Self-Harm Prediction Model Using Machine Learning Technology

Trishala Ahalpara[1], Kalyani Deore[2], Prathamesh Desai[3] and Nida Parkar[4]
[1][2][3] *Department of Computer Engineering, Student,University of Mumbai, Atharva College of Engineering, Malad, Mumbai, India*
[4] *Department of Computer Engineering, Assistant Professor,University of Mumbai, Atharva College of Engineering, Malad, Mumbai, India*

Abstract

Psychological Disorders like self-harm and depression are very common among the people in the age range of 15-30 years. The host category of this age range is mainly the institute going students as it affects their lifestyle and the efficiency of the students to perform well on academic fronts. If the situation persists, they might even commit suicide if they are not diagnosed at an early stage. Machine learning is a powerful tool for predicting such medical situations. Hence the research focuses on predicting whether an institute going student shows any self-harm tendencies. The dataset of 353 students was considered and analyzed for predicting the performance of the techniques used. This research has further applied seven machine learning algorithms and has compared their results on the dataset collected. Out of the seven, the best working algorithm considered on the dataset is the Random Forest Algorithm and hence the model was trained on it. In the model, the researcher's has considered twenty-five attributes out of which it has been reduced to thirteen attributes using random forest classifier feature importance method. Further using Stratified K Fold on the dataset the research has sampled the training data. In the end, fine-tuning the hyperparameters using Grid Search CV the classifier model is trained.

Keywords- *Mental Health, Self-harm, Suicidal, Machine Learning, Data Mining, Random Forest Classifier, Grid Search CV, Feature Importance, Stratified K Folds, Stochastic Gradient Descent (SGD), Logistic Regression, Naïve Bayes Classifier, K Nearest Neighbour (KNN), Decision Tree Classifier, Support Vector Machine (SVM).*

1. INTRODUCTION

Across the world, Suicide/Self-harm is considered one of the most significant issues. Day by day the rate of suicide is increasing steadily. The greatest amounts of people committing suicide are found in the range of 14-25 years. 800,000 people die due to suicide which is approximately equivalent to one person every 40 seconds. Suicide was the second-highest cause of death among the age gap of 15-29 years [11]. The diagnosis of Mental health is a step by step process as the symptoms of mental health are not clear and it just changes the lifestyle of the person so it becomes difficult to consider it. The diagnosis takes place with a set of questionnaire and various sessions with the psychiatrist. The tests are done in order to understand whether the symptoms observed in the individual are due to the mental illness or not. Various types of psychological test are conducted. There are nine types of psychological test that can be considered as the metric used to study psychological disorders. Some of them are Intelligence test, Personality test, Attitude test, Achievement test, Neuropsychological tests, Vocational tests, Direct observation test, Aptitude test, Sexological test, Interest test. Tests like Wechsler Memory Scale, Hamilton Depression Rating Scale, Hamilton Anxiety Scale, Beck Depression Inventory, Schizophrenia Test and Early Psychosis Indicator.

The Mental health issues are considered as insignificant and are left out of the sights of an individual, but on a greater extent these issues are needed to be openly discussed and proper diagnosis should be given to the individuals. Every person once in life faces the issues of depression, self-harm or other psychological disorders due to various problems like family issues, ragging or bullying, sexual abuse and peer pressure in education. Individuals tend to suffer in silence and never openly discuss it with others, at times they are even being ignored or not taken seriously. Discussing such issues with the close ones is one of the important aspects for curing such medical issues because talking about their problems with other people and sharing the issues can help them to understand their problems and take further actions.

The paper broadly focuses on the diagnosis of mental health problem like self-harm, its causes and reasons of students attempting suicide. Analyzing the self-harm tendencies of individuals and providing help in the form of individual motivational videos, articles, quotes and professional help is what the implementation plan focuses upon. The research inhibits the various techniques of machine learning technology that can be applied to the collected dataset to generate the results. The implementation methodology is a step by step process starting with data preprocessing in order to generate the result with maximum accuracy. Further, analysis using various machine learning algorithms like Gaussian Naïve Bayes, Stochastic Gradient Descent, Logistic Regression, KNN, Decision Tree, Random Forest, SVM were used. The results were compared in order to obtain the algorithm which would give precise results. Out of all of these, it has been noted that the Random Forest Classifier best fits with the dataset obtained. Initially, twenty-five attributes were taken into consideration, later on, using the Feature Importance (inbuilt method of random forest classifier) thirteen attributes of utmost importance were generated and the algorithms were applied on dataset considering twenty-five attributes and thirteen attributes respectively. The results and accuracy observed had a slight difference in the twenty-five attributes trained model and thirteen attributes trained model but considering the thirteen attributes rather than considering the twenty-five attributes helps to reduce the cost of the analysis.

2. LITERATURE REVIEW

According to the WHO report, 800,000 people die due to suicide which is approximately equivalent to one person every 40 seconds. The number of psychological problems faced by the students is increasing day by day due to competition in the education, peer pressure, family issues, ragging are few to measure. There are various characteristics to be considered for predicting whether the person is facing any psychological problem or not.

According to the paper 'Prediction of Mental Health Problems among Children', twenty-five attributes were identified and five mental health problems were discussed which are as follows:

- Attention problem
- Anxiety problem
- Attention Deficit Hyperactivity Disorder
- Academic problems
- Pervasive Developmental Disorder (PDD)

The aim was to study such mental health issues and create an analytical model which predicts mental health problems among children using machine learning technology. The paper broadly discusses the parameters that contribute to the children's mental health and further developed a prediction model that can use algorithms such as Multilayer Perceptron, Multiclass Classifier and LAD Tree [1]

The school children also experience behavioural and environmental factors that contribute to their psychological well being. Four classes of well being of school children were identified and two extreme classes of well-being and not well-being were observed. The algorithm used was GMDH (group method of data handling). It predicts on combining the knowledge of network structure and partial description of neurons. The algorithm proceeds in the way that from all the variables, the neurons select the active variables and then the algorithm is used for data mining, prediction and classification tasks. Between the three years of 2015 and 2017, 578 adolescents aged 12-17 years from 18 rural schools were interviewed. The 193 responses from 11 schools formed the basis for the proposed technology. The authors developed a special questionnaire for supervising the risks for reducing the psychological well-being of the pupil of the schools [2]

The mental health of people belonging to IT industry deals with a lot of stress in the company and are more vulnerable to mental health diseases observing severe psychological health issue due to the workload and unending deadlines which in turn make them stay awake overnight and work day and night to meet the deadline on time. In this system, they diagnose the psychological disorder inpatient by comparing the patient's mental health with the DSM-IV-TR. The dataset referred was derived from the survey which was done by setting up a questionnaire. Based on the responses, the conclusion was derived that whether the member of staff needs attention or not. They have used genetic algorithm, classification and machine learning techniques to build a semi-automated system. The future goal is to fully automate the system. The assessment is done by the classifier and the final call will be of analyst for treatment of the patient. People suffering from mental disorders face anxiety disorder which finally develops into depression. They have considered the input for their model and then machine learning psycho-linguistic posts. Various techniques are applied to the model to identify mental health-related features. Their focus was on precursor, cognitive distortion and symptoms of psychological behaviour like anorexia, anxiety and depression. They have applied LIWC (Linguistic Inquiry and Word Count) to extract features and further machine learning is applied. According to research, stress leads to depression, stroke, heart attack, cardiac arrest. They used the EEG signal (electroencephalogram) for analysis where stress is added to the MIST (Montreal Imaging Stress Task). Machine learning with EEG gives feature extraction, selection, classification (Logistic Regression and naïve Bayes. The paper concludes by advising employers to make working environments better for the employees and also to give feedback regularly to the organization [3]

