DOI: 10.14744/DAJPNS.2023.00221
Dusunen Adam J Psychiatr Neurol Sci
2023;36:179-188

RESEARCH ARTICLE

Predicting suicidal thoughts in a non-clinical sample using machine learning methods

Burcu Turk[1], Hasan Halit Tali[2]

[1]Halic University, Department of Psychology, Istanbul, Turkiye
[2]Halic University, Department of Mathematics, Istanbul, Turkiye

ABSTRACT

Objective: When examining the causes of suicide – an important public health problem – various psychological, social, cultural, and biological factors come to light. Given the complex nature of suicide, machine learning techniques have recently been used in psychological and psychiatric research. Machine learning is defined as the programming of computers to improve their performance using sample data or past experience. This study aims to predict suicidal thoughts in a non-clinical sample using supervised learning classification algorithms, one of the machine learning methods. This method is based on the risk and protective factors associated with suicide.

Method: The Personal Information Form, Coping Attitudes Assessment Scale, and Rosenberg Self-Esteem Scale were used as data collection tools. The study comprised 1,940 participants, with ages ranging between 18 and 30 ($\bar{x}$=20.48, SD=2.45).

Results: Using the ensemble learning model with the Hard Voting approach, the prediction rate for a "yes" answer to the question "Have you had suicidal thoughts in the past year?" was determined to be 82%.

Conclusion: This study is believed to contribute to prevention efforts by addressing potential future suicidal thoughts and preventing existing suicidal thoughts from evolving into actions. This contribution considers suicide-related warning signals and associated protective and risk factors.

Keywords: Suicide, suicidal thoughts, machine learning

INTRODUCTION

Suicide is a major public health concern in Turkiye and globally, regardless of age, gender, race, and other factors. Its incidence is on the rise. According to the Turkish Language Association, suicide is defined as "ending one's own life under the influence of social and mental reasons." The World Health Organization defines it as "self-harm with a fatal intent, carried out with awareness of the act's purpose and implications" (1,2).

According to the World Health Organization, nearly 800,000 people commit suicide every year. In 2016, suicide was the second leading cause of death among young people aged 15-29 worldwide (2). Turkish Statistical Institute (TURKSTAT) data reveals that in Turkiye, 3,161 suicides were reported in 2018 and 3,406 in 2019. Breaking down the 2019 data by gender, 77.09% were men, and 22.91% were women. While suicidal behaviors span all age groups, the most prevalent age bracket for suicides in 2019, as per TURKSTAT was 20-24 years old, accounting for 12% (3).

How to cite this article: Turk B, Tali HH. Predicting suicidal thoughts in a non-clinical sample using machine learning methods. Dusunen Adam J Psychiatr Neurol Sci 2023;36:179-188.

Correspondence: Burcu Turk, Halic University, Department of Psychology, Istanbul, Turkiye
E-mail: burcuturk@halic.edu.tr
Received: March 10, 2023; **Revised:** June 13, 2023; **Accepted:** August 30, 2023

The rising global prevalence of suicidal behavior has driven researchers to investigate its underlying causes. From a theoretical standpoint, the psychoanalytic theory attributes suicide to the loss of a loved one. It describes a shift in anger towards oneself by internalizing the lost object of affection as a part of one's identity (4). Durkheim explores the ties between suicide and various societal factors, positing that suicides can be curbed if individuals are adequately integrated into society (5).

Suicidal behavior spans a broad spectrum, from ideation and attempts to complete acts. This intricate progression from thought to action is influenced by many psychological, social, cultural, and biological determinants (4). One significant factor is an individual's health status and perceptions of medical conditions. Studies suggest that individuals with poor health are more likely to have suicidal thoughts or attempt suicide than those with good health (6). Another factor associated with suicide is an individual's ability to cope with stressful events. The lack of appropriate coping responses can increase the likelihood of suicide (7). The fact that people who attempt suicide, have tentative life plans, feel there are insurmountable obstacles to achieving their goals, and believe that they are unlikely to attain these goals can also be viewed as risk factors for suicide (6). Factors such as marital problems, unemployment, academic issues, low socioeconomic status, living alone, lack of social support, a previous history of suicidal thoughts or attempts, low self-esteem, and poor impulse control are also considered risk factors related to suicide (8–10). Additionally, having psychological problems is another factor that predisposes an individual to suicide. Depression, alcohol and substance abuse, psychotic disorders, panic disorder, and antisocial, borderline, histrionic, and narcissistic personality disorders are associated with suicide (6,9).

