

**APPLYING MACHINE LEARNING TECHNIQUE TO UNDERSTAND THE
SUICIDAL BEHAVIOR IN THE CONTEXT OF BANGLADESH AGE GROUP 12
TO 35**

BY

MD. SHOHANUR RAHMAN

ID: 172-15-10052

HRIDOY SHOM

ID: 172-15-9980

NABIL HAYAT

ID: 172-15-10017

This Report Presented in Partial Fulfillment of the Requirements for the Degree
of Bachelor of Science in Computer Science and Engineering

Supervised By

Zerin Nasrin Tumpa

Lecturer

Department of CSE

Daffodil International University



**DAFFODIL INTERNATIONAL UNIVERSITY
DHAKA, BANGLADESH
SEPTEMBER 2021**

APPROVAL

This Project titled “**Applying Machine Learning technique to understand the suicidal behavior in the context of Bangladesh age group 12 to 35**”, submitted by Md. Shohanur Rahman, Id- 172-15-10052 , Hridoy Shom, Id- 172-15-9980 and Nabil Hayat, Id- 172-15-10017 to the Department of Computer Science and Engineering, Daffodil International University, has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of B.Sc. in Computer Science and Engineering and approved as to its style and contents. The presentation has been held on 9th September 2021.

BOARD OF EXAMINERS

Chairman



Dr. Touhid Bhuiyan
Professor and Head

Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

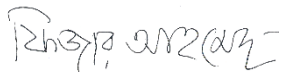
Internal Examiner



Nazmun Nessa Moon
Assistant Professor

Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Internal Examiner



Dr. Fizar Ahmed
Assistant Professor

Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University



External Examiner

Dr. Md Arshad Ali

Associate Professor

Department of Computer Science and Engineering
Hajee Mohammad Danesh Science and Technology
University

DECLARATION

We hereby declare that, this project has been done by us under the supervision of **Zerin Nasrin Tumpa, Lecturer, Department of CSE** Daffodil International University. We also declare that neither this project nor any part of this project has been submitted elsewhere for award of any degree or diploma.

Supervised by:



Zerin Nasrin Tumpa

Designation

Department of CSE

Daffodil International University

Submitted by:

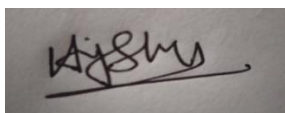


Md. Shohanur Rahman

ID: -172-15-10052

Department of CSE

Daffodil International University



Hridoy Shom

ID: -172-15-9980

Department of CSE

Daffodil International University

Nabil Hayat

Nabil Hayat

ID: -172-15-10017

Department of CSE

Daffodil International University

ACKNOWLEDGEMENT

First we express our heartiest thanks and gratefulness to almighty God for His divine blessing makes us possible to complete the final year project/internship successfully.

We really grateful and wish our profound our indebtedness to **Zerin Nasrin Tumpa**, Lecturer, Department of CSE Daffodil International University, Dhaka. Deep Knowledge & keen interest of our supervisor in the field of “*Machine Learning*” to carry out this project. His endless patience ,scholarly guidance ,continual encouragement , constant and energetic supervision, constructive criticism , valuable advice ,reading many inferior draft and correcting them at all stage have made it possible to complete this project.

We would like to express our heartiest gratitude to Zerin Nasrin Tumpa, Lecturer and Dr. Touhid Bhuiyan , Professor and Head, Department of CSE, for his kind help to finish our project and also to other faculty member and the staff of CSE department of Daffodil International University.

We would like to thank our entire course mate in Daffodil International University, who took part in this discuss while completing the course work.

Finally, we must acknowledge with due respect the constant support and patients of our parents.

Abstract

Suicide is a neglected preventable public health problem across the globe and Bangladesh is not an exception. Suicide can happen at any stage of life, it is the second most frequent and in some countries it is the leading cause of death among people aged 12–35 years. Suicide happens because of multi factorial involvement such as genetic, psychological, social, and cultural risk factors. The purpose of this recent study was to calculate the percentage of committing suicide based on the behavior of a person. A certain age group of people's information is being used to get the result. By understanding several Machine Learning algorithms are being used to develop a model that gives the result. Specific and standard questionnaires were asked to get the data from approximately one thousand people via offline and online. The data was then analyzed and preprocessed. Then seven Machine Learning Algorithms were applied. The result that showing a person's individual percentage of committing suicide and it has a correlation with Depression, Stress and Sexual Harassment.

TABLE OF CONTENT

CONTENTS	PAGE
Board of examiners	I
Declaration	iii
Acknowledgments	iv
Abstract	v
Table of Contents	vi
List of figures	viii
List of Tables	ix

CHAPTER	PAGE
CHAPTER 1: INTRODUCTION	1-4
1.1 Introduction	1
1.2 Motivation	2
1.3 Rationale of the Study	2
1.4 Research Questions	3
1.5 Expected Output	3
1.6 Report Layout	4
CHAPTER 2: BACKGROUND	5-10
2.1 Introduction	5
2.2 Related Works	6
2.3 Literature Review	7
2.4 Research Summary	9
2.5 Scope of the Problem	10
2.6 Challenges	10
CHAPTER 3: RESEARCH METHODOLOGY	11-15
3.1 Introduction	11
3.2 Research Subject and Instrumentation	11
3.3 Ethics	12
3.4 Work Methodology	12
3.5 Data Collection Procedure	13
3.6 Data Preprocessing and Cleaning	14
3.7 Data Test and Training	14
3.8 Use Algorithm	14

CHAPTER 4: EXPERIMENTAL RESULTS AND DISCUSSION	16-24
4.1 Introduction	16
4.2 Experimental Results	16
4.3 Descriptive Analysis	20
CHAPTER 5: SUMMARY, CONCLUSION, RECOMMENDATION and IMPLICATION FOR FUTURE RESEARCH	25-26
5.1 Summary of the Study	25
5.2 Conclusions	25
5.3 Recommendations	25
5.4 Implication for Future Study	26
REFERENCES	27-28
APPENDICES	29-32

LISTS OF FIGURES

FIGURES	PAGE
Figure 1.5.1 Expected Output	3
Figure 3.4.1 Work Methodology	12
Figure 3.5.1 Data Collection Procedure	13
Figure 3.8.1 Used library	15
Figure 4.2.1 Result Using Random Forest Regression	16
Figure 4.2.2 Result Using Linear Regression	17
Figure 4.2.3 Result Using Decision Tree Regression	17
Figure 4.2.4 Result Using Lasso Regression	17
Figure 4.2.5 Result Using Elastic net Regression	18
Figure 4.2.6 Result Using Elastic net CV Regression	18
Figure 4.2.7 Result Using Ridge Regression	18
Figure 4.3.1 Gender Ratio	20
Figure 4.3.2 Age Group	21
Figure 4.3.3 Sexual Harassment	21
Figure 4.3.4 Percentage of people feeling Depressed	22
Figure 4.3.5 Percentage of people feeling Stressed	23
Figure 4.3.6 Percentage of people feeling Happy	23