Suicide risk can't be assessed. In the paper of the prediction of suicide in severe mental illness, they have aimed to develop and validate a predictive model for suicide using data with severe mental illness. They have implemented it on Swedish individuals aged 15-65 with a diagnosis of severe mental illness (schizophrenia-spectrum disorders and bipolar disorder) within a fixed time span. Further measuring the discrimination and clinical risk factors using a web-based probability-based risk calculator (Oxford Mental Illness and Suicide tool) without categorical cut-offs. The most important population with high suicide risks with a severe mental illness namely schizophrenia-spectrum disorders and bipolar disorder and it is around 20 times than in the general population. The algorithm used was based on statistical analysis on multivariable logistic regression. The tool is intended to be used in combination with assessing other health and psychosocial needs, which would allow extra discovery of high-risk persons and in conversation between clinicians, patients and careers. [4]

In the paper of predictive Modeling in e-Mental Health, developments in technology using sensor devices and artificial intelligence help to create new opportunities for mental health care. The predictive model is based on three dimensions namely time required to recover, types of available data (like questionnaire data, ecological momentary assessments, smartphones sensor data) and types of clinical decision. Based on these dimensions, four model types are used to classify existing and future researches Promising methodology used for predictive modelling techniques are Decision Trees, Bayesian approaches, Support Vector Machines and Artificial Neural Networks. The four types of model predicted are based on the data. Type 1 models predict the risk of mental illness and can be used to identify the best treatment options. Type 2 models predict short term changes in health status. Type 3 models predict the outcome of the intervention phase. Type 4 models aim to predict relapse [5]

Another paper on Prediction of suicide causes in India uses machine learning technology like Artificial Neural Network and Support Vector Machine to generate the results. The algorithms were used to extract the causes of suicide. The model is not predicting the causes independently for every age group or male and female separately. The data is extracted from various social media website like twitter, weka tool of data mining was used in this paper. The dataset was provided by the National Crime Record Bureau. When the dataset is used with the neural networks it gives 77.5% accuracy and with SVM model it gives 81.5% accuracy in prediction [6]

In the paper 'Use of Machine Learning Algorithm to predict individuals with Suicide Ideation in the general population,' a model is developed which can predict individuals with suicide ideation using a machine learning algorithm. A national survey was done for collecting the dataset of 35116 Korean Adults, out of which 11628 individuals were selected which had an equal number of individuals with suicidal ideation and non-suicidal ideation. A random forest model was trained on the dataset and 15 features were considered with recursive feature elimination via 10 fold cross-validation. All the analysis was done on the 35116 adult individuals using the R language. The prediction model achieved a good performance in the test set and predicted suicide ideators among the total samples with an accuracy of 0.821, sensitivity of 0.836, and specificity of 0.807 [7]

According to another paper, an individual's everyday posts on social media websites like Reddit which contains many clues to predict the future occurrence of mental illness. Hence samples were collected illustrating on posts to groups concentrating on various kinds of mental ailment and conversation groups concentrating on non-mental health topics. Words drawn from such sources could be used to differentiate several kinds of intellectual ailment like ADHD, Anxiety, Bipolar, and Depression. The paper is categorized into two groups' clinical reddits and non-clinical reddits. Further, the model is mainly trained to focus on words used by clinical reddits to predict future health issues or disorders by focusing on words like life stress which provides insight into a developing mental illness. The limitation is that it is not possible to be unambiguous to predict clinical subreddit have been diagnosed with the disorder. But there are words that are not quite understandable by the non-clinical population, initially, that does not indicate that they also don't suffer from mental illness, hence the paper focuses on finding out delicate signals which can help to find the disorders and mental illness amongst individuals [8]

In the paper of 'Machine Learning Framework for Detection of Psychological Disorders at OSN,' a system for psychological disorders detection that has the ability to extract online social behaviour of an individual was created. It uses a machine learning approach on social media websites like facebook, twitter to get a higher accuracy rate. The workflow is to collect the data, clean and preprocess the data and extracting features. It uses the Naïve Bayes classifier method with a precision of 50% and an accuracy of 79.9% [9]

3. OVERVIEW OF MACHINE LEARNING TECHNIQUES

There are many algorithms available today and one can use any of the techniques to find accurate results. In this research, seven algorithms are considered which can give accurate results for small datasets. The following are the details:

1. Gaussian Naive Bayes: This algorithm is mainly used for the classification task and which is derived from the Bayes theorem.

   $$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

   According to the Bayes theorem, we can find the probability of A occurring given that B has already occurred. This states that both the events are independent of one another and the presence of one particular feature does not affect another.

2. Stochastic Gradient Descent (SGD): In gradient descend optimal solutions are found by tweaking the parameters iteratively in order to minimize a cost function. But this increases the computation time hence SGD is used to pick random instances in the training set at every step and computes the gradients based on that single instance.

3. Logistic Regression: Logistic Regression is commonly used to estimate the probability that an instance belongs to a particular class. If the estimated probability is greater than 50% than the model predicts that the instance belongs to that class and otherwise it predicts that it does not.

4. K-Nearest-Neighbour (KNN): In KNN, the training samples are considered and when a test sample(unknown class label) is given, KNN tries to find those neighbouring training samples that are closest to the test sample using various distance formulas like Euclidean distance.

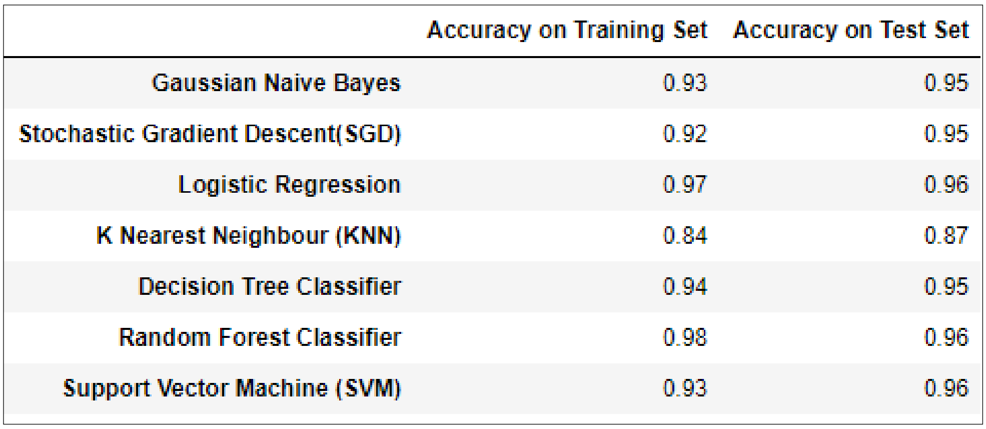
5. Decision Tree Classifier: Decision tree is a popular Machine Learning algorithm that can perform both classification and regression tasks and even multi-output tasks. Decision tree uses

the tree representation to solve the problem in which each leaf node corresponds to a class label and attributes are represented on the internal node of the tree. It uses methods like Gini Index and Information Gain to calculate the feature that contributes to the prediction and generate the decision tree.

