of information about a person's mental state, as individuals may share their thoughts, feelings, and experiences online. NLP algorithms can analyze this data to identify patterns and predict which individuals may be at a high risk of suicide [3].

The paper is organized as follows: Section II summarizes the work done in the similar field by other authors. Section III lists down the problem definition which is attempted to be solved through this work. Section IV focuses on the methodology used in the paper. Section V highlights the pictorial design of the methodology. Section VI discusses the various algorithms used for model training. Section VII includes the results and discussions of the work outcome. Section VIII includes the conclusion of the paper.

II. RELATED WORKS

It is worth mentioning that suicide is ranked within top ten causes of mortality worldwide. In order to forecast suicide-related sentiment in social networks, Birjali et al. discuss the application of machine learning and semantic sentiment analysis [1]. The authors have suggested methods and algorithms for sifting through social media posts to find and forecast suicidal sentiment. In a psychiatric clinical research database, Fernandes et al. used natural language processing (NLP) methods to find suicidal ideation and attempts [2]. The authors share their research on the use of NLP to identify pertinent language patterns linked to suicide risk. Natural language processing is investigated by Coppersmith et al. [3] for identifying suicidal risk on social media networks. The writers look into techniques for analysing social media posts in order to spot signs of suicidal ideas and actions. In order to identify suicidal ideation on social media, Ramrez-Cifuentes et al. take a multidimensional strategy [4]. To detect suicidal ideas in online encounters, the authors integrate multimodal analysis, relational analysis, and behavioural analysis.A machine learning system is presented by Wani and Nasti [5] for the prediction and detection of suicidal ideation in online social content. The authors talk about the creation of a model that forecasts and finds indications of suicide intentions in digital communication using machine learning techniques.

The detection of suicidal ideation can be accomplished using a variety of machine learning techniques, according to Ji et al. [6]. The authors review the literature in the area and explain the advantages and disadvantages of several machine learning techniques for detecting suicidal thoughts and risk variables in text data. The study by Kumar et al. [7] focuses on identifying suicidal risk in social media content. To identify warning indicators of suicide ideation and risk in

979-8-3503-8320-1/23/$31.00 ©2023 IEEE

Authorized licensed use limited to: Technical University Kosice. Downloaded on April 19,2025 at 21:04:02 UTC from IEEE Xplore. Restrictions apply.

user-generated social media posts, they investigate several NLP and ML techniques. Aldhyani et al. are particularly interested in identifying and examining suicide ideation that is expressed on social media sites [8]. To automatically recognise and examine content that suggests the presence of suicidal thoughts or intents in social media posts, the authors use deep learning and machine learning models. In order to facilitate early intervention, Gaur et al. explore the creation of a knowledge-aware system to evaluate the seriousness of suicide risk [9]. The authors probably introduce a system that makes use of knowledge graphs and other methods to assess a person's level of suicide risk, enabling quick preventive action. A multifaceted strategy is presented by Sinha et al. to locate and investigate suicide ideation on Twitter [10]. In order to better comprehend this phenomenon on social media, the writers probably examine several approaches and strategies used to identify and analyse tweets that reveal suicidal thoughts or intents. Chiroma et al. concentrate on text classification for locating tweets related to suicide [11]. They discuss text classification and machine learning methods used to identify tweets with suicidal content, which helps us understand how such material is distributed on social media.

III. PROBLEM DEFINITION

The problem definition of using NLP as a method to predict risk of suicide using social media is to develop a system that can automatically analyze text data from social media platforms and identify individuals who may be at a high risk of suicide. This involves several sub-problems such as

- Identification and extracting relevant information from social media text data: This includes identifying specific keywords and phrases related to suicide, as well as extracting mental state of an individual and risk factors for suicide.
- Data preprocessing requires careful consideration of many different factors and decisions that can affect quality of data and performance of the model.
- Selecting the most optimum hyperparameters for the machine learning model is a challenging problem because it needs an intellectual knowledge of the model and the problem domain, plus an iterative process of trial and error.
- It is important to note that the goal of this problem definition is not to replace professional assessment and treatment, but rather to provide a tool that can support decision-making and potentially help identify individuals who may be at a high risk of suicide.