Recent psychological and psychiatric research has started using machine learning techniques due to the complexity of making clinical decisions surrounding suicide (11). Machine learning involves programming computers to improve performance using sample data or past experiences. It encompasses algorithms designed to enhance computer effectiveness (12). Machine learning methods can be categorized into three types: supervised, unsupervised, and reinforcement learning. These methods, based on machine learning, consider the interaction between data units and are employed for classification, diagnosis, and taking preventive measures by making statistical inferences (13,14). Reviewing studies on suicide using machine learning methods, Oh, Yun, Hwang & Chae (2017) (15) estimated suicide rates among 573 participants. The study's findings showed that the model's overall accuracy rate was 93.7% for one month, 90.8% for one year, and 87.4% for a lifetime suicide attempts. Lin, Nagamine, Yang, Tai, Lin & Sato (2020) (16) applied machine learning algorithms such as logistic regression, decision tree, random forest, gradient boosting regression tree, support vector machine, and multi-layer sensor on data from 3,546 military personnel aged 18-50 years. This data included medical examinations, blood tests, and chest x-rays, drawing from the premise that military personnel experience heightened psychological stress and a greater risk of suicide attempts compared to the general population. The accuracy of all six machine learning algorithms for predicting the existence of suicidal ideation was found to be over 98%. Aldhyani, Alsubari, Alshebami, Alkahtani & Ahmed (2022) (17) attempted to detect suicidal ideation through social media posts using a convolutional neural network and a bidirectional long short-term memory (CNN-BILSTM) model, along with the machine learning XGBoost model. According to their research results, the BILSTM model performed better than the XGBoost model, with an accuracy rate of 91.5%. In supervised learning, one of the methods used in this study and rooted in machine learning, algorithms are trained. Models are created using the values of input parameters, referred to as attributes or independent variables, and the labeled data derived from the output parameter values, known as the target variable or dependent variable, which corresponds to these values. Essentially, a relationship is identified between the values of the attributes and the corresponding target variable using a previously observed and known dataset (18). Supervised learning algorithms vary based on the type of dependent variable. Since the dependent variable in this study is categorical, supervised learning classification algorithms will be employed.

In this context, the study aims to predict suicidal risk and behavior, focusing on adolescents and young, more susceptible adults, especially within specific age groups, using supervised learning classification algorithms. This method is a subset of machine learning methods that harness the risk and protective factors associated with suicide. Moreover, regarding warning signals related to suicide and associated risk factors, the study strives to contribute to prevention efforts by thwarting the evolution of potential or existing suicidal thoughts into actions and offers a foundation for future research.

METHOD

Participants were included in the study after obtaining approval from the Haliç University Non-Interventional Clinical Research Ethics Committee on 20.03.2020, under reference number 53. Participants were informed about the research before the scales were administered and that participation was voluntary. The informed consent form and scales were presented to the participants online, and the data were collected via forms created on the Internet.

The study population comprised young adults between the ages of 18-30. For sampling, the convenience sampling method was used. This method is based on collecting data from volunteers who can be easily accessed, keeping the study's objectives in mind (19). Within the scope of the research, a total of 1,940 participants were reached: 199 (10.25%) male and 1,741 (89.75%) female. This sample was strong enough to represent the broader population. The ages of the participants ranged from 18 to 30 ($\bar{x}$=20.48, SD=2.45). The average age of male participants was 21.37±2.73, while the average age for female participants was 20.38±2.39.

Personal Information Form, Coping Attitudes Assessment Scale, and Rosenberg Self-Esteem Scale were used as data collection tools.