LISTS OF TABLES

TABLES	PAGE
Table 2.2.1 A comparative study	6
Table 4.2.1 Accuracy of algorithms	19
Table 4.3.1 Gender distribution	20

CHAPTER 1

INTRODUCTION

1.1 Introduction

Suicide is the act of intentionally causing one's own death. It is a way for people to get rid of suffering or pain. People who committed suicide are defined 'died by suicide.' When someone tries to end his own life, but somehow failed to make that happen, then it is called a 'suicide attempt'. According to a study, approximately 49% of suicide victims were between the ages of 20 and 35, and among them 57% were women. Suicide is a major crisis for families, relatives, and friends, and it is seen negatively almost everywhere around the world.

Suicide is a neglected preventable public health problem across the globe and Bangladesh is not an exception. Suicide is a common cause of unnatural death and a long-standing social concern in Bangladesh. Each year, Bangladeshis account for 2.06% of all suicide deaths globally. According to Bangladesh, the rate of suicides is increasing rapidly. In 2017, 11,095 people in Bangladesh committed suicide, implying that 30 people commit suicide per day [1]. Suicide claims the lives of approximately one million people each year around the world, with a global mortality rate of 10.05 per 100,000 and 128.08 per 100,000 in, Bangladesh.

Although suicide can happen at any stage of life, it is the second most frequent, and in some countries, it is the leading cause of death among people aged 12–35 years. Suicide happens because of multifactorial involvement such as genetic, psychological, social, and cultural risk factors. A survey shows that more people in Bangladesh died by committing suicide rather than Covid-19 from last year. Approximately the death rate is 70% more than from Covid-19. Mental health therapy and aid, particularly for young people, are lacking in Bangladesh [2]. We have seen that the most common methods used for suicide are different from one country to another. The most common methods of suicide are pesticide poisoning, firearms, and hanging.

We aim to develop a model applying machine learning algorithms for predicting suicidal behavior in the context of Bangladesh age group 12-35. This model will increase awareness among the people about suicide and also we think it will help us to reduce the rate of suicide in Bangladesh.

1.2 Motivation

Suicide affects people of different nationalities, ethnicities, religions, genders, and socio-economic groups all across the world. Other innate features, such as mental illnesses and birth defects, can raise a person's susceptibility for depression, whether it's a one-time incident or a long-term condition. The huge rise in the suicide rate has been one of the most talked-about subjects. Suicide poses a significant and growing threat to native countries, especially young people.

To reduce the number of suicide deaths, countries must address several common underlying issues that combine to make someone more prone to choose suicide as an outlet. Depression rates are one important element to consider, but there are others to consider as well.

1. Make people aware of the bad impact of suicide.
2. Aware people of suicidal myth and fact.
3. Increase suicide resistance and bring people together in the fight against suicide.

1.3 Rationale of the Study

The effect of our exploration on suicidal behavior has a long haul impact on the. Our research territories are-

1. Developing a system that can help to reduce the rate of suicidal death.
2. Understand and predict the danger of suicide worldwide, suicide is a global problem and it is increasing rapidly.
3. Allow people to recognize the warning signs and take action as soon as possible to prevent suicide.

1.4 Research Questions

Some questions emerged in our mind as we were researching about Suicidal Behavior. When we try to find the answers of these question we came to know a lots of things about suicide. After getting so many information the similar time of work is not very common in our country that we are going to implement through our model. The questions are as follows:

1. Why do people attempt suicide?
2. What are the most common methods of suicide?
3. Why do men complete suicide more often than women do?
4. Is suicide related to impulsiveness?
5. What biological factors increase the risk for suicide?
6. Can the risk for suicide be inherited?
7. Is it possible to predict suicide?
8. Do people attempt suicide to prove something or to get sympathy?
9. Why do people attempt suicide when they appear to feel better?
10. Is there such a thing as 'rational' suicide?

1.5 Expected Output

This research is about developing a model for predicting suicidal behavior is being tried using some different Algorithms of Machine Learning. And also we aim to make people more conscious and make people more aware of suicide and its behavior so that people can early get a chance to recover from their suicidal tendency. Our expected output is a probability of a person committing suicide. From that a sample of our expected output is given below:

43.15	18.975	72.	37.5	7.525	48.2	2.675	32.3	17.4	26.625
6.95	10.525	46.05	32.325	21.05	8.75	30.5	45.	48.45	3.375
15.85	26.2	8.775	3.575	25.55	60.65	6.3	46.15	32.05	16.675
27.55	15.6	23.2	34.1	36.25	63.95	3.025	16.5	45.5	65.6
32.225	22.5	7.5	27.45	22.5	39.95	6.825	39.65	19.025	8.525]

Figure: 1.5.1 Expected Output

1.6 Report Layout

Chapter 1: Introduction

In this section, we discussed our purpose for conducting this research, as well as the study's rationale, research questions, and expected outcomes.

Chapter 2: Background

We will explain the related work of our research and comparative studies between our developed model and current relevant models, as well as the problems and obstacles that we have encountered, in this chapter.

Chapter 3: Research Methodology

We'll talk about our research technique in this chapter, which means we'll talk about how we conducted our research and the terminology we used.

Chapter 4: Experimental results and Discussion

In this chapter, we will discuss about our research experimental results and outputs. We will talk about which results we have found by implementing the algorithms of machine learning over our collected data.

Chapter 5: Summary, Conclusion, Recommendation and Implication for Future Research

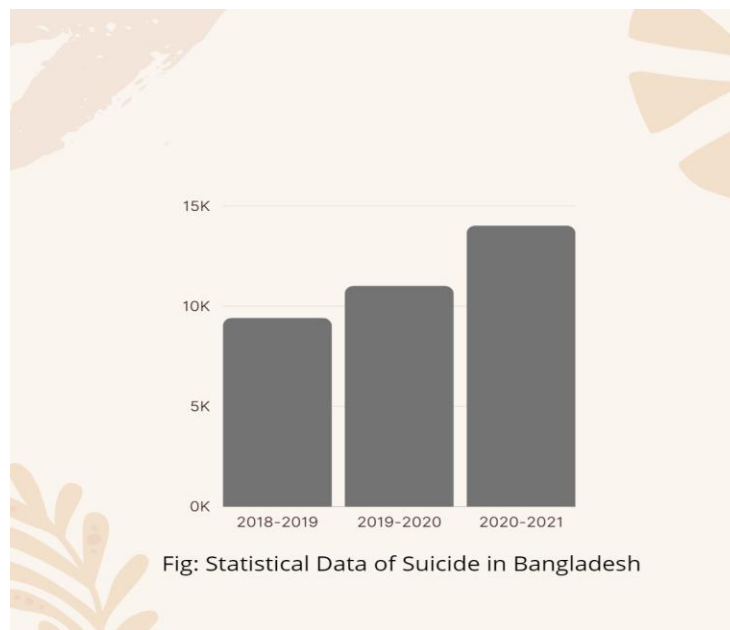
In this chapter, we will discuss about summary, conclusion, recommendation and scope for further development of our research.

