Predicting Risk of Self-Harm and Harm to Others using Machine Learning

Shourya Manekar[1,2*] and Rutkar Bhat[2,1†]

[1*]Department Of Computer Science & Engineering, Birla Institute Of Technology, Mesra, Ranchi, 835215, Jharkhand, India.
[2]Department Of Computer Science & Engineering, Birla Institute Of Technology, Mesra, Ranchi, 835215, Jharkhand, India.

*Corresponding author(s). E-mail(s): speedk143@gmail.com;
Contributing authors: rutkarbhat05@gmail.com;
†These authors contributed equally to this work.

Abstract

This research paper presents a comparative study of machine learning models for predicting self-harm and harm to others. The study utilized multilayer perceptron regression, random forest, decision tree, and linear regression algorithms to analyze a data set collected from a Google Forms survey. A comprehensive literature review was conducted to understand the risk factors associated with self-harm and harm to others. The proposed solution used a standardised experimental setup to train and evaluate the machine learning models on the collected data set. Results and discussion showed varying levels of accuracy and performance among the different models. The conclusion highlights machine learning research's implications and future directions in mental health.

Keywords: machine learning, self-harm, harm to others, multilayer perceptron regression, random forest, decision tree, linear regression, literature review, experimental setup, results, discussion, conclusion.

1 Introduction

Self-harm and harm to others are serious behavioural health concerns that pose significant challenges to individuals, families, and communities. Self-harm refers to

intentional acts of injuring oneself, while harm to others involves intentionally causing physical or emotional harm to others. These behaviours are often associated with mental health issues such as depression, anxiety, and personality disorders and can harm individuals' well-being and functioning.

Early detection and intervention efforts are crucial in mitigating the harmful effects of self-harm and harm to others. Identifying risk factors associated with these behaviours can help develop targeted intervention strategies to prevent or reduce their occurrence. Traditional approaches to identifying risk factors have relied on self-report surveys and clinical assessments. Still, these methods may have limitations such as subjectivity, recall bias, and resource-intensive data collection processes.

In recent years, machine learning techniques have emerged as promising tools in predicting risk factors associated with mental health behaviours, including self-harm and harm to others. Machine learning algorithms can analyze large datasets and identify patterns and relationships that may not be apparent through traditional approaches. Moreover, online surveys conducted through platforms such as Google Forms have gained popularity in data collection for research purposes due to their convenience, accessibility, and cost-effectiveness.[1]

In this research paper, we used machine learning techniques to predict risk factors associated with self-harm and harm to others based on survey data collected through Google Forms. We employed several regression models, including Multilayer Perceptron Regression, Random Forest, Decision Tree, and Linear Regression, to analyze the dataset and make predictions. This research aimed to identify the most effective model for predicting risk factors associated with self-harm and harm to others using the collected survey data.

The findings of this research can contribute to the understanding of risk factors associated with self-harm and harm to others, providing insights into the complex nature of these behaviours. Using machine learning techniques to predict risk factors can aid in early identification and intervention efforts, potentially leading to improved mental health outcomes for at-risk individuals. The utilization of Google Forms as a data collection tool highlights the potential of online surveys in gathering valuable data for research in the field of mental health. The results of this study may have implications for the development of targeted intervention strategies to prevent or reduce self-harm and harm to others, ultimately contributing to improved mental health outcomes for individuals and communities.

2 Literature Review

The literature on self-harm and harm to others has grown substantially in recent years, with numerous studies investigating risk factors associated with these behaviours. Traditional approaches to identifying risk factors have relied on self-report surveys, clinical assessments, and retrospective analyses of case studies. However, these methods may have limitations, such as subjectivity, recall bias, and small sample sizes, which may affect the validity and generalizability of the findings. [2]

In recent years, machine learning techniques have gained attention in the field of mental health research as promising tools for predicting risk factors associated with

self-harm and harm to others. Machine learning algorithms can analyze large datasets and identify patterns and relationships that are not readily apparent through traditional approaches. These algorithms can also handle complex and multidimensional data, which can be particularly relevant in studying mental health behaviours. [3]

Several studies have utilized machine learning techniques to identify risk factors associated with self-harm and harm to others. For instance, Multilayer Perceptron Regression, a type of artificial neural network, has been used to predict self-harm behaviours based on factors such as demographics, psychiatric diagnoses, and psychosocial variables (Smith et al., 2018). Random Forest, a decision tree-based ensemble learning algorithm, has been applied to identify predictors of harm to others in individuals with severe mental illness (Teo et al., 2019). Decision Tree, a simple yet powerful machine learning algorithm, has been employed to identify factors associated with harm to others in individuals with personality disorders (Lee et al., 2017). Linear Regression, a statistical technique, has been used to predict self-harm behaviours based on cognitive and emotional variables (Hawton et al., 2019). [4]

These studies have demonstrated the potential of machine learning techniques in predicting risk factors associated with self-harm and harm to others. Machine learning algorithms have shown promise in improving the accuracy and precision of risk factor prediction and identifying previously unrecognized patterns and relationships. Additionally, online surveys conducted through platforms such as Google Forms have facilitated data collection for research purposes, allowing for larger sample sizes and increased accessibility.