Personal Information Form

Developed by the researchers, this form was used to determine the demographic characteristics of the participants. It contained questions about variables such as gender, age, marital status, educational status, and socioeconomic level, along with questions concerning risk and protective factors potentially associated with suicide.

Coping Attitudes Evaluation Scale

This scale, initially developed by Carver, Scheier, and Weintraub in 1989 with 15 sub-dimensions spanning 60 items, was revised by Zuckerman and Gagne in 2003 to feature five factors and 40 items. These factors were named Self-Help, Approach, Accommodation, Avoidance and Self-Punishment. The scale employs a 4-point Likert-type format. Pearson product-moment correlation coefficients, used to assess the relationship between the Turkish and English versions of the scale, registered at r=0.932 in terms of the total score, which is significant at p<0.0001. This indicates a high level of consistency between the two versions, ensuring linguistic equivalence. The Cronbach Alpha Coefficient, denoting the internal consistency of the Coping Attitudes Evaluation Scale, was found to be 0.979. These findings suggest that the Coping Attitudes Assessment Scale is valid and reliable for individuals in Turkiye and can be used in its 5-factor, 32-item form. The highest possible score on the scale is 128, while the lowest is 40. A higher score indicates a greater coping attitude, while a lower score signifies a lesser coping attitude (20).

Rosenberg Self-Esteem Scale (RSE)

This scale, recognized as a reference for measuring self-esteem in research, was developed by Morris Rosenberg in 1963. The reliability studies for the scale were conducted with 5,024 high school students in the USA. In Turkiye, the RSE's reliability and validity studies were undertaken by Cuhadaroglu (1986) (21) on a high school sample group of 205 people. In the Turkish validity and reliability study of the RSE, scores of 0-1 points were categorized as high self-esteem, 2-4 points as medium self-esteem, and 5-6 points as low self-esteem. The RSE has 12 sub-areas, with its first ten items measuring self-esteem. Hence, only the first ten items were used in this study. Positive and negative items are sequenced alternately. Items 1, 2, 4, 6, and 7 are positive, while items 3, 5, 8, 9, and 10 are negatively charged. A low score on the scale indicates high self-esteem, while a high score indicates low self-esteem.

This study used the supervised learning method from the range of machine learning methods. Using input values, algorithms like ExtraTrees, GradientBoosting, CatBoost, XGBoost, Logistic Regression and Voting classification estimated whether individuals had suicidal thoughts within the past year. Additionally, inferences were drawn by calculating accuracy and precision values using the confusion matrix. The programming language used for machine learning is Python 3.9.16, and the platform for coding is Spyder 5.4.2 in Anaconda.

Encoding of Categorical Variables

Most machine learning algorithms require the categorical data in the dataset to be numerical. Numerous methods exist for this transformation, and in this study, all categorical data were converted to numerical form using the Label Encoding method.

Label Encoding

In this technique, each category of data is assigned a unique integer starting from zero, based on its alphabetical order (22).

Data Scaling

Disparities in the units of input parameter measurement can adversely affect a model's success. In this study, the values of input parameters were standardized to mitigate this problem. To standardize, where x is the original data value, the average of the values in the attribute, where the x value was located, was represented as $\bar{x}$, and the standard deviation was σ, this formula was used (22):

$$z = \frac{x - \bar{x}}{\sigma}$$

Model Creation

Model building consists of training the algorithm and testing the model. As such, the dataset is partitioned into training and testing sets. When constructing a model with a training set, the model's performance is gauged using the test set (22,23).

The model's performance is influenced by the choice of the training and test sets. There are different performance evaluation methods for this selection. In this study, both holdout and k-fold cross-validation methods were used.

Holdout

This method divides the dataset into two parts: a one-time training set and a test set. Typically, 2/3 of the dataset is chosen as the training set, and the remaining 1/3 as the test set.

k-Fold Cross Validation

In this method, the dataset is divided into k equal subsets. The closest feasible division is performed if the dataset cannot be evenly divided into k subsets. Each k subset is selected once as the test set, with the remaining k-1 subsets selected as the training set. Model performance is ascertained by averaging the performance measures for k iterations (24).