Chapter 2

Background

2.1 Introduction

Machine learning is concerned with the creation of algorithms that can learn and generate statistical models for data analysis and prediction. Without being particularly trained for a task, ML algorithms should be able to learn on their own and generate correct predictions based on the data presented. As a result, we applied machine learning algorithms to make suicidal behavior predictions. Suicide has become one of the major social problems all across the world. Bangladesh is also facing this problem. Every day we are losing so many young lives. The suicide mortality rate is an indicator of Sustainable Development Goal, which is to reduce premature death from non-communicable diseases by one-third by 2030 through prevention and treatment, while also promoting mental health and well-being [3].



Source: The Daily Star

2.2 Related Works

Table 2.2.1 A comparative study

Research	Data Collection Process	Applied Technique
Predicting Depression in Bangladeshi Undergraduates using Machine Learning[4]	Online Survey	K-Nearest Neighbor , Random Forest, Support vector machine
Prevalence and Predisposing Factors of Suicidal Ideation Among the University Students in Bangladesh: A Single-Site Survey[5]	Paper and Pencil Survey	Logistic Regression
SUICIDAL IDEATION IN RELATION TO DEPRESSION, LONELINESS AND HOPELESSNESS AMONG UNIVERSITY STUDENTS[6]	Data were collected from 10 different departments under to faculty of DU	Regression Analysis
Risk factors of suicidality among married adults: A cross-sectional survey in Rajshahi City, Bangladesh[7]	A self-developed semi-structured questionnaire was used for data collection in this study.	Logistic Regression
Depression Detection from Bangla Facebook Status using Machine Learning Approach[8]	Using social Media Information	Multinomial Naïve Bayes, Logistic Regression, Linear SVC,SGDC Classifier

Alcohol-Related Risk of Suicidal Ideation, Suicide Attempt, and Completed Suicide: A Meta-Analysis[9]	Data were collected using a self-report questionnaire completed by 17,530 students.	Logistic regression
Gender differences in suicide-related communication of young suicide victims [10]	Interview data from a psychological autopsy study of 35 youths who died by suicide in the Netherlands in 2017 were analyzed.	Qualitative analyses were performed to examine explicit SRC throughout the youths' lives and implicit SRC during the last months prior to suicide.
Declines in the Lethality of Suicide Attempts Explain the Decline in Suicide Deaths in Australia [11]	National data on suicide deaths and deliberate self-harm for the period 1994–2007 were obtained from the Australian Institute of Health and Welfare.	Negative binomial regression was used to estimate the size and significance of method-specific time-trends in attempts and lethality.

2.3 Literature Review

This paper gives an overview of the current literature on suicide in hospitals with two particularly important subjects in mind: long-term and short-term suicide projections. As clinical practice dominates the concept of a short-term suicide risk, problems relating to dynamic risk factors are highlighted [12].

There is sufficient proof that AUD increases the risk of suicide, suicide attempt and completed suicide significantly. AUD can therefore be regarded as an important predictor of suicide and a major premature death source [9].

The boys and girls unique SRC patterns posed corresponding challenges to the next family for interpretation of communications and an adequate response to SRC. The early debut of the SRC

for girls underlines the importance of girls early screening and prevention work, and the late beginning and ambiguity in SRC for boys and kids encourages them to be clear about the idea of a suicide or their intention [10].

Here review and meta-analysis will provide the first complete systematic overview to sum up current literature on the suicide risk in sleep apnea patients. This study will help to evaluate the high-risk suicide population and propose an intervention plan [13].

The remarkable decline in suicide mortality in Australia since 1997, in particular in the exhaust and suspension attempts by motor vehicles [11].

In order to evaluate smartphone addiction, the six-point Smartphone Application-Based Addiction (SABAS; Csibi et al. 2018). The size scale uses the size of six to measure each element (1 = strongly opposed to 6 = very strongly supported) and includes items like "My smartphone was my most important thing in life" (or friends). The larger the score, the higher the risk of developing smartphone apps. Good psychometric and internal consistency were previously found (Chen et al. 2019; Lin et al. 2019). In the current study Cronbach's alpha was good (0.72) [5].

This research focuses on getting important insights into why university students in Bangladesh, in particular undergraduate students, suffer from depression, in addition to predicting depression in university students with a view to recommending their psychiatrist. The data were gathered in cooperation with psychologists, consultants and professors. The survey was conducted. After three algorithms were used to train and test the data set, the optimal way to prevent depression in Bangladesh grads was determined [4].

The aim of this study was to evaluate the association between suicide ideation and sadness, alone and desperation among university students. The sample of the study includes 112 students (51 men and 61 women) selected using the purposeful technique of sampling. Standard questionnaires were used to acquire data from cross-sectional survey participants [6].

The mood illness is depression. This mental disease is a quiet murderer. The way people feel, think and act is significantly affected. Every year, many people commit suicide because of depression throughout the world [8].

The incidence of bullying is as follows: "If a kid or group of youngsters say something terrible and unpleasant about him or her, it is bullying that will. It's also harassing when a kid's tingling again and again doesn't like him or her [14].

A. Wadood and G. Hossain also noted the major risk for suicides as a suicidal ideation. This study has been carried out by the relevant government and non-governmental bodies to identify vulnerable groupings and individuals and to take actions to prevent and minimize suicide in the country[7].

For this research suicidal behavior is being predicted for reducing the rate of suicidal death in our country Bangladesh. If we can predict early suicidal behavior then it will be easier to reduce the rate of suicide. To begin, we gathered information through questionnaires and the internet. To prepare and clean the acquired data, data preprocessing techniques were applied. When the data was provided to the algorithm for prediction, it was generalized and output was generated by the algorithm.