6. Random Forest Classifier: The Random Forest is an ensemble of Decision Trees. It introduces extra randomness when growing trees, instead of searching for the best feature when splitting a node; it searches for the best feature among the random subset of features. The algorithm results in greater tree diversity, which trades a higher bias for a lower variance, generally yielding an overall better model.

7. Support Vector Machine (SVM): An SVM model is a representation of the examples as points in space, mapped so that the examples of the separate categories are divided by a clear gap that is as wide as possible. In addition to performing linear classification, SVMs can efficiently perform a non-linear classification, implicitly mapping their inputs into high-dimensional feature spaces.

The following are the results generated of algorithms applied on the collected dataset. As per the results it has been noted that Random Forest Classifier works best on testing dataset with accuracy score of 96% and training dataset with accuracy 98%, followed by Logistic Regression, Support Vector Machine (SVM) and Decision Tree Classifier.



	Accuracy on Training Set	Accuracy on Test Set
Gaussian Naive Bayes	0.93	0.95
Stochastic Gradient Descent(SGD)	0.92	0.95
Logistic Regression	0.97	0.96
K Nearest Neighbour (KNN)	0.84	0.87
Decision Tree Classifier	0.94	0.95
Random Forest Classifier	0.98	0.96
Support Vector Machine (SVM)	0.93	0.96

Figure 1: Accuracy Score of algorithms on collected datasets.

4. IMPLEMENTATION

This section represents the implementation method used for this particular study of predicting self-harm tendencies among institute going students using a predictive approach.

The research broadly focuses on asking questions that can consider important features for our analytical model. Keeping in mind various psychological disorders, personality types and behaviours of the individuals we have designed a well-defined questionnaire suggested by a psychological professional that can actually throw some light to self-harm tendencies among individuals. Further, machine learning technology is used to generate the results.

1) Dataset: In order to understand the roots and causes of self-harm tendencies there are various categories of attributes considered and types of questions asked based on demographic, personal, present and past experiences. The dataset collected for this research is based on the survey conducted among students in the age range of 15 to 30 years old. Following is the brief information about the attributes considered.

A. Demographic Questions Asked:
   1. Age: There are different age groups in dataset from 15 to 30 years old.
   2. Gender: The value in this field is Male, Female or Prefer not to say.
   3. Bodyweight: Bodyweight is entered in kilograms.
   4. Sexuality: Categorical groups like Straight, Bisexual and Other are given.

B. Personal Questions Asked:
   1. Employment type: For students it is Unemployed, Part-time and Full-time
   2. Education Status: Categories like Current Student, Pass out, Repeater and Dropout is used
   3. Number of friends: Having support system like friends talks a lot about a person's psychological health.
   4. Income: There are various Income groups like 0, below 1 lakh, 2 to 5 lakh and 5 lakh above.
   5. Relationship Status: Categories like Single, In-relationship, Married is considered.
   6. Have you ever been into any kind of physical relationship: Virginity Status like virgin and non-virgin are categorised.
   7. How often do you abuse alcohol: Categories of Regularly, Weekly, Yearly and don't drink is considered.
   8. How often do you abuse drugs: Categories of Regularly, Weekly, Yearly and don't do drugs is considered.
   9. Suffering from chronic disease: Having a chronic disease or not.
   10. Family Issues: Suffering from any kind of family issues or not.

C. Present Questions Asked:
   1. Do you have trouble sleeping at night: Yes or no options to find out whether the student is Insomniac or not.
   2. Social Fear: Having a social fear or not.
   3. Communication with opposite sex: Able to interact with opposite genders or not.
   4. Lack of Appetite: Suffering from lack of appetite or not.
   5. Anger Issues: Having Anger issues or not.
   6. Self loathing: Harbouring any kind of self-loathing feelings
   7. Feeling lost in life: It has been noted that people who suffer from psychological disorders like suicide often feel lost in life as they come across being void.
   8. Suffering from trauma: Facing traumatic incidents also affects a person's mental health.
   9. Ragging/Bullying: Ever been victim of ragging or bullying.
   10. Antidepressant: Using any kind of medication.

D. Past Questions Asked:
   1. Self-harm: Plasticising self harm techniques once, few times, prefer not to say or never.

The above questions are the brief description of parameters that is considered for our model. These questions are not asked directly but implicitly to the individual students. Proper rephrasing and implicit questionnaire is used to conduct the survey in the data collection process. As it has been suggested by psychological professionals that direct questions intimidates the individuals.

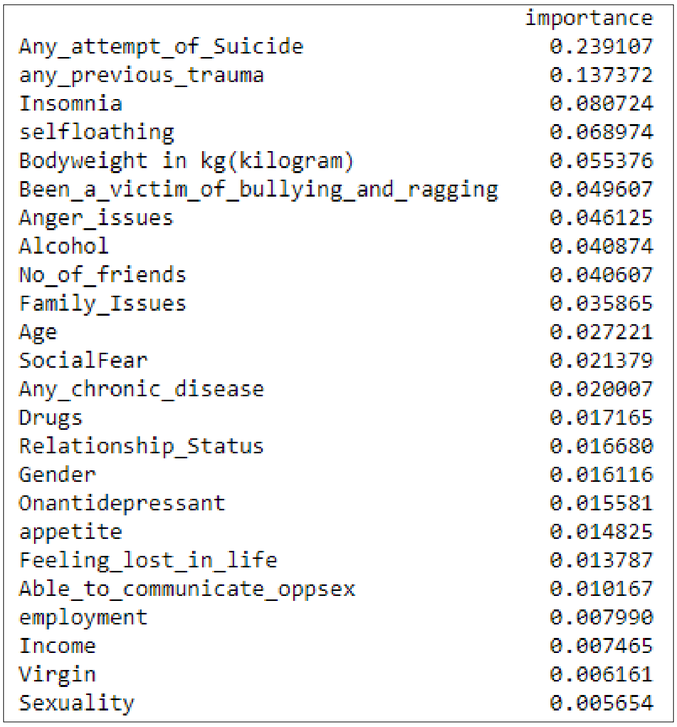
2. Number of Dataset: The research has considered 353 entries in our dataset for our training and testing model and this dataset is collected among institute going students over the periods of two months.

3. Data Pre-processing: Data Pre-processing and Data Cleaning Strategies are used to make the available data efficient and free from error. Data preprocessing methods like data dropping of null values and unwanted columns and replacing the missing values is done.

4. Data Encoding: Data Encoding is the method used to encode the categorical non-numerical data and using this data for analysis of the model. Most of the machine learning algorithm works well with encoded numerical data.

5. Data Splitting: Data Spitting is splitting of the dataset into testing and training set. For the research purpose 40 percent of data is considered as testing data and 60 percent of data is considered as the training data.

6. Feature Importance: Calculating the feature importance of each and every parameters of the dataset based on self harm rate and using feature importance method of random forest classifier is really important step to calculate efficiency of each feature. From the analysis of the data, below mentioned is the feature importance of each and every attribute.