The Performance of the Model

A confusion matrix was used to measure the classification algorithms' performance in this study (Table 1). Given that the dependent variable in the study encompasses two categories, the matrix used is of size 2x2. The definitions and formulas (25) for this matrix are:

A confusion matrix juxtaposes prediction values on one axis with actual values on the other. This matrix comprises true positives, true negatives, false positives, and false negatives.

If the model correctly classifies an instance of the positive class as positive, it is labeled a true positive. Conversely, if the model incorrectly classifies a negative instance as positive, it is termed a false positive.

Table 1: Confusion matrix

	Prediction (positive)	Prediction (negative)
Actual (positive)	True positive (TP)	False negative (FN)
Actual (negative)	False positive (FP)	True negative (TN)

Similarly, when the model accurately classifies a negative instance as negative, it is deemed a true negative. However, if the model misclassifies a positive instance as negative, it is designated a false negative.

Accuracy: This represents the fraction of correctly classified samples relative to the total sample count. With this metric, model performance is broadly assessed, essentially determining the success of the model.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

Precision: Reflects the accuracy rate of instances that the model classifies within the positive group.

$$Precision = \frac{TP}{TP + FP}$$

Additionally, this study aimed to enhance the model performance by considering the correlation of each independent variable with the dependent variable and employing the backward stepwise model selection method. Initially, in stepwise model selection approach, a model containing all independent variables is computed. Subsequently, these variables are incrementally removed from the model to attempt to improve model performance (18).

RESULTS

The research group comprised 1,940 participants, of which 199 (10.25%) were male, and 1,741 (89.75%) were female. Participant ages ranged from 18 to 30 ($\bar{x}$=20.48, SD=2.45). The average age for male participants was 21.37±2.73, while that of female participants was 20.38±2.39.

It was determined that 542 female participants (31%) and 64 male participants (32%) had suicidal thoughts in the past year. The demographic characteristics of the participants are presented in Table 2.

Table 2: Descriptive statistics of demographic variables

Variable	Group	n	%
Gender	Male	199	10.3
	Female	1,741	89.7
Marital status	Single	1,906	98.3
	Married	34	1.7
Working status	Working	201	10.4
	Not working	1,739	89.6
Socioeconomic level	Lower	215	11.0
	Middle	1,623	83.7
	Upper	102	5.3
Education level	Primary school	1	0.1
	Secondary school	16	0.8
	High school	1,273	65.6
	University	609	31.4
	Postgraduate	41	2.1
Exposure to domestic violence	Yes	659	34.0
	No	1,281	66.0
Witnessing domestic violence	Yes	694	35.8
	No	1,246	64.2
Exposure to violence outside the family	Yes	486	25.1
	No	1,454	74.9
Do you get the support you expect from your family and find your relationships satisfying?	Yes	1	0.1
	No	16	0.8
	Undecided	1,273	65.6
	Partially	609	31.4
Do you see the support you expect from your friends and find your relationships satisfying?	Yes	1	0.1
	No	16	0.8
	Undecided	1,273	65.6
	Partially	609	31.4
Have you had suicidal thoughts at any point in your life?	Yes	1,028	53.0
	No	912	47.0
Have you ever attempted suicide at any point in your life?	Yes	233	12.0
	No	1,707	88.0
Have you had suicidal thoughts in the past year?	Yes	606	31.2
	No	1,334	68.8
Have you attempted suicide in the past year?	Yes	45	2.3
	No	1,895	97.7
Have you ever deliberately engaged in self-harming behavior?	Yes	848	43.7
	No	1,092	56.3
Has anyone in your family attempted suicide?	Yes	184	9.5
	No	1,756	90.5
Have you ever known someone close to you who attempted suicide?	Yes	596	30.7
	No	1,344	69.3