2.4 Research Summary

It is being attempted to construct a model for predicting suicide behavior like depression, which will be shown in future years using various and effective machine learning algorithms. There have been some similar efforts that have used various types of algorithms on various forms of collected data and information. From the Table 2.2.1, we can see that someone used some data from surveying, self-developed semi-structured questionnaire, and social media information to predict different kinds of problems related to suicide. There are no such projects similar to ours. In this research suicidal behavior is being used in our data set and from there we try to predict the probability of committing suicide. For this purpose, we have used several machine learning regression algorithms such as Linear Regression, Random Forest Regressor, Decision Tree Regressor, Lasso, Ridge, Elastic Net, and Elastic Net CV. After that, we used Mean Absolute Error,

Mean Squared Error, Root Mean Squared Error to find out the error rate in Machine Learning model.

2.5 Scope of the Problem

There is always the possibility of an issue in any research. As a result, there is no exception in our work. Our gathered data should be relevant, and our data set should be easy to work with. While working with algorithms, we must correctly train our gathered data. If we don't perform this correctly, our output will be inappropriate and unrepresentative of a real-life scenario. We must work appropriately and attentively if we are to decrease the problem.

2.6 Challenges

There is no way to complete work without encountering difficulties. Challenges might arise at any point during the process of accomplishing any job. The primary problem we faced in our research was obtaining data about suicide from people. That is to say, the most difficult aspect of our research is gathering data from reliable sources and analyzing the results. From a short distance, proper and appropriate data must be collected.

CHAPTER 3

RESEARCH METHODOLOGY

3.1 Introduction

This section outlines the investigation's research methodology. It contains details about the members, including the test's inclusion criteria, the participants, and how they were tested. We describe the examination structure, as well as the rationale behind it, that was chosen with the research's final goal in mind. The equipment that was used to collect data is also displayed, as well as the approach that was used to conduct this investigation. The methods used to break down the data are also discussed. Finally, the requirements for implementation that were followed throughout the process are highlighted.

3.2 Research Subject and Instrumentation

We are researching the prediction of suicidal behavior. We are doing this using different machine learning algorithms. So our research subject is - Applying Machine Learning techniques to understand suicidal behavior in the context of Bangladesh age group 12 to 35.

A survey is a list of questions designed to separate explicit information from a specific group of people in a human subject's research. Telephone, mail, the internet, and even face-to-face surveys are used to do research. Surveys are used to advance knowledge in domains such as social research and demography. Structured questionnaires were created and pre-tested for our study. It was utilized as a research tool when it was completed. As a result, surveys/questionnaires are the most important tool in our research work, which can also be referred to as a research instrument. Following the survey, certain preprocessing techniques were used to generalize the acquired data. The preprocessed data was then prepared for further analysis. Then the prepared data has been used in machine learning (ML) algorithms for predicting suicidal behavior.

3.3 Ethics

Participants were advised that all their data would be kept anonymous and confidential, and information on the nature and purpose of the study, procedure was provided to them and the right to withdraw your information. The detailed study procedure, participatory rights, data protection and confidentiality were established under the 1975 Helsinki Declaration before they could access the questions.

3.4 Work Methodology

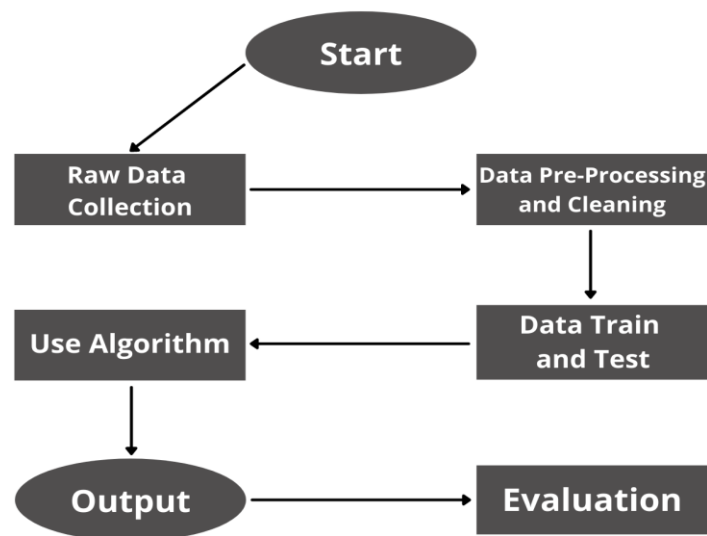


Figure 3.4.1 Work Methodology

3.5 Data Collection Procedure

A systematic procedure of obtaining observations or measurements is known as data collection. Data collecting, also known as information gathering, is the process of gathering and analyzing data on elements of interest in a structured manner that enables one to react to stated research questions, test hypotheses, and assess outcomes. All fields of study, including physical and sociologies, humanities, business, and so on, rely on the information gathering aspect of research.

We have faced so many challenges while collecting data. As there are many ways to collect data but we have collected our data in two different ways. They are given below:

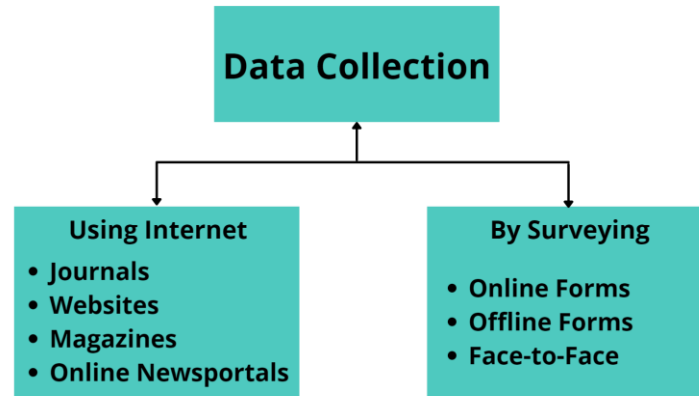


Figure 3.5.1 Data Collection Procedure

Internet

For enhancing and expanding our dataset, we used the internet, including websites, publications, journals, and online newspapers.

Surveys

To acquire data and information, we have surveyed over more than 500 people. We produced a question sheet with all of the research questions we would require. The question paper inquired about age, gender, educational qualifications, and a few other characteristics. We can add their details to our dataset by collecting their information.

3.6 Data Preprocessing and Cleaning

We have cleansed the data to remove as many outliers as possible before using our prediction algorithms on the data, to enhance our accuracy. First, we examined whether any of the data rows were missing. Each question must be completed to continue to the next step in order to finish our survey. However, some of the questions for some students were not answered due to technical problems with the Google survey system. We therefore removed all the rows that lacked replies.

3.7 Data Test and Training

Generally, when a machine learning algorithm split to fit in a model, the ratio of the data remains 25-30 for test purpose and rest for training purpose.