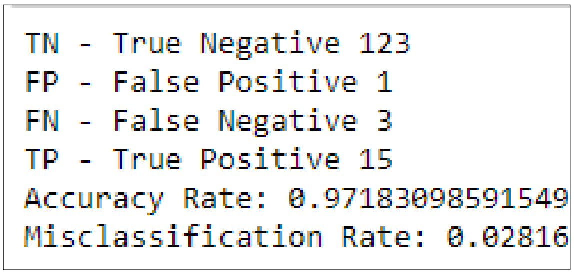
```
                                     importance
Any_attempt_of_Suicide                 0.239107
any_previous_trauma                    0.137372
Insomnia                               0.080724
selfloathing                           0.068974
Bodyweight in kg(kilogram)             0.055376
Been_a_victim_of_bullying_and_ragging  0.049607
Anger_issues                           0.046125
Alcohol                                0.040874
No_of_friends                          0.040607
Family_Issues                          0.035865
Age                                    0.027221
SocialFear                             0.021379
Any_chronic_disease                    0.020007
Drugs                                  0.017165
Relationship_Status                    0.016680
Gender                                 0.016116
Onantidepressant                       0.015581
appetite                               0.014825
Feeling_lost_in_life                   0.013787
Able_to_communicate_oppsex             0.010167
employment                             0.007990
Income                                 0.007465
Virgin                                 0.006161
Sexuality                              0.005654
```

Figure 5: Feature Importance of all parameters considered.

7. Training Model: Setting a threshold value of 0.02 and using select from model 13 features are considered out of 25. These 13 features again are used to train the final classifier that acts as the base classifier for our model.

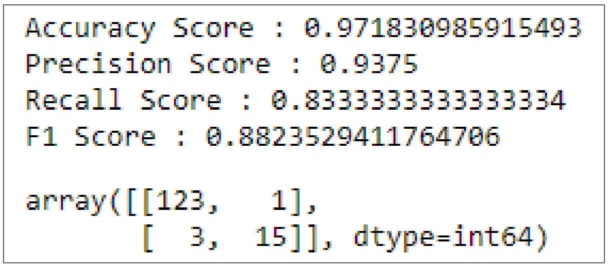
8. Fine-tuning: Next step is to fine-tuning the model with Grid Search CV to generate accurate estimators and hyperparameters for training the model and also using Stratified K fold method to generate various folds to train the model at a different instance each time to avoid recurrence and maintaining the efficiency of the model.

9. Confusion Matrix: Confusion Matrix is used to analyze the performance of the trained model on the test data. The below-mentioned diagram is the confusion matrix of the test data set which were calculated based on below-mentioned equations.



```
TN - True Negative 123
FP - False Positive 1
FN - False Negative 3
TP - True Positive 15
Accuracy Rate: 0.97183098591549
Misclassification Rate: 0.02816
```

Figure 7: TN, FP, FN, TP Rates



```
Accuracy Score : 0.971830985915493
Precision Score : 0.9375
Recall Score : 0.8333333333333334
F1 Score : 0.8823529411764706

array([[123,   1],
       [  3,  15]], dtype=int64)
```

Figure 8: Accuracy Score of Trained Model and Confusion Matrix.

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

$$\textbf{\textit{Recall}} = \frac{TP}{TP + FN}$$

$$\textbf{\textit{F - measure}} = \frac{2*Recall*Precision}{Recall + Precision}$$

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

5. DATA ANALYSIS AND RESULTS

1) Random Forest Tree: A random decision tree is generated to better visualize the dataset and to understand the contribution of the features in making such list of decision trees for the Random Forest Classifier.

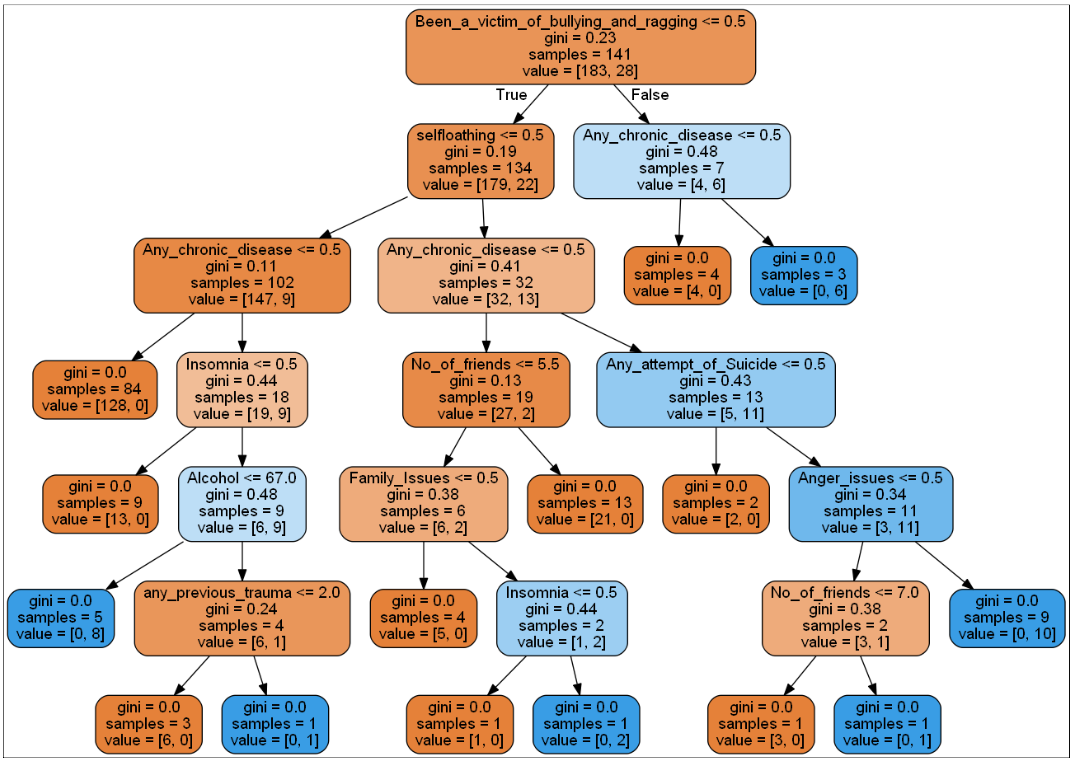


Figure 9: Random Forest Decision Tree representation.

2) Feature Importance: From the feature importance plot graph it has been noted that Any_attempt_suicide, Any_previous_trauma, Insomnia, Self-loathing and Anger issues are the top five attributes contributing to the self-harm tendencies of the students.