Table 2 (cont): Descriptive statistics of demographic variables

Variable	Group	n	%
Health status	Bad	67	3.5
	Medium	570	29.4
	Good	1,303	67.1
Stress situation	Lower	120	6.2
	Middle	704	36.3
	Upper	1,116	57.5
Do you think you lead a purposeful and meaningful life?	Yes	407	21.0
	No	417	21.5
	Partially	214	11.0
	Undecided	902	46.5
Do you look to the future with hope and enthusiasm?	Yes	504	26.0
	No	443	22.8
	Partially	185	9.5
	Undecided	808	41.6
Total		**1,940**	**100.0**

Evaluation

Initially, the collected data was converted into a data frame of size (1940,117) using the Python programming language. Using the Label Encoding method, all categorical data were made numerically. For this study, given the goal to determine if individuals had suicidal thoughts in the past year, this attribute was shaped into a data frame of size (1940,1) and set as an output parameter. From the original (1940,117) sized dataset, attributes such as "Have you had a suicide attempt at any time in your life?", "Have you had suicidal thoughts at any point in your life?", "Have you attempted suicide in the past year?" were extracted. The resulting data frame of dimensions (1940, 113) was established as input parameters. Since the study's objective is to predict whether individuals had suicidal thoughts in the past year by probing into associated risk and protective factors, these three questions were excluded from input parameters raised for due diligence. The values of these input parameters were then standardized. Models were formulated using algorithms like ExtraTrees, GradientBoosting, CatBoost, XGBoost, and Logistic Regression with the final input and output parameter values. The average success and the standard deviations of these accuracy values were tabulated in Table 3 using the 5-fold cross-validation approach.

After determining the model performance values through 5-fold cross-validation, the dataset was randomly split using the holdout method: 67% of the dataset (or 1,299 observations) was used as the training set, and the remaining 33% (or 641 observations) as the test set. Models were generated using algorithms like ExtraTrees, GradientBoosting, CatBoost, XGBoost, and Logistic Regression based on the training set's input and output parameter values. The accuracy and precision metrics of the models are detailed in Table 4. As all categorical data was previously converted to numerical format via the Label Encoding method, precision values reflecting the model's accuracy are provided in Table 4 for classes 0 and 1. In this context, class 0 corresponds to a 'yes' answer to the question "Have you had suicidal thoughts in the past year?" and class 1 denotes a 'no' response.

Table 3: Model performance values with 5-fold cross-validation method

Average success	Average standard deviation	Algorithm
0.7557	0.0174	ExtraTrees
0.7655	0.0222	GradientBoosting
0.7588	0.0243	XGB
0.7598	0.0206	CatBoost
0.7381	0.0303	LogisticRegression

Since the average accuracy values of the models found with 5-fold cross-validation closely match those derived from the holdout method, subsequent analyses were conducted based on a fixed set of training and tests, chosen by the holdout method. This study aimed to identify individuals who have had suicidal thoughts in the past year using machine learning algorithms. This identification may aid in

Table 4: Model performance values with the holdout method

Accuracy	Precision (0)	Precision (1)	Algorithm
0.75	0.69	0.77	ExtraTrees
0.76	0.68	0.79	GradientBoosting
0.76	0.64	0.80	XGB
0.77	0.68	0.79	CatBoost
0.74	0.62	0.78	LogisticRegression