Having this in mind, we split our whole data set at a ratio of 70 to 30. Where 70 percent of the data is used for training a model while remaining 30 percent of the data utilized for testing.

3.8 Use Algorithm

When it comes to algorithms, we should add that seven algorithms have been used to predict Suicidal attempting Percentage. The algorithms are Linear Regression, Random Forest Regressor, Decision Tree Regressor, Lasso, Ridge, Elastic Net, and Elastic Net CV. We employ a variety of ways to differentiate the mean absolute error, root mean absolute error, and root means square error. We'll need to import certain library functions into our IDE to complete all of these tasks. Jupyter notebook has been used as our main working platform for this suicidal behavior prediction. Which libraries have been used is given below-

```
import pandas as pd
import numpy as np
import statistics
from sklearn import metrics
from sklearn.ensemble import RandomForestRegressor
from sklearn.linear_model import LinearRegression
from sklearn.tree import DecisionTreeRegressor
from sklearn.linear_model import Lasso
from sklearn.linear_model import Ridge
from sklearn.linear_model import ElasticNet, ElasticNetCV
from sklearn.model_selection import train_test_split
from pandas import Series, DataFrame
```

Figure 3.8.1 Used Library

In our project, we have used pandas and NumPy libraries. For data frames, Pandas is utilized. Sklearn.model selection is used to import the models or algorithms that we wish to utilize, as well as the train test split function, which is used to train and test the dataset.

CHAPTER 4

EXPERIMENTAL RESULTS AND DISCUSSION

4.1 Introduction

The prescient intensity of the determining model surveyed by Linear Regression, Random Forest Regressor, Decision Tree Regressor, Lasso, Ridge, Elastic Net, and Elastic Net CV is improved by including new parameters for determining the probability of suicidal behavior. We have collected data from numerous places so that our model could give us the most accurate result. Suicide was found to be strongly linked to depression and loneliness. As a result, the suicide percentage rate predicted by this model will aid in the early detection of suicidal behavior. The prescient intensity of defining suicidal behavior prediction models is essentially improved by combining these two elements in the model.

Our model is capable of predicting current-day suicidal behavior. Early warning is critical in this issue, as it allows people to seek mental help from others and preserve their lives.

4.2 Experimental Results

Here are some results that we got after we ran our model and some discussion on it given below.

The output of Random Forest Regression:

[43.15 19.025 71.45 37.5 7.4]

Figure 4.2.1 Result using Random Forest Regression

Using Random Forest Regressor we have predicted the percentage of committing suicide. Here, we can see the first person's rate of committing suicide is 43.15 percent.

The output of Linear Regression:

```
[44.66755966 26.00477221 65.06026905 14.6284923 17.91906851]
```

Figure 4.2.2 Result using Linear Regression

We also use Linear regression to predict the rate of committing suicide and its results are a little bit different from the previous one. Here, we can see the first person's rate of committing suicide is 44.66755966 percent.

The output of Decision Tree Regression:

```
[45. 20. 75. 37.5 7.5]
```

Figure 4.2.3 Result using Decision Tree Regression

We have also used Decision Tree regression to predict the rate of committing suicide. Here, we can see the first person's rate of committing suicide is 45 percent which is quite similar to the Linear Regressor.

The output of Lasso Regression:

```
[39.16878404 28.84452805 63.51261855 4.32688623 20.99075894]
```

Figure 4.2.4 Result using Lasso Regression

We have also used Lasso regression to predict the rate of committing suicide. Here, we can see the first person's rate of committing suicide is 39.16878404 percent which has got the lowest percentage among the last four algorithms we have used.

The output of Elastic net Regression:

```
[44.62517598 26.04107204 65.03491404 14.54755837 17.96493028]
```

Figure 4.2.5 Result using Elastic net Regression

We have also used Elastic net Regression to predict the rate of committing suicide. Here, we can see the first person's rate of committing suicide is 44.62517598 percent which is quite similar to the Linear Regressor.

The output of Elastic net CV Regression:

```
[38.26165703 29.53636185 62.32869572 4.07646244 21.5740584 ]
```

Figure 4.2.6 Result using Elastic net CV Regression

We have also used Elastic net CV Regression to predict the rate of committing suicide. Here, we can see the first person's rate of committing suicide is 38.26165703 percent which is quite similar to Lasso Regressor.

The output of Ridge Regression:

```
[38.90576085 29.30071767 62.45659308 5.13242252 21.35456171]
```

Figure 4.2.7 Result using Ridge Regression

We have also used Decision Tree regression to predict the rate of committing suicide. Here, we can see the first person's rate of committing suicide is 38.90576085 percent which is quite similar to the Elastic net CV Regression and Lasso Regression.

The error of Algorithms:

Table 4.2.1 Accuracy of Algorithms

Algorithm	Error
Random Forest Regressor	Mean Absolute Error: 2.181290849673202
	Mean Squared Error: 11.611803513071896
	Root Mean Squared Error: 3.4076096479896134
Linear Regressor	Mean Absolute Error: 12.353946289678158
	Mean Squared Error: 226.4277634979771
	Root Mean Squared Error: 15.04751685488264
Decision Tree Regressor	Mean Absolute Error: 2.3284313725490198
	Mean Squared Error: 18.52532679738562
	Root Mean Squared Error: 4.304105806945923
Lasso Regressor	Mean Absolute Error: 12.341110250119206
	Mean Squared Error: 226.31152676524064
	Root Mean Squared Error: 15.043654036345048
Ridge Regressor	Mean Absolute Error: 12.352715633809753
	Mean Squared Error: 226.3529477900315
	Root Mean Squared Error: 15.045030667633466
Elastic net Regressor	Mean Absolute Error: 12.364365940851611
	Mean Squared Error: 227.8455011926725
	Root Mean Squared Error: 15.094552036833438
Elastic net CV Regressor	Mean Absolute Error: 12.342333695026126
	Mean Squared Error: 226.39163283652277
	Root Mean Squared Error: 15.046316254702438

Here we can see that in the Random Forest Regressor the value of mean absolute error, Mean Squared Error, and Root Mean Squared Error is less than all others algorithms. So we can say that this Random Forest Regressor algorithm gives us the most accurate result.

4.3 Descriptive Analysis

Following our survey, we discovered some Descriptive Analysis statistics data that revealed some differentiation, as shown below:

Gender Ratio:

Gender is a very important feature in our model. We tried to figure out male or female who are more likely willing to share their information with us.