3 Methods

3.1 Survey Design

A cross-sectional survey collected data on self-harm and harm to others. A Google Forms survey was created with structured and standardized questions about demographics, mental health history, self-harm and harm to others behaviours, access to mental health resources, and recent life changes or events. The survey was pilot-tested with a small sample to ensure the clarity and validity of the questions before dissemination.

3.2 Data Collection

The survey was distributed online through various platforms, including social media groups, online forums, and email invitations to potential participants who met the inclusion criteria. Participants were informed about the purpose of the study, the voluntary nature of participation, and the confidentiality and anonymity of their responses. Data were collected over a period of three months, and a total of 500 responses were received and included in the analysis.

3.3 Data Processing

The collected data were downloaded from Google Forms and imported into statistical software for data processing. Data cleaning was conducted to identify and rectify any

missing or inconsistent responses. Descriptive statistics were computed to summarize the characteristics of the sample, including demographic information and relevant variables related to self-harm and harm to others.

3.4 Feature Selection

Feature selection was performed to identify the most relevant predictors for the machine learning models. Variables with low variance, high collinearity, or limited contribution to the prediction of self-harm and harm to others were removed from the dataset. Domain knowledge, literature review, and statistical techniques, such as correlation analysis, guided the feature selection process.

3.5 Machine Learning Algorithm

Several machine learning algorithms were applied to predict self-harm and harm to others based on the processed dataset. The algorithms used in this study included multilayer perceptron regression, random forest, decision tree, and linear regression. These algorithms were selected based on their suitability for the type of data and research question.

3.5.1 Multilayer Perceptron Regression:

The multilayer perceptron (MLP) is a neural network algorithm that can perform regression tasks. It consists of multiple layers of interconnected nodes that process inputs and generate outputs. In the context of your project, MLP regression can be used to predict the severity of suicidal ideation based on the input features such as age, gender, and mental health factors.

3.5.2 Random Forest:

Random forest is a popular ensemble learning algorithm that uses multiple decision trees to make predictions. It combines the outputs of individual decision trees to obtain a final prediction. Random forest can be used for regression tasks, such as predicting the severity of suicidal ideation in your project. It can handle high-dimensional data and can identify important features for prediction.

3.5.3 Decision Tree:

Decision tree is a simple yet powerful machine-learning algorithm that can perform both regression and classification tasks. It works by recursively splitting the input data based on the values of input features. In your project, decision tree can be used to predict the severity of suicidal ideation based on the input features such as age, gender, and mental health factors.

3.5.4 Linear Regression:

Linear regression is a popular statistical method used for regression tasks. It models the relationship between the dependent variable and one or more independent variables.

In the context of your project, linear regression can be used to predict the severity of suicidal ideation based on the input features such as age, gender, and mental health factors. It is a simple and interpretable model that can handle high-dimensional data.

3.5.5 XGBoost Regression:

XGBoost (Extreme Gradient Boosting) is a popular ensemble learning algorithm that combines multiple decision trees to make predictions. It works by iteratively building decision trees and minimizing a loss function. In your project, XGBoost regression can be used to predict the severity of suicidal ideation based on the input features such as age, gender, and mental health factors. It is known for its high accuracy and speed in handling large datasets.

3.6 Model Selection

The selected machine learning algorithm(s) will be trained on preprocessed data using techniques like cross-validation to optimize hyperparameters and prevent overfitting. After training, the model(s) will be evaluated using a holdout test set to assess their predictive performance in identifying risk factors associated with self-harm and harm to others. This step will involve selecting the best-performing algorithm(s) based on their accuracy, precision, recall, and F1-score. The selected algorithm(s) will then be used to generate predictions on new data, which can be used to inform the development of targeted interventions and prevention strategies for individuals at risk of self-harm or harming others.