Table 5: Voting classifier confusion matrix

	Predicted (0)	Predicted (1)
Actual (0)	55	150
Actual (1)	14	422

prevention efforts by addressing potential warning signals and risk factors related to suicide, and curbing future suicidal thoughts or preventing current suicidal ideation from manifesting as actions. Therefore, for this study, the accuracy values of the models and the prediction accuracy rate of observations in class 0 are essential. However, the average accuracy of the model values in Table 4 is 0.76, while the average of the precision (0) values is 0.66. In the next part of the study, studies were carried out to increase the precision value (0), which was relatively low at 0.66. Naturally, the goal was to maintain the average accuracy value of 0.76 as closely as possible. To this end, an ensemble learning model was created combining the Logistic Regression and CatBoost algorithms. "Ensemble learning models are co-learning models that combine different algorithms with different collaboration models to increase prediction and prediction performance and reduce bias and variance problems" (23). This ensemble learning model utilizes the Voting Classifier's Hard Voting approach. In the Hard Voting method, n distinct models are created over the training set, and n different prediction sets are generated using these models for all observations. Then, for an observation, the class label that the majority of the models predicted is taken as the forecast for that class. Here, one model was developed using the CatBoost algorithm, and two others were created by adjusting the threshold values of the Logistic Regression algorithm. Using these three models, an ensemble learning model was established via the Hard Voting approach. Logistic regression is a classification algorithm based on determining to which class a sample likely belongs. Classification in logistic regression is expressed as:

Table 6: Performance values for voting classifier model

Accuracy	Predicted (0)	Predicted (1)
0.74	0.80	0.74

Table 7: Voting classifier confusion matrix (final situation)

	Predicted (0)	Predicted (1)
Actual (0)	59	146
Actual (1)	13	423

Table 8: Performance values for voting classifier model (final situation)

Accuracy	Predicted (0)	Predicted (1)
0.75	0.82	0.74

$$f(z) = \frac{1}{1 + e^{-z}}$$

This is done with the help of the threshold value in the sigmoid function. Typically, this threshold value is set at 0.5 (26). Given that the output parameter values in this study consisted of classes 0 and 1, it was determined that if the f(z) function for the first model, created by the Logistic Regression algorithm, exceeded a threshold value of 0.28, the dependent variable belonged to the class 1; otherwise, it was assigned to class 0. The confusion matrix for the ensemble learning model derived in this manner can be found in Table 5, with the performance values presented in Table 6.

Subsequently, the ensemble learning model's performance was further refined. In doing so, the correlations between each of the independent variables and the dependent variable were taken into consideration. The independent variables "Has anyone in your family attempted suicide?", "Has anyone you know closely attempted suicide?", and "How would you evaluate your health status?" were excluded from the model using the backward stepwise model selection technique. The confusion matrix for the final ensemble learning model is presented in Table 7, and the performance values are in Table 8.

DISCUSSION

This study was conducted to predict suicidal ideation as a precursor to individuals displaying suicidal behavior using supervised learning classification algorithms. This approach is one of the machine learning methods used to discant the risk and protective factors related to suicide.

This study compiled a dataset from 1,940 respondents who completed the Personal Information Form, Coping Attitudes Assessment Scale, and Rosenberg Self-Esteem Scale. The supervised learning method, employing algorithms such as ExtraTrees, GradientBoosting, CatBoost, XGBoost, Logistic Regression and Voting Classification were utilized to create models predicting whether participants had suicidal thoughts in the preceding year. Initially, 5-fold cross validation was implemented, followed by the holdout method for model performance evaluation. Given that the average accuracy values from the 5-fold cross-validation closely resembled those derived from the holdout method, analyses were made based on a consistent set of training and test data chosen by the holdout method. The Confusion Matrix was used to measure model performances, and model comparisons hinged on accuracy and precision values.

Reviewing relevant studies on predicting suicidal ideation via machine learning methods, Amini et al. (2016) (27) compared various prediction methods with 71.9% to 75.2% accuracy rates. Lin et al. (2020) (16) reported that all six machine learning algorithms they employed surpassed 98% accuracy, while Aldhyani et al. (2022) (17) noted between 91.5% and 95%. In our study, the accuracy of the ExtraTrees, GradientBoosting, CatBoost, XGBoost, and Logistic Regression machine learning algorithms oscillated between 74% and 77%. The hit rate for observations classified as having had suicidal thoughts in the last year was 66%.

A model was created using the CatBoost algorithm to enhance this rate and achieve more accurate determinations. Additionally, two other models were developed by adjusting the threshold values of the Logistic Regression algorithm. Combining these three models under the Hard Voting approach yielded an ensemble learning model. This model showed that the hit rate for observations classified as having suicidal thoughts in the last year increased to 82%. This refined ensemble learning model has been identified as the most successful one. Consistent with the research findings, we can infer that predictions of suicidal ideation are in alignment with existing literature.