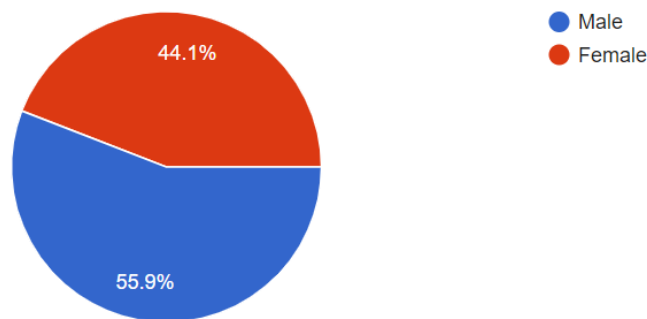


Figure 4.3.1 Gender ratio

The figure 3.4.1 shows that there is 44.1 percent of female and 55.9 percent of male shared their information with us.

Table 4.3.1 Gender distribution

Gender	Percentage %
Male	56.7%
Female	44.1%

Age Group

We also tried to get the information of a particular age group of people. We are implementing our model on this age group. So it was really an important task for us.

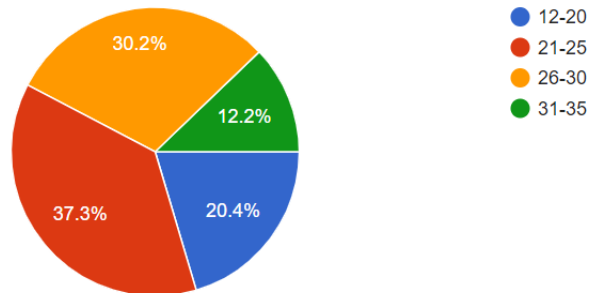


Figure 4.3.2 Age Group

From this figure, we can say that the most people are from the age group of 21-25 who shared their information with us. Then the age group 26-30 standing in second position holding the percentage of 30.2. Age group 12-20 holding 20.4 percent and age group 31-35 holding 12.2 percent.

Sexual Harassment

We can say that sexual harassment is an important issue that has a huge impact on suicide. From our dataset we also got some information where people are being harassed sexually.

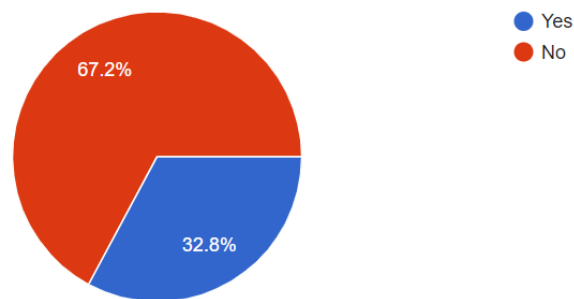


Figure 4.3.3 Sexual Harassment

From the above figure, we can see that 32.8% of People have faced sexual harassment and 67.2% did not. This feature helps us in a very effective way to make our model efficient.

Depression

Depression is one of the best feature we selected to make our model more effective. Depression leads a person to do insane activates and suicide is one of them. So we picked depression as our core feature to predict the result. It helps us to make a more accurate model.

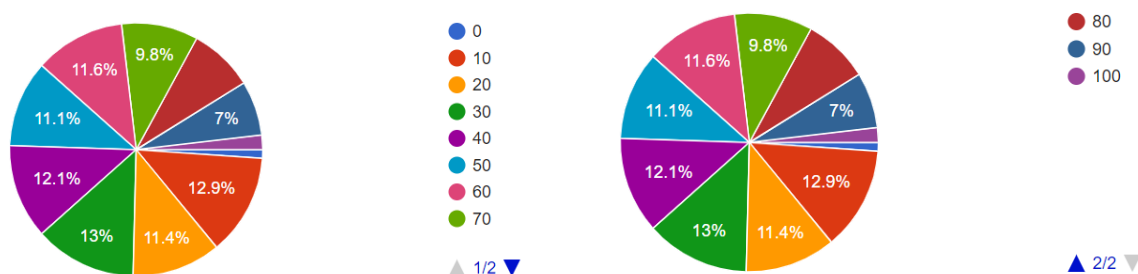


Figure 4.3.4 Percentage of People Feeling Depressed

We have also collected data based on Depression and we can see here that 8.2% of people feel they are 80% depressed in their life and 7% of people feel they are 90% depressed in their life. Depression plays a vital role in committing suicide. So in our model, we used these data as a feature for predicting the percentage of committing suicide.

Stress

Stress is a kind of mental thing that every person having in their life. Sometimes this stress can be very dangerous. People who are having stress at a very high level at the end they got depressed. So we can say stress leads people to get depressed. And depression leads people to commit suicide. So we decided to take stress as aver important feature to measure the percentage of committing suicide.

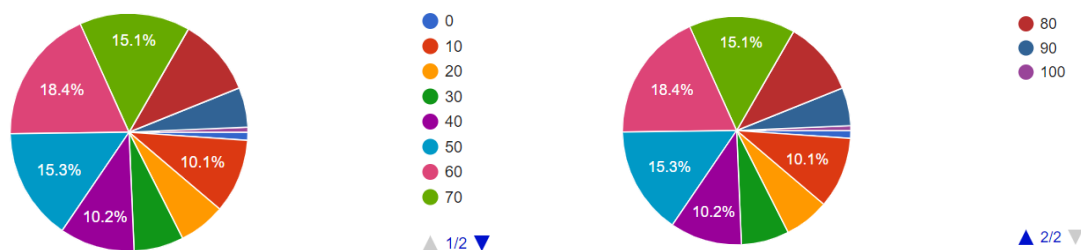


Figure 4.3.5 Percentage of People Feeling Stressed

We have also collected data based on Stress and we can see here that 10.6% of people feel they are 80% stressed in their life and 5.4% of people feel they are 90% stressed in their life. Stress also plays a vital role in committing suicide. So in our model, we also used these data as a feature for predicting the percentage of committing suicide.

Happiness

When a person leads a happy life then their stress level also decreased. We took happiness percentage of a person so that we can evaluate that and fit it in our model to calculate the committing suicidal percentage. A person who is not happy in life he or she definitely will be in stress. So we think being happy in life also makes a person to not think about any insane activities like suicide. In our model happiness is a core feature that we used to predict that percentage of committing suicide.