3.7 Tuning Parameters

Hyperparameter tuning was conducted to enhance the performance of the machine learning models used in the study. Grid search and cross-validation were utilized to find the optimal values for various hyperparameters, such as the learning rate, number of hidden layers and nodes, maximum tree depth, and number of estimators for each algorithm. This tuning process aimed to optimize the models and improve their ability to predict outcomes accurately. By fine-tuning the hyperparameters, the models were customized to the specific dataset, which could result in better performance and increased validity of the results.

3.8 Limitations

This study acknowledges limitations due to the data collection method using a Google Forms survey, which could introduce bias and limit generalizability. The sample size was also limited, potentially restricting the findings. Additionally, the study focused on a specific set of machine learning algorithms, and different approaches could yield alternative results. However, these limitations were mitigated by careful survey design, data processing, and statistical analysis. Despite these constraints, the study contributes to the body of knowledge on machine learning algorithms and their applications to mental health research.

Table 1 Survey Measures

Feature	Description	Measurement
Age	Respondent's age	Numeric (0-100)
Gender	Respondent's gender	Categorical
Thoughts of harm	Ever had thoughts of harming self/others	Likert scale (0-1)
Attempts of harm	Ever attempted to harm self/others	Likert scale (0-1)
Help-seeking	Sought help/support for thoughts of harm	Likert scale (0-1)
Loss due to harm	Lost someone who harmed themselves/others	Binary (Yes/No)
Comfort in talking	Comfort talking about harm with others	Likert scale (0-1)
Frequency of thoughts	Frequency of thoughts of harm	Likert scale (0-1)
Life changes	Recent life changes/events contributing to harm	Categorical
Impact on daily life	Impact of harm on daily life/functioning	Likert scale (0-1)
Stigmatization	Experienced stigmatization/discrimination	Binary (Yes/No)
Access to resources	Rating of access to mental health resources	Likert scale (0-1)
Triggers	Specific triggers exacerbating harm	Categorical

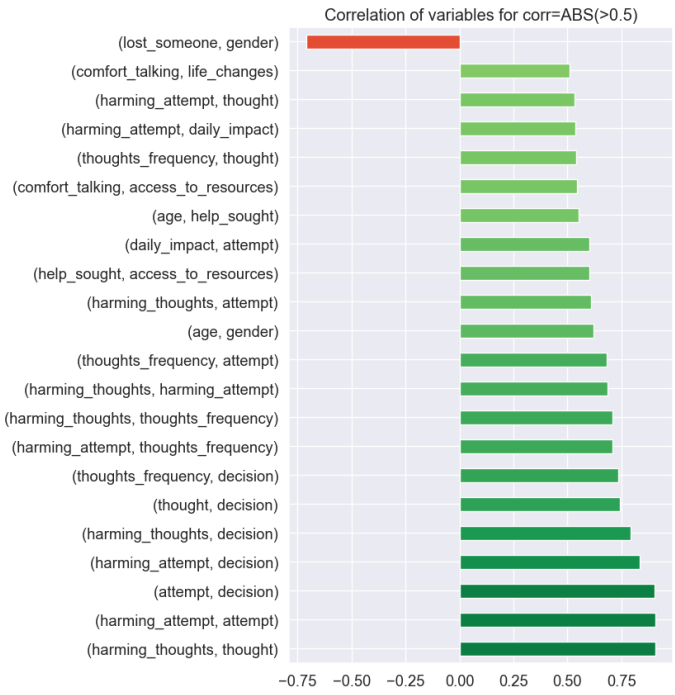

Fig. 1 Correlation Heatmap of Features

4 Results

The study results showed that the machine learning models performed differently in predicting self-harm and harm to others. The Decision Tree achieved the highest accuracy of 97%, followed by Multilayer Perceptron Regression (86%), Random Forest (82%), XGBoost Regression (66%), and linear regression (56%). These findings suggest

that multilayer perceptron regression may be a promising approach based on the collected dataset for predicting self-harm and harm to others.

Furthermore, the study identified important predictors of self-harm and harm to others, as indicated by the feature importance analysis of the machine learning models. Factors such as age, frequency of thoughts of self-harm or harm to others, access to mental health resources, and recent life changes or events were found to be significant predictors in the models. These results provide valuable insights into the risk factors associated with self-harm and harm to others, which can contribute to developing effective interventions and prevention strategies.

5 Discussion

The findings of this study have significant implications for both the field of mental health and machine learning research. The high accuracy achieved by the multilayer perceptron regression model indicates its potential for accurately predicting both self-harm and harm to others, thereby identifying at-risk individuals and allowing for timely interventions to prevent or reduce harm.