IV. METHODOLOGY

A, Data Acquisition

For this, a relevant dataset containing social media posts with suicidal content is required. The collection of such data is a time-consuming task and requires expertise. To address this, publicly available datasets from reputable sources such as Kaggle was searched. Additionally, labelled datasets specifically designed for detecting suicidal ideation were searched for. These labelled datasets are an important resource for training and testing machine learning models for detection of suicidal ideation in social media posts as well as blogs. The quality and relevance of the data can be ensured by combining publicly available and labelled datasets, leading to the development of a more accurate model for early detection and prevention of suicide. The dataset used from Kaggle is “Suicide and Depression Detection” and the salient features of the dataset is given below in Table 1:

Characteristics	**Description**
Dataset Name	Suicide and Depression Detection
Data Source	Reddit
Description	A collection of posts from the "SuicideWatch" and "depression" subreddits of the Reddit platform
Number of Instances	232074
Features/Attributes	text, class
Target Variable(s)	class
Data Format	.csv
Data Size	166.9MB
Data Collection Period	12/15/2008 to 01/01/2021
Data Collection Method	Web scrapping
Missing Values	0%
License	CC BY-SA 4.0

Table 1: Dataset description

B. Data Preprocessing

Before feeding the dataset into machine learning models, it must be preprocessed. Several techniques are used in this preprocessing step, including removing punctuation, removing stopwords, and lemmatizing the text. Removing punctuation will help to simplify the text and remove extra noise that could affect the model's accuracy. Stopwords like "and," "the," "a," and so on have no meaning and can also affect the model's accuracy, so they will be removed. Lemmatization is the process of reducing words to their base or root form, which improves the model's efficiency and accuracy. This preprocessing step is essential for ensuring that the text data is in a format that is suitable for machine learning models to analyze and make predictions.

C. *Feature Extraction*

The project on suicidal thoughts will employ the feature extraction method. This entails locating and picking out pertinent dataset elements that are most suggestive of suicidal content in social media posts. To extract text data features, the Bag-of-Words and Term Frequency-Inverse Document Frequency (TF-IDF) approaches will be used. Before being used as input for machine learning models, the retrieved features will be transformed, creating a predictive model that can recognise suicidal intent in social media posts. In order to ensure that the model effectively captures the underlying patterns and characteristics of suicidal content in social media posts, feature extraction is essential. This helps to increase the model's overall accuracy.

D. Model Training

Machine learning model training will come after data pretreatment and feature extraction. This procedure comprises feeding the model with the preprocessed data and letting it absorb knowledge from the relationships and patterns in the data. The models will be constructed using a variety of regression techniques, including logistic regression, linear regression, and support vector machines. The models will be trained on a labelled dataset, and measures like accuracy, precision, and recall will be used to assess their performance. Choosing the best model and fine-tuning it to attain the desired performance are all possible steps in the training process. Hyperparameter adjustments may also be necessary. Our system for detecting suicidal

Authorized licensed use limited to: Technical University Kosice. Downloaded on April 19,2025 at 21:04:02 UTC from IEEE Xplore. Restrictions apply.

ideation can be made more accurate and reliable by appropriately training the machine learning models.

E. Model Evaluation

A confusion matrix will be used to assess how well the machine learning models performed in the suicide ideation project. This matrix will show how the model's anticipated results stack up against the actual outcomes. Four categories—true positives, true negatives, false positives, and false negatives—will make up the matrix. We can calculate the models' accuracy, precision, recall, and F1 score by looking at these areas. The best model can be chosen and then adjusted for the best outcomes using the information provided here.