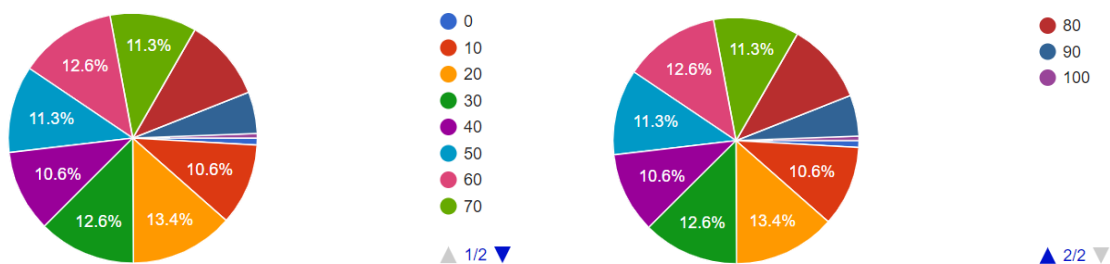


Figure 4.3.6 Percentage of People Feeling Happy

We have also collected data based on Happiness and we can see here that 10.7% of people feel they are 80% happy in their life and 5.4% of people feel they are 90% happy in their life. Happiness plays a vital role for reducing suicide. So in our model, we used these data as a feature for predicting the percentage of committing suicide.

CHAPTER 5

SUMMARY, CONCLUSIONS, RECOMMENDATIONS, AND IMPLICATIONS FOR FUTURE RESEARCH

5.1 Summary of the Study

Suicide is one of the most common social issues in recent days. Bangladesh is having a very serious effect on this problem. So many lives are getting lost every year. So we have to make people more conscious of this social issue. How can we make people more conscious about suicide, this question arose when we started our research? Lastly, we concluded that we could work with the suicidal behavior that would be so much beneficial for the people of society. So we develop a model for detecting suicidal behavior using different types of machine learning algorithms Linear Regression, Random Forest Regressor, Decision Tree Regressor, Lasso, Ridge, Elastic Net and Elastic Net CV have been used for predicting the suicidal behavior over our collected data and our dataset. The most difficult aspect of our research was gathering data.

5.2 Conclusions

Suicide rates are currently at an all-time high, indicating that suicide is one of our country's most serious issues. When depression and loneliness are included as factors, the model becomes more accurate. Suicide is a type of problem for which no cure is available. If we could acquire an early warning about someone committing suicide, it would help us minimize the suicide rate. This is very similar to the findings of our study. We created this model for early suicide prediction, which could help to lower the suicide rate.

5.3 Recommendations

Machine Learning (ML) is the logical analysis of computations and factual models that computer systems use to carry out a certain task without using explicit instructions, relying instead on instances and guesswork. Artificial intelligence is seen as a subset of it. Machine Learning

computations construct a numerical model based on test data, referred to as training data, to determine expectations or choices without being explicitly updated to carry out the operation. Our model for predicting suicide behavior was developed using seven different machine learning algorithms, to determine which is better in terms of prediction across our collected dataset. The various algorithms are as follows: Linear Regression, Random Forest Regressor, Decision Tree Regressor, Lasso, Ridge, Elastic Net, and Elastic Net CV. While researching our dataset, we discovered the mean absolute error, root mean squared error, and mean squared error from several techniques. By comparing the values of the error we get from the above algorithms we can recommend you Random Forest regressor algorithm which gives us the most accurate result.

5.4 Implication for Further Study

Using several machine learning methods, a model for predicting suicidal behavior rate was constructed in this study. However, we have a different strategy in mind for the future. For example, we will strive to construct a website and application in the future that will use Machine Learning algorithms to anticipate the behavior of a person who may commit suicide. I believe it will be more beneficial for individuals to discover the true cause of suicide and to take the necessary safeguards to avoid suicide.

References

- [1] A. R. Roney, "Suicide on the rise in Bangladesh | Dhaka Tribune," *Dhaka Tribune*, 2018. <https://www.dhakatribune.com/bangladesh/nation/2018/03/27/suicide-rise-bangladesh> (accessed Sep. 05, 2021).
- [2] S. N. Sakib, "Bangladesh: Suicide claims more lives than coronavirus," 2021. <https://www.aa.com.tr/en/asia-pacific/bangladesh-suicide-claims-more-lives-than-coronavirus/2175200> (accessed Sep. 05, 2021).
- [3] W. H. Organization, "Suicide," *WHO*, 2021. <https://www.who.int/news-room/fact-sheets/detail/suicide> (accessed Sep. 05, 2021).
- [4] A. A. Choudhury, M. R. H. Khan, N. Z. Nahim, S. R. Tulon, S. Islam, and A. Chakrabarty, "Predicting Depression in Bangladeshi Undergraduates using Machine Learning," *Proc. 2019 IEEE Reg. 10 Symp. TENSYP 2019*, pp. 789–794, Jun. 2019, doi: 10.1109/TENSYP46218.2019.8971369.
- [5] M. A. Mamun, I. Rayhan, K. Akter, and M. D. Griffiths, "Prevalence and Predisposing Factors of Suicidal Ideation Among the University Students in Bangladesh: A Single-Site Survey," *Int. J. Ment. Heal. Addict.* 2020, pp. 1–14, Oct. 2020, doi: 10.1007/S11469-020-00403-Z.
- [6] M. Pervin, N. F.-D. U. journal of Biological, and U. 2016, "Suicidal ideation in relation to depression, loneliness and hopelessness among university students," *banglajol.info*, 2016, Accessed: Sep. 05, 2021. [Online]. Available: <https://www.banglajol.info/index.php/DUJBS/article/view/28495>.
- [7] A. Wadood, R. Karim, A. A. M. Hussain, M. Rana, and G. Hossain, "Risk factors of suicidality among married adults: A cross-sectional survey in Rajshahi City, Bangladesh," *PLoS One*, vol. 16, no. 5, p. e0251717, May 2021, doi: 10.1371/JOURNAL.PONE.0251717.
- [8] M. Billah and E. Hassan, "Depression Detection from Bangla Facebook Status using Machine Learning Approach," *Int. J. Comput. Appl.*, vol. 178, no. 43, pp. 9–14, Aug. 2019, doi: 10.5120/IJCA2019919314.
- [9] N. Darvishi, M. Farhadi, T. Haghtalab, and J. Poorolajal, "Alcohol-Related Risk of Suicidal Ideation, Suicide Attempt, and Completed Suicide: A Meta-Analysis," *PLoS One*, vol. 10, no. 5, p. e0126870, May 2015, doi: 10.1371/JOURNAL.PONE.0126870.
- [10] E. Balt *et al.*, "Gender differences in suicide-related communication of young suicide victims," *PLoS One*, vol. 16, no. 5, p. e0252028, May 2021, doi: 10.1371/JOURNAL.PONE.0252028.
- [11] M. J. Spittal, J. Pirkis, M. Miller, and D. M. Studdert, "Declines in the Lethality of Suicide Attempts Explain the Decline in Suicide Deaths in Australia," *PLoS One*, vol. 7, no. 9, p. e44565, Sep. 2012, doi: 10.1371/JOURNAL.PONE.0044565.
- [12] C. Cassells, B. Paterson, D. Dowding, and R. Morrison, "Long- and Short-Term Risk Factors in the Prediction of Inpatient Suicide," <http://dx.doi.org/10.1027/0227-5910.26.2.53>, vol. 26, no. 2, pp. 53–63, Jan. 2005, doi: 10.1027/0227-5910.26.2.53.
- [13] Y. Yang, A. E. Ssentongo, Y. Pan, M. Ciarletta, V. M. Chinchilli, and P. Ssentongo, "Risk of suicidal ideation, suicide attempts, and suicide deaths in persons with sleep apnea: Protocol for a systematic review and meta-analysis," *PLoS One*, vol. 15, no. 7, p. e0235379, Jul. 2020, doi: 10.1371/JOURNAL.PONE.0235379.
- [14] S. Fujikawa *et al.*, "The association of current violence from adult family members with adolescent bullying involvement and suicidal feelings," *PLoS One*, vol. 11, no. 10, Oct. 2016, doi: 10.1371/JOURNAL.PONE.0163707.