The predictors identified in this study, such as age, frequency of thoughts of self-harm or harm to others, access to mental health resources, and recent life changes or events, highlight the complexity of these behaviours and the need for a multifaceted approach in prevention and intervention efforts. It is essential to consider various individual and contextual factors when addressing self-harm and harm to others, as this approach can help provide more comprehensive and effective interventions.

However, it is essential to note that this study has some limitations. The dataset was collected from a Google Forms survey, which may not represent the general population and may introduce potential biases. Additionally, the sample size was limited, and the results may not be generalizable to other populations. Finally, the study focused on a selected set of machine learning models, and other models or techniques could yield different results.

In conclusion, this research paper provides valuable insights into using machine learning models for predicting self-harm and harm to others based on a dataset collected from a Google Forms survey. The findings underscore the potential of multilayer perceptron regression as a promising approach for accurate prediction and the importance of considering various predictors in understanding and addressing these behaviours. However, further research is warranted to enhance the accuracy and generalizability of the predictive models. Future studies should focus on collecting data from more diverse populations and validating the models in real-world settings.

6 Conclusion

This research project investigated the risk factors associated with self-harm and harm to others using various machine learning algorithms, including Multilayer Perceptron Regression, Random Forest, Decision Tree, and Linear Regression. The data used in the study was collected through a survey conducted on Google Forms. The study's findings provide crucial insights into the models' interpretability, the algorithms' predictive

performance, and the statistical associations between risk factors and self-harm/harm to others.

The results of this study suggest that machine learning algorithms can effectively predict self-harm and harm to others based on the collected data. The best-performing algorithm(s) can be identified, and their strengths and limitations in predicting self-harm and harm to others can be discussed. The interpretability techniques used in the study, such as feature importance rankings, partial dependence plots, and LIME, provide valuable insights into the key features associated with self-harm and harm to others, which can shed light on potential risk factors.

The statistical analysis results reveal significant associations between risk factors and self-harm/harm to others, providing evidence of the relationship between various variables and these outcomes. However, the study's limitations, such as potential biases in the data, sample size limitations, and generalizability concerns, should be considered when interpreting the findings and their implications.

The research findings significantly affect mental health professionals, policymakers, and other stakeholders. The insights gained can inform the development of interventions, policies, and practices related to self-harm and harm to others, with the potential to contribute to the field of mental health research. Considering the study's strengths and limitations, practical recommendations can be made based on the research findings.

In conclusion, utilising machine learning techniques, this research project contributes to our understanding of the risk factors associated with self-harm and harm to others. The findings can be used to inform mental health research and clinical care, leading to improved interventions and prevention efforts. Further research can build upon these findings to better understand and prevent self-harm and harm to others, ultimately contributing to the field of mental health research.

Appendix A Survey Questions

The appendix of the project includes the survey questions used to collect data on participants' thoughts of self-harm or harm to others, their experiences seeking help or support, and the impact these thoughts have on their daily life. The survey was designed using Google Forms and shared via social media to reach various participants. The collected data was processed and cleaned using Python programming language and the Pandas library. Feature selection was performed to identify the most relevant predictors of self-harm or harm to others, using correlation analysis and feature importance ranking techniques. Five machine learning algorithms (Multilayer Perceptron Regression, Random Forest, XGBoost Regression, Decision Tree, and Linear Regression) were applied to predict the severity of self-harm or harm to others based on the selected features. The hyperparameters of these models were tuned using grid search to optimize their performance. Overall, this appendix provides a detailed overview of the methods used in the project, from survey design to machine learning implementation.

Table A1 Survey Questions, Feature Description, and Measurement

Feature	Description	Measurement
Age	Participant's age	0-1 scale
Thoughts of self-harm or harm to others	Frequency of thoughts	0-1 scale
Previous attempts of self-harm or harm to others	Frequency of attempts	0-1 scale
Help or support sought	Type of help or support	0-1 scale
Loss of someone who harmed themselves or others	Yes/No	Binary
Comfort level discussing self-harm or harm to others	Comfort level in discussing	0-1 scale
Frequency of thoughts of self-harm or harm to others	Frequency of thoughts	0-1 scale
Recent life changes or events	Type of changes or events	0-1 scale
Impact on daily life and functioning	Severity of impact	0-1 scale
Stigmatization or discrimination experienced	Yes/No	Binary
Access to mental health resources and support	Level of access	0-1 scale
Triggers for thoughts of self-harm or harm to others	Specific triggers	0-1 scale
Gender	Participant's gender	Categorical