It was noted that the probability of correctly identifying a "yes" response to the question "Have you had suicidal thoughts in the last year?" was 66%. To increase this 66% rate and make more accurate determinations, a model was established using the CatBoost algorithm. Two models were created by changing the threshold values of the Logistic Regression algorithm, and an ensemble learning model was obtained using the Hard Voting approach with these three models. The performance of the obtained model improved, with an accuracy determined to be 75%. It was observed that 82% answered "yes" to the question, "Have you had suicidal thoughts in the past year?" In line with the research findings, we can deduce that suicidal ideation can be predicted in congruence with the literature. In future research on the topic, it is believed that a balanced number of female and male participants might elevate the accuracy rate estimate. Moreover, the limitations of the study include accessing data through an appropriate sampling method, not using a clinical sample, and evaluating participants irrespective of diagnosis. It is recommended that future research on this topic take these limitations into account. Other limitations include the cross-sectional nature of our research, online data collection, and reliance on self-report scales. In similar research on this topic, including data such as health examinations, blood tests, and chest X-rays led to accuracy rates up to 98%. When questions about the risk and protective factors associated with suicide, along with the results of the analysis, are combined within health institution applications, the detection of suicidal ideation is anticipated to improve, leading to proactive prevention measures. Alternatively, in sessions such as those with psychologists, psychological counselors, or family counselors, where hospital admission is not necessary, and there is no access to procedures like blood tests or chest X-rays, it is believed that suicidal thoughts can be identified primarily through indirect questions about protective and risk factors. Direct questions related to suicide might be avoided to achieve quicker results. Thus, with this determination in place, it will facilitate the referral of individuals to the appropriate units and ensure swift interventions.

While it is notable that studies utilizing machine learning techniques in the mental health field have proliferated in recent years, this study will likely be of value to the mental health sector. It is especially so given the relative scarcity of such studies in Turkiye, and the human eye can discern details that may be overlooked. Individuals in the risk group can be identified. Within this context, the current study might be seen as crucial in contributing to suicide prevention research and the literature and being an interdisciplinary study.

Contribution Categories		Author Initials
Category 1	Concept/Design	B.T.
	Data acquisition	B.T., H.H.T.
	Data analysis/Interpretation	H.H.T.
Category 2	Drafting manuscript	B.T., H.H.T.
	Critical revision of manuscript	B.T., H.H.T.
Category 3	Final approval and accountability	B.T., H.H.T.
Other	Technical or material support	H.H.T.

Ethical Approval: Halic University Non-Interventional Clinical Research Ethics Committee granted approval for this study (date: 20.03.2020, number: 53).

Informed Consent: Informed consent was obtained from all participants.

Peer-review: Externally peer-reviewed.

Conflict of Interest: The authors declare that they have no conflict of interest.

Financial Disclosure: The authors declare that they have no financial support.