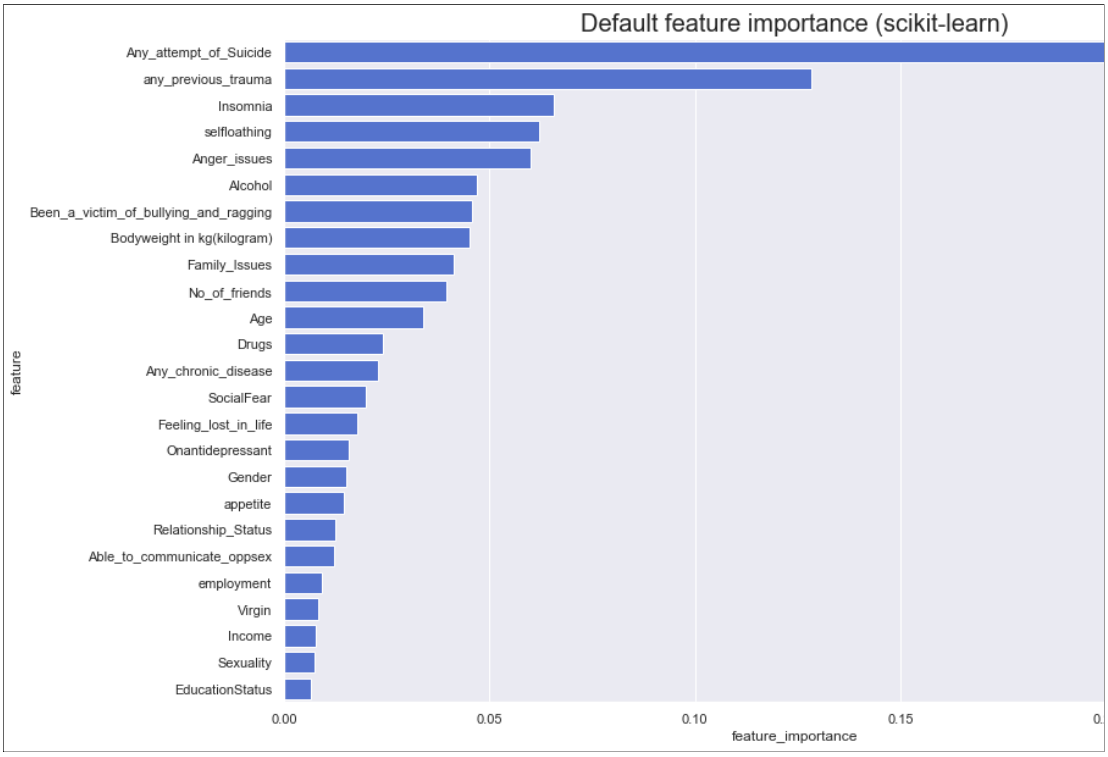
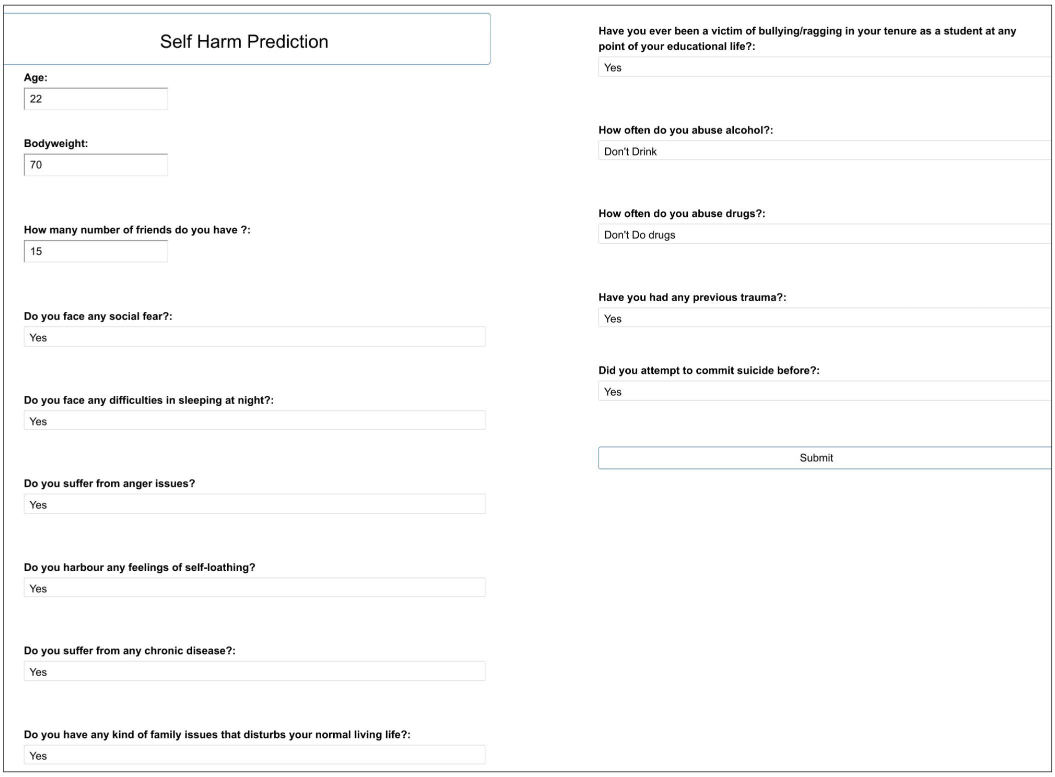


Figure 10: Graphical Representation of Feature Importance.

3) Test Cases:

Results generated for true prediction on Self-harm



Self Harm Prediction

Age:
22

Bodyweight:
70

How many number of friends do you have ?:
15

Do you face any social fear?:
Yes

Do you face any difficulties in sleeping at night?:
Yes

Do you suffer from anger issues?
Yes

Do you harbour any feelings of self-loathing?
Yes

Do you suffer from any chronic disease?:
Yes

Do you have any kind of family issues that disturbs your normal living life?:
Yes

Have you ever been a victim of bullying/ragging in your tenure as a student at any point of your educational life?:
Yes

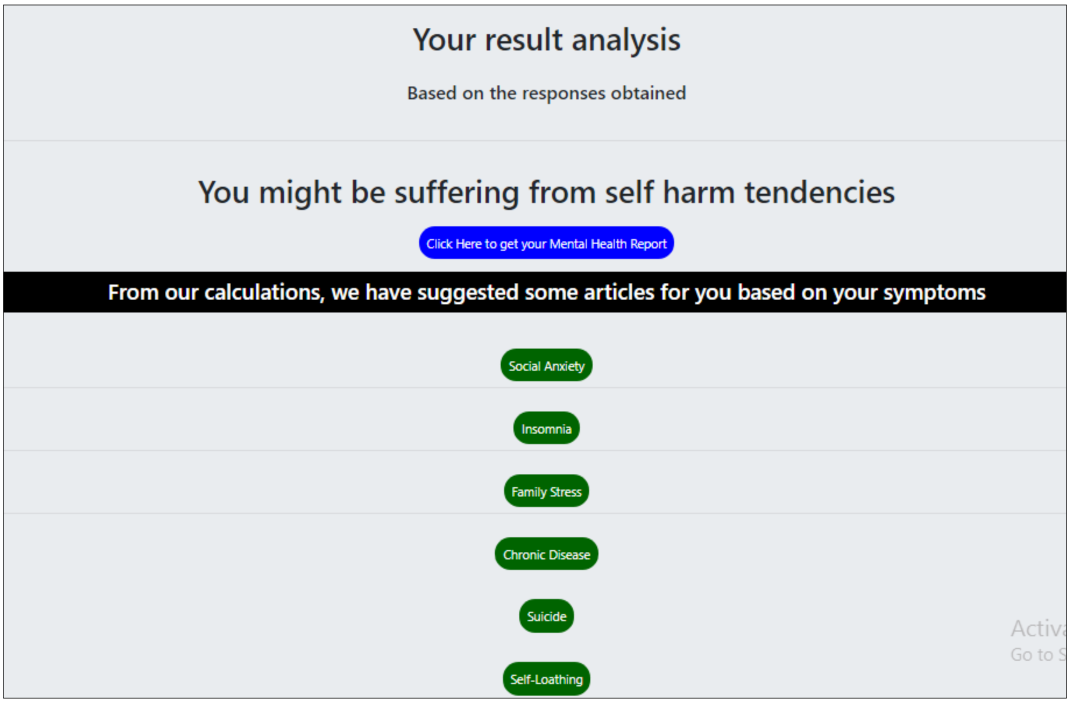
How often do you abuse alcohol?:
Don't Drink

How often do you abuse drugs?:
Don't Do drugs

Have you had any previous trauma?:
Yes

Did you attempt to commit suicide before?:
Yes

Figure 11: Inputs for Self-harm Prediction.