APPENDICES

Appendix

```
import pandas as pd
import numpy as np
import statistics
from sklearn import metrics
from sklearn.ensemble import RandomForestRegressor
from sklearn.linear_model import LinearRegression
from sklearn.tree import DecisionTreeRegressor
from sklearn.linear_model import Lasso
from sklearn.linear_model import Ridge
from sklearn.linear_model import ElasticNet, ElasticNetCV
from sklearn.model_selection import train_test_split
from pandas import Series, DataFrame
```

```
df["suicide"] = "0"
|
def add(h, s, d):
    if (h >= 50):
        s = s/2
        d = d/2

    b = abs(s + d - h)
    c = b / 2
    return c

df['suicide'] = df.apply(lambda row : add(row['Happy_in_Life_(Percentage)'],
                                         row['Stress_in_Life_(Percentage)'], row['Depressed_(Percentage)']), axis = 1)
```

```
y = df.sucide
|
df_features = ['Gender', 'Age', 'Education', 'Marital_Status', 'Financial_Problem', 'Family_Fued', 'Sexual_Harrasment',
'Body_Shaming', 'Study_Pressure', 'Have_Caring_Parents', 'Lonely', 'Happy_in_Life_(Percentage)', 'Stress_in_Life_(Percentage)',
'Depressed_(Percentage)']

X = df[df_features]

#trin set
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3, random_state=1)

#for RandomForestRegressor
clf = RandomForestRegressor()

clf = clf.fit(X_train,y_train)

y_pred = clf.predict(X_test)
print(clf.predict(X.head()))

#error rate
print('Mean Absolute Error:', metrics.mean_absolute_error(y_test, y_pred))
print('Mean Squared Error:', metrics.mean_squared_error(y_test, y_pred))
print('Root Mean Squared Error:', np.sqrt(metrics.mean_squared_error(y_test, y_pred)))
|

#for LinearRegression
clf = LinearRegression()

clf = clf.fit(X_train,y_train)

y_pred = clf.predict(X_test)
print(clf.predict(X.head()))

#error rate

print('Mean Absolute Error:', metrics.mean_absolute_error(y_test, y_pred))
print('Mean Squared Error:', metrics.mean_squared_error(y_test, y_pred))
print('Root Mean Squared Error:', np.sqrt(metrics.mean_squared_error(y_test, y_pred)))
```



```
#for DecisionTreeRegressor
clf = DecisionTreeRegressor()

clf = clf.fit(X_train,y_train)

y_pred = clf.predict(X_test)
print(clf.predict(X.head()))

#error rate

print('Mean Absolute Error:', metrics.mean_absolute_error(y_test, y_pred))
print('Mean Squared Error:', metrics.mean_squared_error(y_test, y_pred))
print('Root Mean Squared Error:', np.sqrt(metrics.mean_squared_error(y_test, y_pred)))
|
#for Lasso
clf = Lasso()

clf = clf.fit(X_train,y_train)

y_pred = clf.predict(X_test)
print(clf.predict(X.head()))

#error rate

print('Mean Absolute Error:', metrics.mean_absolute_error(y_test, y_pred))
print('Mean Squared Error:', metrics.mean_squared_error(y_test, y_pred))
print('Root Mean Squared Error:', np.sqrt(metrics.mean_squared_error(y_test, y_pred)))
```

```
#for Ridge
clf = Ridge()

clf = clf.fit(X_train,y_train)

y_pred = clf.predict(X_test)
print(clf.predict(X.head()))

#error rate

print('Mean Absolute Error:', metrics.mean_absolute_error(y_test, y_pred))
print('Mean Squared Error:', metrics.mean_squared_error(y_test, y_pred))
print('Root Mean Squared Error:', np.sqrt(metrics.mean_squared_error(y_test, y_pred)))

|
#for ElasticNet
clf = ElasticNet()

clf = clf.fit(X_train,y_train)

y_pred = clf.predict(X_test)
print(clf.predict(X.head()))

#error rate

print('Mean Absolute Error:', metrics.mean_absolute_error(y_test, y_pred))
print('Mean Squared Error:', metrics.mean_squared_error(y_test, y_pred))
print('Root Mean Squared Error:', np.sqrt(metrics.mean_squared_error(y_test, y_pred)))

#for ElasticNetCV
clf = ElasticNetCV()

clf = clf.fit(X_train,y_train)

y_pred = clf.predict(X_test)
print(clf.predict(X.head()))

#error rate

print('Mean Absolute Error:', metrics.mean_absolute_error(y_test, y_pred))
print('Mean Squared Error:', metrics.mean_squared_error(y_test, y_pred))
print('Root Mean Squared Error:', np.sqrt(metrics.mean_squared_error(y_test, y_pred)))
```

Defense Final Summer 2021 (2)

ORIGINALITY REPORT

23%

SIMILARITY INDEX

18%

INTERNET SOURCES

14%

PUBLICATIONS

12%

STUDENT PAPERS

PRIMARY SOURCES

1	www.science.gov Internet Source	2%
2	doku.pub Internet Source	2%
3	journals.plos.org Internet Source	2%
4	Submitted to Daffodil International University Student Paper	2%
5	www.questia.com Internet Source	1%
6	Mohammed A. Mamun, Istihak Rayhan, Khaleda Akter, Mark D. Griffiths. "Prevalence and Predisposing Factors of Suicidal Ideation Among the University Students in Bangladesh: A Single-Site Survey", International Journal of Mental Health and Addiction, 2020 Publication	1%
7	www.centropiaggio.unipi.it Internet Source	1%