V. DESIGN

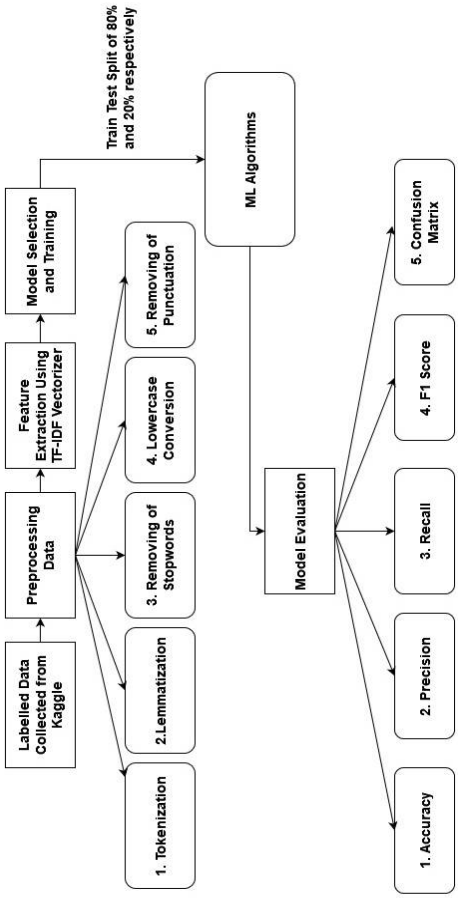

Fig 1. Design Diagram

VI. MODEL TRAINING

a. Random Forest

- A model that could determine whether a specific social media post was suggestive of suicide ideation was developed using the random forest algorithm. The actions were as follows:
- Train-test split: The train_test_split function from scikit-learn was used to divide the data into training and testing sets.
- TF-IDF vectorization: The scikit-learn tfidfvectorizer tool was used to convert the text input into numerical features. Because machine learning algorithms need numerical data to generate predictions, this was important.
- The scikit-learn RandomForestClassifier function was used to fit the random forest model to the training set of data. The countvectorizer function was used to transform the training data into a format that the model could use once it had been trained with a random state of 42.
- Predictions and evaluation: The model was used to create predictions based on test data, and the accuracy_score function from scikit-learn was used to determine the model's accuracy score. The number of true positives, true negatives, false positives, and false negatives for the model were also displayed on the confusion matrix.

b. Logistic Regression

- Logistic regression was also used to predict suicide ideation from social media posts.
- The following steps were taken:
- Train-test split: The data was split into training and testing sets using the train_test_split function
- TF-IDF vectorization: The text data was transformed into numerical features using the tfidfvectorizer function.
- Fitting the model: The logistic regression model was fit to the training data using the LogisticRegression function from scikit-learn.
- Predictions and evaluation: The model was used to make predictions on the test data and the accuracy score of the model was calculated using the accuracy_score function. The confusion matrix was also printed to show the number of true positives, true negatives, false positives, and false negatives for the model.

c. K Nearest Neighbors

- A KNN model was developed to forecast suicidal thoughts. The actions were as follows:
- Dividing the dataset: Utilising the train_test_split tool from sklearn, divide the dataset into training and test sets.
- Feature extraction: The scikit-learn tfidfvectorizer tool was used to convert the text data into numerical features.
- Launching the KNN model The desired number of neighbours was set as the value of k when creating an instance of the KNeighborsClassifier class.
- Develop the model: used the fit approach to train the KNN model using the training data.
- Assess the model: Evaluation of the KNN model's performance using metrics including the confusion matrix, precision, recall, and F1-score.
- Fine-tune hyperparameters: Using the grid search method, the KNN model's hyperparameters were adjusted to enhance performance.
- Make predictions: Using the predict approach, the trained KNN model was used to make predictions on fresh, unforeseen data.

d. Simple Neural Network

- A pre-trained GloVe word embedding file is loaded and used to create an embeddings dictionary that maps each word to its 100-dimensional vector representation.
- The neural network architecture consists of an embedding layer, a flatten layer, and a dense output layer with sigmoid activation.

Authorized licensed use limited to: Technical University Kosice. Downloaded on April 19,2025 at 21:04:02 UTC from IEEE Xplore. Restrictions apply.

- The model is trained on the training set using binary cross-entropy loss and the Adam optimizer.
- The performance of the model is evaluated on the test set.

e. Convolutional Neural Network

- A pre-trained GloVe word embedding file is loaded and used to create an embeddings dictionary that maps each word to its 100-dimensional vector representation.
- An embedding matrix is created containing the 100-dimensional GloVe word embeddings for all words.
- The neural network architecture consists of an embedding layer, a 1D convolutional layer with 128 filters and a filter size of 5, a global max pooling layer, and a dense output layer with sigmoid activation.
- The model is trained on the training set using binary cross-entropy loss and the Adam optimizer.
- The performance of the model is evaluated on the test set using the evaluate() method.

f. Support Vector Machines

- SVMs are a subset of supervised learning algorithms used in regression and classification.
- In binary classification, SVMs look for the hyperplane that divides the data into the two groups most effectively.
- The margin, or the distance between the hyperplane and the nearest data points of each class (known as support vectors), is maximised while choosing the hyperplane.
- In order to translate the data into a higher-dimensional space where linear separation is feasible while yet allowing for nonlinear decision boundaries, SVMs can also be used.
- SVMs are very helpful in high-dimensional spaces and where there is a distinct line dividing classes.
- Convex optimisation techniques, such as the Lagrange multiplier approach, are used to solve the optimisation problem to determine the best hyperplane.