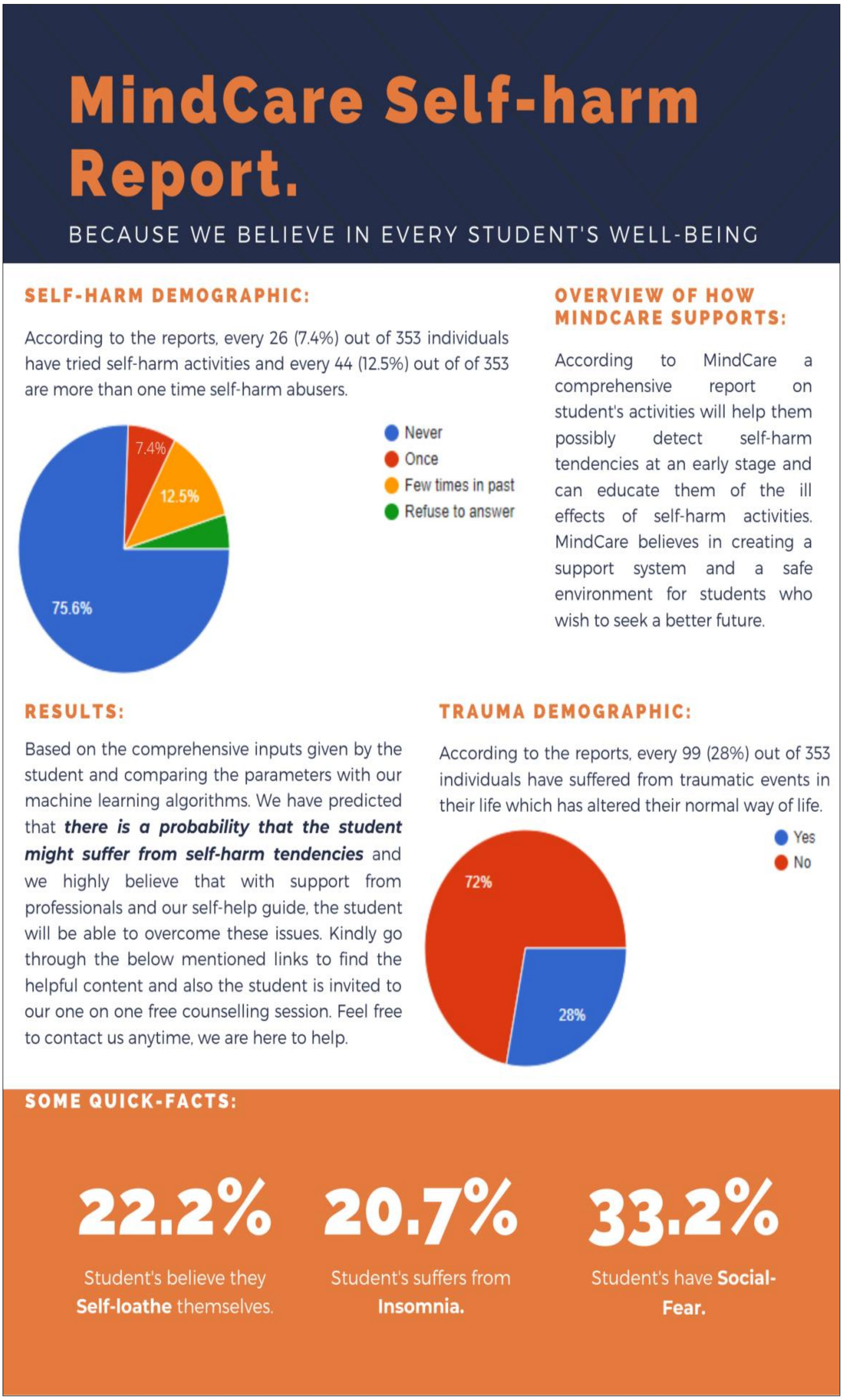
Your result analysis

Based on the responses obtained

You might be suffering from self harm tendencies

From our calculations, we have suggested some articles for you based on your symptoms

Figure 12: Output for Self-harm Prediction.



MindCare Self-harm Report.

BECAUSE WE BELIEVE IN EVERY STUDENT'S WELL-BEING

SELF-HARM DEMOGRAPHIC:

According to the reports, every 26 (7.4%) out of 353 individuals have tried self-harm activities and every 44 (12.5%) out of of 353 are more than one time self-harm abusers.

7.4%
12.5%
75.6%
● Never
● Once
● Few times in past
● Refuse to answer

OVERVIEW OF HOW MINDCARE SUPPORTS:

According to MindCare a comprehensive report on student's activities will help them possibly detect self-harm tendencies at an early stage and can educate them of the ill effects of self-harm activities. MindCare believes in creating a support system and a safe environment for students who wish to seek a better future.

RESULTS:

Based on the comprehensive inputs given by the student and comparing the parameters with our machine learning algorithms. We have predicted that ***there is a probability that the student might suffer from self-harm tendencies*** and we highly believe that with support from professionals and our self-help guide, the student will be able to overcome these issues. Kindly go through the below mentioned links to find the helpful content and also the student is invited to our one on one free counselling session. Feel free to contact us anytime, we are here to help.

TRAUMA DEMOGRAPHIC:

According to the reports, every 99 (28%) out of 353 individuals have suffered from traumatic events in their life which has altered their normal way of life.

SOME QUICK-FACTS:

22.2%

Student's believe they **Self-loathe** themselves.