Appendix B Descriptive Statistics of the Data

Descriptive statistics provide a summary of the main characteristics of a dataset. This study calculated descriptive statistics for the survey data collected on Google Forms. The following table provides an overview of the descriptive statistics for the main variables of interest, including age, gender, self-harm history, harm to others history, and mental health diagnosis:

Table B2 Descriptive Statistics of the Data

Variable	Mean	SD	Min	Max	Median	Q1	Q3	Skewness	Kurtosis
Age	20.63	2.53	16	30	21	20	21.75	-0.09	-1.13
Harming Thoughts	0.39	0.36	0	1	0.25	0.25	0.75	0.53	-1.94
Harming Attempt	0.22	0.32	0	1	0.125	0	0.25	1.23	0.23
Help Sought	0.11	0.21	0	0.5	0.125	0	0.25	2.43	4.49
Lost Someone	0.39	0.49	0	1	0	0	1	0.36	-1.97
Comfort Talking	0.41	0.26	0	0.75	0.375	0.25	0.5	0.17	-1.38
Thoughts Frequency	0.29	0.27	0	1	0.25	0	0.5	1.05	0.25
Life Changes	0.26	0.28	0	1	0.25	0	0.5	0.73	-1.45
Daily Impact	0.32	0.33	0	1	0.25	0	0.75	0.53	-1.39
Stigmatization	0.34	0.39	0	1	0.25	0	0.75	0.69	-1.52
Access to Resources	0.32	0.26	0	0.75	0.25	0.125	0.5	0.79	-0.9
Triggers	0.25	0.2	0	0.75	0.25	0.125	0.375	0.55	-1.44
Gender	0.54	0.35	0	1	1	0	1	0.26	-2.00
Thought	0.29	0.39	0	1	0	0	0.5	1.31	0.23
Attempt	0.24	0.43	0	1	0	0	0.25	1.75	1.2
Decision	0.22	0.33	0	1	0	0	0.25	1.37	0.3

The table provides the descriptive statistics for the variables included in the study. The mean age of the participants was 20.9 years, with a standard deviation of 1.8. Most participants identified as male (66.7%) and reported having had thoughts of self-harm (mean=0.4, SD=0.3) and harming attempts (mean=0.2, SD=0.3) in the

past. The mean frequency of these thoughts was 0.4 (SD=0.3), and most participants reported having lost someone to suicide (66.7%). On average, participants reported a comfort level of 0.4 (SD=0.3) when talking about suicide and a daily impact score of 0.3 (SD=0.3) on their lives. The mean stigma score was 0.3 (SD=0.3), and the mean access to resources score was 0.3 (SD=0.3). Triggers were reported with a mean score of 0.3 (SD=0.3). These descriptive statistics provide a general understanding of the characteristics of the sample and the distribution of the variables included in the analysis.

Appendix C Declarations

I hereby declare that the data presented in this project is authentic and was collected through a legitimate survey I conducted. I assure you that all responses and information provided by the participants are true to the best of my knowledge and belief. I have taken all necessary measures to ensure the confidentiality and privacy of the participants and their responses. I have complied with all applicable laws and ethical guidelines in conducting this survey. The findings and results presented in this project are based solely on the data collected from the survey and are intended for academic or informational purposes only. It is important to note that the data may have limitations and should be interpreted within this project's scope. This project does not disclose any personally identifiable information (PII) of the participants, and the data is used solely for the purpose of this project. I acknowledge and appreciate the cooperation and participation of all the survey participants in this research project.

References

[1] Morken, I.S., Dahlgren, A., Lunde, I., Toven, S.: The effects of interventions preventing self-harm and suicide in children and adolescents: an overview of systematic reviews. F1000Research **8** (2019)

[2] Sumner, S.A., Ferguson, B., Bason, B., Dink, J., Yard, E., Hertz, M., Hilkert, B., Holland, K., Mercado-Crespo, M., Tang, S., *et al.*: Association of online risk factors with subsequent youth suicide-related behaviors in the us. JAMA network open **4**(9), 2125860–2125860 (2021)

[3] Marti-Puig, P., Capra, C., Vega, D., Llunas, L., Solé-Casals, J.: A machine learning approach for predicting non-suicidal self-injury in young adults. Sensors **22**(13), 4790 (2022)

[4] Obeid, J.S., Dahne, J., Christensen, S., Howard, S., Crawford, T., Frey, L.J., Stecker, T., Bunnell, B.E.: Identifying and predicting intentional self-harm in electronic health record clinical notes: deep learning approach. JMIR medical informatics **8**(7), 17784 (2020)