VII. RESULTS AND DISCUSSION

This section presents the classification outcomes of the logistic regression and random forest models that were trained and tested using textual characteristics taken from Reddit postings. Table 2 details how the accuracy, recall, precision, and F1-scores calculated from the confusion matrices were used to assess the classification outcomes using textual characteristics. The different models are compared and it is shown in Fig 2.

Algorithms	Accuracy	Recall	Precision	F1 score
Random Forest	0.8946	0.9022	0.8891	0.8956
Logistic Regression	0.9373	0.9283	0.9448	0.9365
KNN	0.62480	0.612	0.7073	0.5931
SNN	0.8693	0.8582	0.8768	0.8674
CNN	0.9209	0.9256	0.9165	0.9210
SVM	0.9402	0.9363	0.9431	0.9397

Table 2. Performance Evaluation

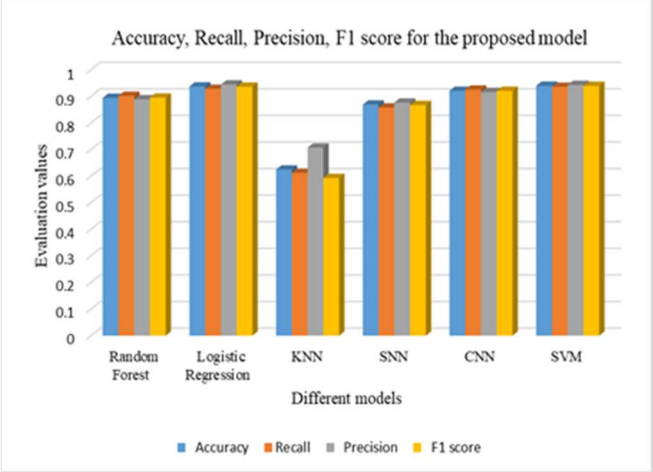

Fig 2. Comparison of different models

Confusion Metrics

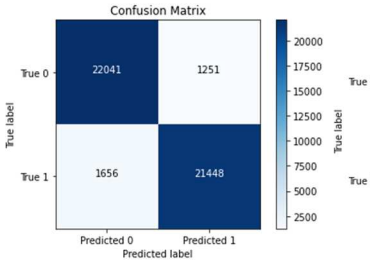

Fig 3. Logistic Regression

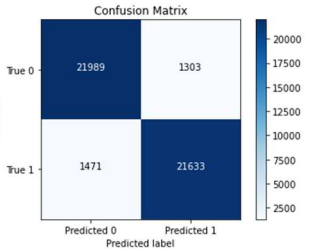

Fig 4. Random Forest

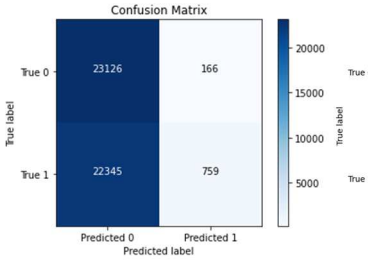

Fig 5. KNN

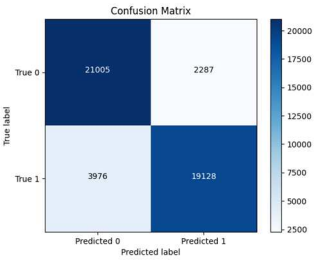

Fig 6. SNN

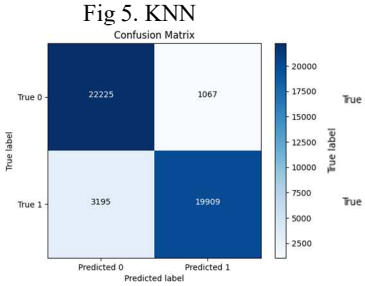

Fig 7. CNN

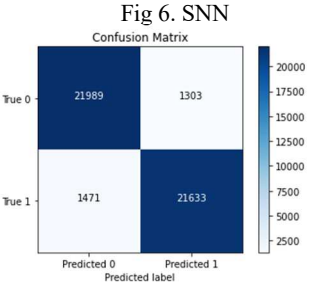

Fig 8. SVM

VIII. CONCLUSION

The goal of the study is to create a machine learning model that can foresee suicidal ideation in social media posts. The "Suicide and Depression Detection" dataset from Kaggle, which included instances of social media posts categorised as indicative or not of suicide and depression, was used for the research. The feature extraction was carried out using the TF-IDF vectorization technique after the data had been pre-processed using the spacy library to weed out extraneous information. For model training, a number of machine learning techniques were employed, including support vector machines, logistic regression, K closest neighbours,

Authorized licensed use limited to: Technical University Kosice. Downloaded on April 19,2025 at 21:04:02 UTC from IEEE Xplore. Restrictions apply.

basic neural networks, and random forests. Based on the accuracy score as well as additional measures like precision, recall, and F1-score, each method was assessed.

The experiment's findings lead us to the conclusion that the logistic regression and SVM algorithms are the most appropriate for our needs. The accuracy metrics for SVM and logistic regression were 94.02% and 93.7%, respectively, while the CNN model also provided us with a respectable accuracy score of 92.09%.

REFERENCES