20.7%

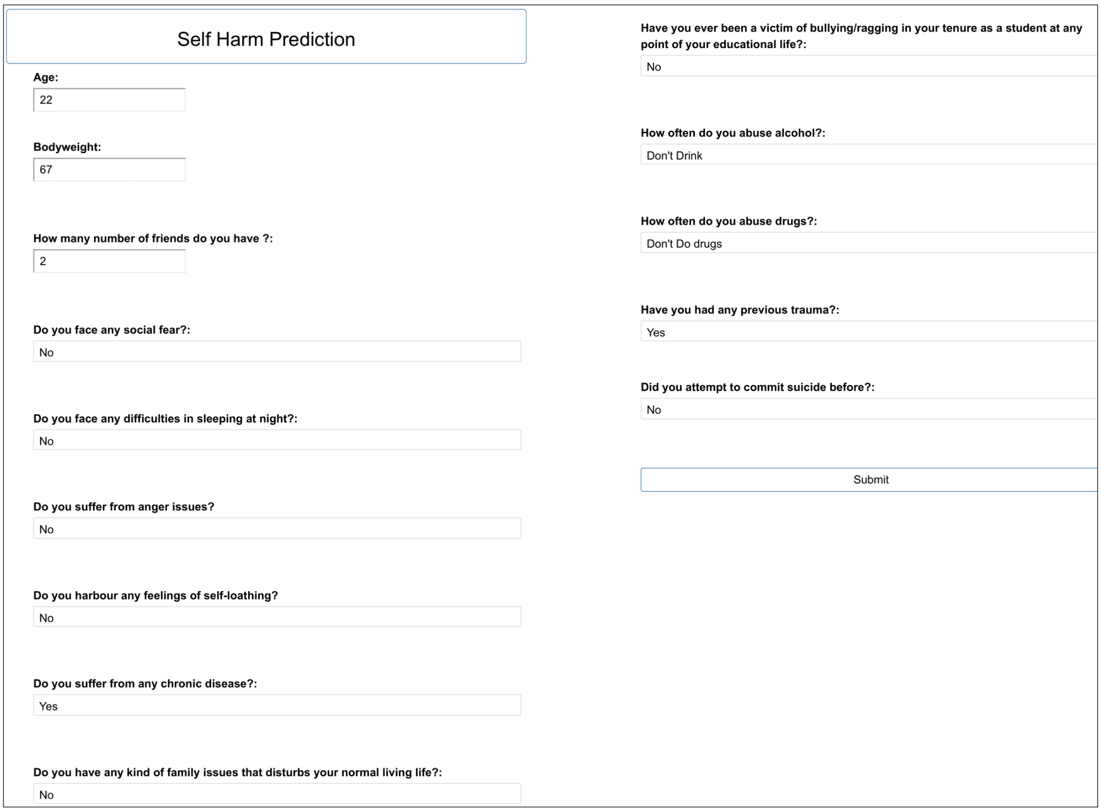
Student's suffers from **Insomnia.**

33.2%

Student's have **Social-Fear.**

Figure 13: Report for Self-harm Prediction.

Results generated for false prediction on Self-harm



Self Harm Prediction

Age:
22

Bodyweight:
67

How many number of friends do you have ?:
2

Do you face any social fear?:
No

Do you face any difficulties in sleeping at night?:
No

Do you suffer from anger issues?
No

Do you harbour any feelings of self-loathing?
No

Do you suffer from any chronic disease?:
Yes

Do you have any kind of family issues that disturbs your normal living life?:
No

Have you ever been a victim of bullying/ragging in your tenure as a student at any point of your educational life?:
No

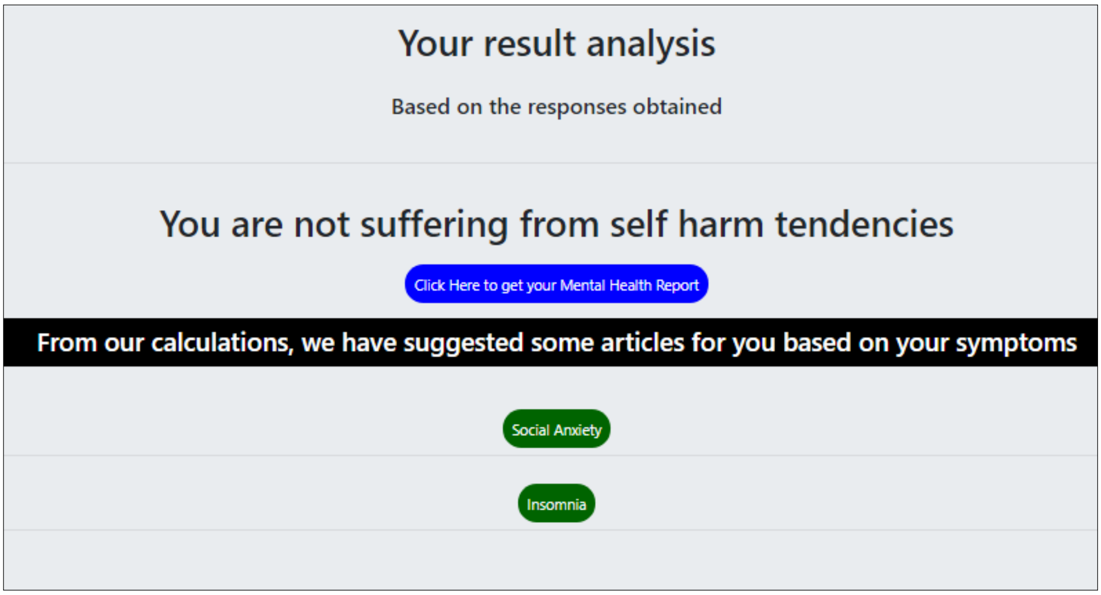
How often do you abuse alcohol?:
Don't Drink

How often do you abuse drugs?:
Don't Do drugs

Have you had any previous trauma?:
Yes

Did you attempt to commit suicide before?:
No

Figure 14: Inputs for not a self-harm prediction.



