Contents lists available at ScienceDirect

Asian Journal of Psychiatry

journal homepage: www.elsevier.com/locate/ajp

Letter to the editor

A machine learning model to predict suicidal tendencies in students

Dear editor,

Suicide is a significant global mental and public health issue, causing 800,000 deaths per year. The Asian continent accounts for more than half of the ensuing 800,000 deaths (Suryadevara and Tandon, 2018). Suicide is the second-leading cause of death in the age group between 15 and 29 years (Esmaeili et al., 2022). Some educational institutions have adopted measures to provide psychological support and timely intervention for at-risk students (Cherian et al., 2022). While some studies (Miller et al., 2009) appear to confirm the positive impact of such programs, the efficacy of such superficial measures has been contradicted by others (Farahbakhsh et al., 2022). A recurring criticism of said programs is that often the staff designated to conduct such programs are not trained adequately to properly identify the behavioral markers of depression in students. Several researchers have submitted that there is a need for a standard, dependable identification model to counteract suicides in youth (Farahbakhsh et al., 2022; Fleischmann and De Leo, 2014).

Machine Learning (ML) is a branch of Artificial Intelligence that learns any patterns of interest from large amounts of data. ML techniques are used to build predictive modeling tools that can forecast future occurrences. Application of ML techniques for the prediction and classification of mental illnesses is already underway, gaining popularity at an exponential level (Cho et al., 2019; Shatte et al., 2019; Tandon and Tandon, 2019). Intelligent screening systems designed using machine learning models can provide crucial aid for timely diagnosis and prevention of self-harm. The efficiency of an intelligent screening model is contingent on identifying the correct markers that are relevant and essential for identifying an underlying condition.

Anxiety and depression are two common mental health conditions that are prevalent in adolescents and young adults. When left untreated, they potentially lead to self-harm and suicide. Patient Health Questionnaire-9 (PHQ-9) and Generalized Anxiety Disorder-7 (GAD-7) are questionnaires used to assess short-term depression and anxiety levels, respectively. Some existing ML and statistical models rely solely on depression as a variable to predict suicidal or self-destructive tendencies in individuals and presume that such a singular variable fully captures the complex psychological reality of an individual. The PHQ-9 is used fundamentally to measure the mental state, implicitly assuming that each item carries the weight of its score. These approaches not only suffer from methodological fallibility but also have scope for potential bias.

In order to overcome the said limitation, an ML model is built in which all combinations of anxiety indicators are analysed, yielding a more reliable, valid result. The data for the model was collected from a cross-sectional survey carried out using google forms. The survey included PHQ-9 and GAD-7. Three hundred and forty-six responses were recorded from students at undergraduate and postgraduate levels from south India. The model is premised on two fundamental assumptions. The first is that students will be inclined to share sensitive psychological states in a questionnaire rather than the clinicians/counselors. The second assumption is that individuals are more likely to answer anxiety related questions more accurately than the depression-based questions (PHQ) owing to the ignominy surrounding the depression. Therefore, the PHQ part of the collected dataset is dropped completely with the exception to item- 9 (Thoughts that you are better off dead). Item 9 is the best representative item of suicidal/self-harm ideation (SI). For the purpose of this analysis, the SI is treated as a binary value meaning that any degree of response to item 9 is taken as an indication of SI thoughts. The objective of the present study is to develop a predictive model to accurately predict the dependent variable SI.

This dataset containing label-encoded GAD-7 answers and score levels with their corresponding SI values is randomly split into 80:20 ratio. It is standard in a machine learning problem to use the major portion (80%) of the data to train the model, and the rest of the data is used to test the accuracy with which the trained model performs. We experimented with four state-of-the-art machine learning algorithms: K-Nearest neighbor (KNN), Support vector regression (SVR), Decision trees (DT), and Random Forest (RF).

KNN is a supervised, non-parametric machine learning method. Despite being frequently used for categorization, it performs well in terms of predictions because the algorithm has the characteristic of being sensitive to the local structure of the data. Decision tree is a supervised machine learning algorithm that works well for classification and regression problems. It is popularly used to forecast future outcomes of a probabilistic event, suicidal ideation in this case. Finding the optimum fit line is the fundamental tenet of Support Vector Regression. The random forest algorithm employs ensemble learning on the decision trees. KNN, SVR, DT, RF displayed accuracy of 0.74, 0.65, 0.739, 0.971 respectively. In the current experiment, RF has outperformed the rest of the algorithms by a significant margin. The RF model learned the intricate pattern of GAD score combinations leading to SI in this dataset.

By combining two standard questionnaires, we have attempted to build a machine learning framework to predict suicidal ideations in students. 25% of respondents in this survey reported self-harm thoughts, substantiating the dire need for a dependable system. Factoring in the possibility that teenagers are reluctant to disclose sensitive personal information to parents and clinicians, this technique can be a reliable measure of adolescent mental health. Educational institutions can use

Abbreviations: Suicide, Suicide; Machine learning, ML; Random forest, RF; Self harm, Self harm; Students, Students.

https://doi.org/10.1016/j.ajp.2022.103363
Received 19 July 2022; Received in revised form 20 October 2022; Accepted 10 November 2022
Available online 26 November 2022