[1] Birjali, M., Beni-Hssane, A., & Erritali, M. (2017). Machine learning and semantic sentiment analysis based algorithms for suicide sentiment prediction in social networks. Procedia Computer Science, 113, 65-72.

[2] Fernandes, A. C., Dutta, R., Velupillai, S., Sanyal, J., Stewart, R., & Chandran, D. (2018). Identifying suicide ideation and suicidal attempts in a psychiatric clinical research database using natural language processing. Scientific reports, 8(1), 7426.

[3] Coppersmith, G., Leary, R., Crutchley, P., & Fine, A. (2018). Natural language processing of social media as screening for suicide risk. Biomedical informatics insights, 10, 1178222618792860.

[4] Ramírez-Cifuentes, D., Freire, A., Baeza-Yates, R., Puntí, J., Medina-Bravo, P., Velazquez, D. A., ... & Gonzàlez, J. (2020). Detection of suicidal ideation on social media: multimodal, relational, and behavioral analysis. Journal of medical internet research, 22(7), e17758.

[5] Wani, E. G. H., & jan Nasti, S. (2020). A Machine Learning Framework For Prognostication Of Suicidal Ideation And Detection In Online Social Content. Webology (ISSN: 1735-188X), 17(4).

[6] Ji, S., Pan, S., Li, X., Cambria, E., Long, G., & Huang, Z. (2020). Suicidal ideation detection: A review of machine learning methods and applications. IEEE Transactions on Computational Social Systems, 8(1), 214-226.

[7] Kumar, A., Trueman, T. E., & Abinesh, A. K. (2021). Suicidal risk identification in social media. Procedia Computer Science, 189, 368-373.

[8] 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.

[9] Gaur, M., Alambo, A., Sain, J. P., Kursuncu, U., Thirunarayan, K., Kavuluru, R., ... & Pathak, J. (2019, May). Knowledge-aware assessment of severity of suicide risk for early intervention. In The world wide web conference (pp. 514-525)

[10] Sinha, P. P., Mishra, R., Sawhney, R., Mahata, D., Shah, R. R., & Liu, H. (2019, November). # suicidal-A multipronged approach to identify and explore suicidal ideation in twitter. In Proceedings of the 28th ACM international conference on information and knowledge management (pp. 941-950).

[11] Chiroma, F., Liu, H., & Cocea, M. (2018, July). Text classification for suicide related tweets. In 2018 International Conference on Machine Learning and Cybernetics (ICMLC) (Vol. 2, pp. 587-592). IEEE.

[12] World Health Organization. (2014). Preventing suicide: A global imperative. World Health Organization.

Authorized licensed use limited to: Technical University Kosice. Downloaded on April 19,2025 at 21:04:02 UTC from IEEE Xplore. Restrictions apply.