REFERENCES

1. Turk Dil Kurumu. İntihar. http://sozluk.gov.tr/. Accessed Dec 31, 2019.
2. World Health Organization. Preventing Suicide. A Global Imperativehttps://apps.who.int/iris/bitstream/handle/10665/131056/9789241564779_eng.pdf;jsessionid=E8D1917991BC49910FB4AC2C725E94F2 ?sequence=1. Accessed Feb 12, 2020
3. Turkiye Istatistik Kurumu. İntihar İstatistikleri 2018. https://tuik.gov.tr/. Accessed Feb 13, 2020.
4. Can SS, Sayil I. Repeated suicides. Kriz Derg 2004; 12:53-62. [Turkish] [CrossRef]
5. Oktik N, Top A, Sezer S, Bozver U. The sociological study of suicide and attempted suicide in the Muğla region. Kriz Derg 2003; 11:1-19. [Turkish]
6. Gurkan B, Dirik G. Predictors of suicide behavior and ideation in university students: Reasons for living and ways of coping. Turk Psychol Articles 2009; 12;58-69. [Turkish]
7. Karatas Z, Celikkaleli O. Suicide probability in emerging adulthood: An analysis in terms of coping with stress, anger and gender. Mersin University J Facul Educ 2018; 14:450-462. [Turkish]
8. Ekici G, Savas HA, Citak S. Psychosocial factors increasing suicide risk (Absence of social security, immigration, and other stressors). Anatol J Psychiatry 2001; 2:204-212. [Turkish]
9. Culha Atesci F, Kuloglu M, Tezcan E, Yildiz M. Axis I and Axis II diagnoses in suicide attempters. Turk J Clin Psychiatry 2002; 5:22-27. [Turkish]
10. Eskin M. Suicidal ideation, attempts and attitudes in youth. Turk J Psychol 2017, 32:93-111. [Turkish]
11. Burke TA, Ammerman BA, Jacobucci R. The use of machine learning in the study of suicidal and non-suicidal self-injurious thoughts and behaviors: A systematic review. J Affect Disord 2019; 245:869-884. [CrossRef]
12. Alpaydin E. Machine Learning. Fourth ed., Istanbul: Bogazici Universitesi Yayınevi, 2018. [Turkish]
13. Uyulan C, Tekin Erguzel T, Tarhan N. The use of deep learning algorithms on EEG based signal analysis. J Neurobehavor Sci 2019; 6:108-124. [Turkish]
14. Yilmaz Aksehirli O, Ankarali H, Aydin D, Saracli O. An alternative approach in medical diagnosis: Support vector machines. Turkiye Klinikleri J Biostat 2013; 5:19-28. [Turkish]
15. Oh J, Yun K, Hwang JH, Chae JH. Classification of suicide attempts through a machine learning algorithm based on multiple systemic psychiatric scales. Front Psychiatry 2017; 8:192. [CrossRef]
16. Lin GM, Nagamine M, Yang SN, Tai YM, Lin C, Sato H. Machine learning based suicide ideation prediction for military personnel. IEEE J Biomed Health Inform 2020; 24:1907-1916. [CrossRef]
17. Aldhyani THH, Alsubari SN, Alshebami AS, Alkahtani H, Ahmed ZAT. Detecting and analyzing suicidal ideation on social media using deep learning and machine learning models. Int J Environ Res Public Health 2022; 19:12635.
18. Arslan I. Data Science with Python. First ed., Istanbul:Pusula 20 Teknoloji ve Yayincilik, 2019. [Turkish]
19. Buyukozturk S, Cakmak-Kilic E, Akgun OE, Karadeniz S, Demirel F. Scientific Research Methods. Eighteenth ed., Ankara:Pegem Akademi, 2014. [Turkish]
20. Dicle AN, Ersanli K. Turkish standardization and the validity and reliability studies of the cope. J Acad Soc Sci 2015; 3:111-126. [Turkish]
21. Cuhadaroglu F. Self-esteem in adolescents. Postgraduate Thesis, Hacettepe University Faculty of Medicine, Ankara, 1986. [Turkish]
22. Celti C. Machine learning methods and some applications. Master Thesis, Halic University Institute of Graduate Education, Istanbul, 2022. [Turkish]
23. Sorhun E. Machine Learning with Python. Istanbul:Abakus Yayincilik, 2021.
24. Balaban EM, Kartal E. Data Mining and Machine Learning. First ed., Istanbul:Caglayan Kitabevi, 2015. [Turkish]
25. Turk B, Tali, HH. Violence prediction on somatization and emotional self awareness with machine learning methods. Bull Legal Med 2020; 25:99-105. [Turkish] [CrossRef]

26. Atalar A. Logistic regression analysis and machine learning applications. Master Thesis, Halic University Institute of Graduate Education, Istanbul, 2022. [Turkish]

27. Amini P, Ahmadinia H, Poorolajal J, Moqaddasi Amiri M. Evaluating the high-risk groups for suicide: A comparison of logistic regression, support vector machine, decision tree and artificial neural network. Iran J Public Health 2016; 45:1179-1187.