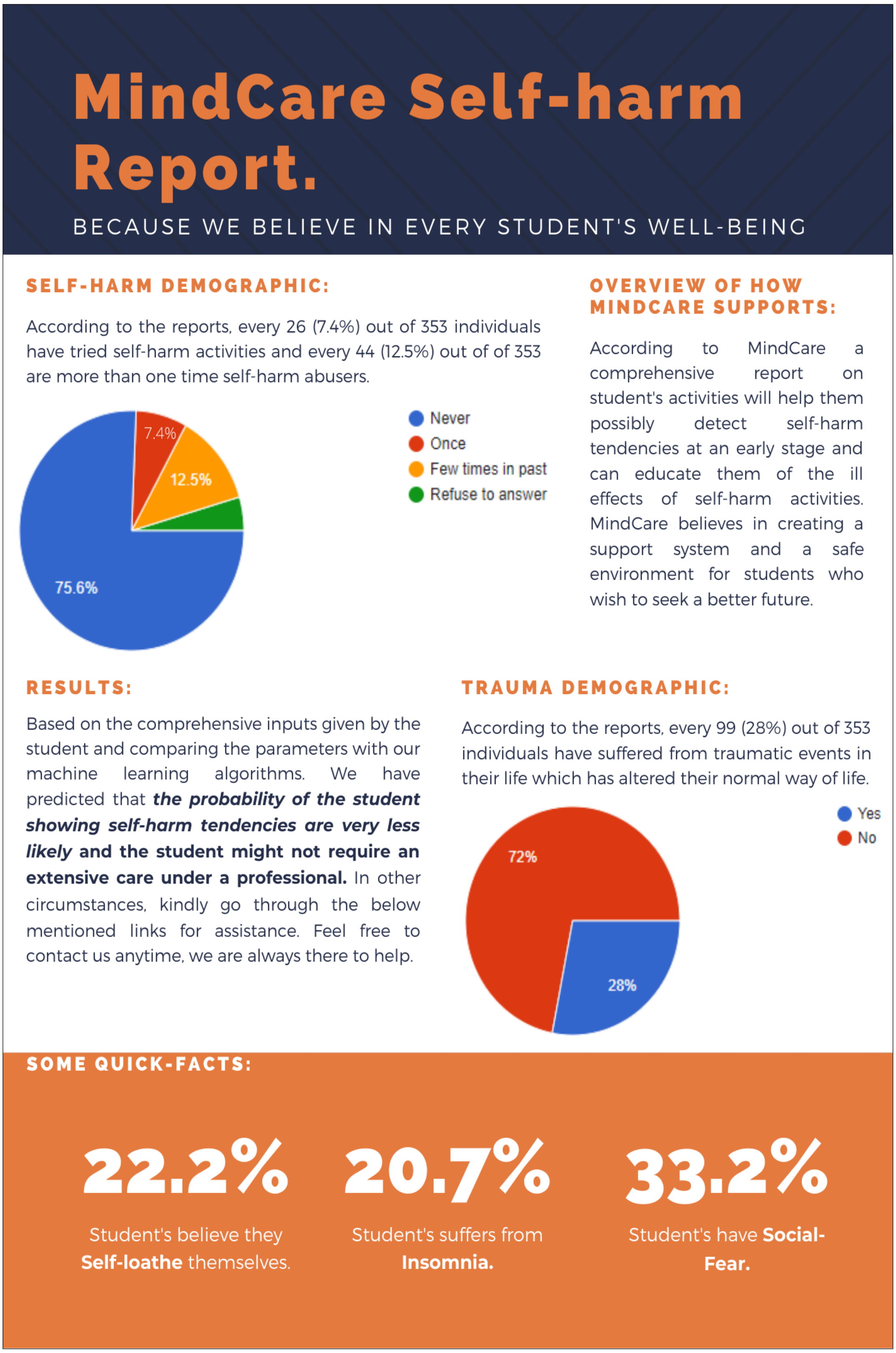
Your result analysis

Based on the responses obtained

You are not suffering from self harm tendencies

From our calculations, we have suggested some articles for you based on your symptoms

Figure 15: Output for not a Self-harm Prediction.



MindCare Self-harm Report.

BECAUSE WE BELIEVE IN EVERY STUDENT'S WELL-BEING

SELF-HARM DEMOGRAPHIC:

According to the reports, every 26 (7.4%) out of 353 individuals have tried self-harm activities and every 44 (12.5%) out of of 353 are more than one time self-harm abusers.

OVERVIEW OF HOW MINDCARE SUPPORTS:

According to MindCare a comprehensive report on student's activities will help them possibly detect self-harm tendencies at an early stage and can educate them of the ill effects of self-harm activities. MindCare believes in creating a support system and a safe environment for students who wish to seek a better future.

RESULTS:

Based on the comprehensive inputs given by the student and comparing the parameters with our machine learning algorithms. We have predicted that ***the probability of the student showing self-harm tendencies are very less likely* and the student might not require an extensive care under a professional.** In other circumstances, kindly go through the below mentioned links for assistance. Feel free to contact us anytime, we are always there to help.

TRAUMA DEMOGRAPHIC:

According to the reports, every 99 (28%) out of 353 individuals have suffered from traumatic events in their life which has altered their normal way of life.

SOME QUICK-FACTS:

22.2%

Student's believe they **Self-loathe** themselves.

20.7%

Student's suffers from **Insomnia.**

33.2%

Student's have **Social-Fear.**

Figure 16: Report for not a Self-harm Prediction.

6. FUTURE WORK

In future, the model can be used for any age group and can be used for the prediction of self-harm tendencies in a particular geographical region, domain, age, gender and can be further utilized for the understanding of the causes of self-harm. It can be further developed and deployed in every institute to know the well-being of students and also to help them in order to cope up with the situation using self-help articles, videos and recommendation. Also, the student can be provided with psychological help if required through interactive web-based software.

Using multi-classification algorithms there is a scope of constructing a machine learning model that can predict various types of psychological disorders. But the limitation faced in constructing such a model is the lack of available database and the complexity of studying the features that not just contribute to one disorder but the one which contributes to multiple disorders.

7. CONCLUSION

In this research, we have studied and analyzed the self-harm tendencies of institute going students using a robust questionnaire prescribed by psychological professional. The dataset collected was analyzed using various algorithms like Gaussian Naives Bayes, Stochastic Gradient Descent, Logistic Regression, Decision Trees, K nearest neighbour (KNN), Support Vector Machine (SVM) and Random Forest Classifier and based on the results generated, random forest classifier was selected as the best training model. Further methods like feature importance, Grid Search CV were used along with Stratified K fold method to predict the accurate results using the best hyperparameters. The trained model results in an accuracy score of 97%, a precision score of 93.75%, F1 score of 88.23% and recall score of 83.33%. With the low false-positive score from the confusion matrix of the model indicates that the trained model is highly accurate in predicting self-harm tendencies among individuals. The only limitation that can be faced in the future of creating such a prediction model is that the collected dataset with biased results. Sometimes models are trained based on a biased dataset which may lead to overfitting conditions. This can be avoided by using methods like regularization, early stopping, ensembling and removing features. Another most effective method is to collect more dataset so that the model is trained on a versatile dataset full of new examples for a machine learning model to learn.

REFERENCES

[1] Ms. Sumathi M.R; Dr.B. Poorna, “Prediction of Mental Health Problems Among children Using Machine Learning Techniques,” IJACSA, 2016.

[2] S.V.Tyulyupo; A.A.Andrakhanov; B.A. Dahieva; A.V.Tyryshkin, “Adolescents Psychological well-being Estimation Based on a Data Mining Algorithm,” IEEE CSIT 2018,11-14 September, 2018.

[3] Sandhya P and Mahek Kantesaria, “Prediction of Mental Disorder for employees in IT Industry,” International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN: 2278-3075, Volume-8 Issue-6S, April 2019.

[4] Seena Fazel; Achim Wolf; Henrik Larsson; Susan Mallett; Thomas R.Fanshawe, “The prediction of suicide in severe mental illness: development and validation of a clinical prediction rule,” Translational Psychiatry 9,Article no:98.

[5] Dennis Becker; Ward van Breda; Burkhardth Funk; Mark Hoogendoorn; Jeroen Ruwaard; Heleen Riper, “Predictive modeling in e-mental health: A common language framework,” Internet Interventions Volume 12, June 2018.

[6] Imran Amin, Sobia Syed, “Prediction of Suicide Causes in India using Machine Learning,” Journal of Independent Studies and Research – Computing Volume 15 Issue 2 July-December 2017.

[7] Seunghyong Ryu, Hyeongrae Lee, Dong-Kyun Lee, and Kyeongwoo Park, “Use of a Machine Learning Algorithm to Predict Individuals with Suicide Ideation in the General Population,” Psychiatry Investigation 2018, Print ISSN 1738-3684 / On-line ISSN 1976-3026.

[8] Thorstad, Robert & Wolff, Phillip. (2019). Predicting Future Mental Illness from Social Media: a Big Data Approach. 10.31234/osf.io/arf4t.

[9] Punam B. Nalinde, Anita Shinde, “Machine Learning Framework for Detection of Psychological Disorders at OSN,” International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN: 2278-3075, Volume-8 Issue-11, September 2019.