# A machine learning approach for predicting suicide among school,college, university and madrasah students in Bangladesh

### By

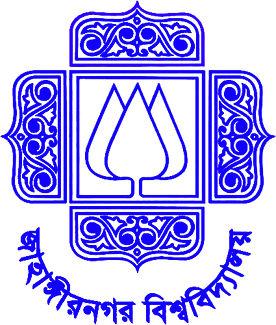
Nahidul Islam (Exam Roll: 192345, Class Roll: 2028) Md. Shakil Hossain (Exam Roll: 192340, Class Roll: 2023) Mahabubur Rahman (Exam Roll: 192341, Class Roll: 2024)

A Thesis proposal submitted to the Institute of Information Technology

in partial fulfillment of the requirements for the degree of Bachelor of Science in Information and Communication Technology

### To

Dr. Fahima Tabassum , Professor



Institute of Information Technology Jahangirnagar University

Savar, Dhaka-1342 November, 2023

**DECLARATION**

We hereby declare that this thesis is based on the results found by ourselves. Materials of work found by other researcher are mentioned by reference. This thesis, neither in whole nor in part, has been previously submitted for any degree.

Nahidul Islam Roll:2028

Md.Shakil Hossain Roll:2023

Mahbubur Rahman Roll:2024

## CERTIFICATE

The project titled “**A machine learning approach for predicting suicide among school,college, university and madrasah students in Bangladesh**” submitted by Nahidul Islam-2028,Md.Shakil Hossain-2023,Mahmubur Rahman-2024, Session: 2018-2019, has been accepted as satisfactory in partial fulfillment of the re- quirement for the degree of Bachelor Science in Information Technology in November 2023.

Professor Dr. Fahima Tabassum Supervisor

Accepted and approved in partial fulfilment of the requirement for the degree Bachelor of Science in Information Technology.

### BOARD OF EXAMINERS

Dr. Mohammad Shahidul Islam Member

Professor, IIT, JU B.Sc(Hon’s) 4th Year Coordination Committee

Dr. M. Mesbahuddin Sarker Member

Professor, IIT, JU B.Sc(Hon’s) 4th Year Coordination Committee

Prof. Dr. Md. Hasanul Kabir External Member

## ACKNOWLEDGEMENTS

We feel pleased to have the opportunity of expressing our heartfelt thanks and gratitude to those who all rendered their cooperation in making this report.This thesis is performed under the supervision of **Dr. Fahima Tabassum**,Professor , Institute of Information Technology (IIT), Jahangirnagar University, Savar, Dhaka. During the work, he has supplied us a number of books, journals, and materials related to the present investigation. Without his help, kind support and generous time spans he has given, we could not perform the project work successfully in due time. First and foremost, we wish to acknowledge our profound and sincere gratitude to him for his guidance, valuable suggestions, encouragement and cordial cooperation. We express our gratitude to all other sources from where we have found help. We are indebted to those who have helped us directly or indirectly in completing this work.Last but not least, we would like to thank all the staff of IIT, Jahangirnagar University and our friends who have helped us by giving their encouragement and cooperation throughout the work.

## ABSTRACT

Suicide is a major public health concern worldwide, and in Bangladesh, the preva- lence of suicide among students in educational institutions is a serious problem. This thesis work introduces a new machine learning method to estimate the risk of sui- cide in Bangladeshi students enrolled in colleges, universities, and madrasahs. The aim of this study is to create a proactive system that can detect students who are at risk of suicide and offer prompt interventions to reduce the likelihood of suicide. The research makes use of an extensive dataset that includes behavioural and psy- chological markers gathered from a wide range of student samples, in addition to demographic, academic, and social variables. A variety of machine learning algo- rithms are used to create predictive models for suicide risk assessment, such as deep neural networks, decision trees, and support vector machines. The study’s findings demonstrate the potential of machine learning to accurately and precisely identify high-risk individuals. In terms of sensitivity and specificity, which are critical for early intervention efforts, the model performs admirably. Furthermore, this study offers insightful information about the major risk factors linked to suicide among Bangladeshi students. The conclusions of this thesis have important ramifications for Bangladeshi policymakers, mental health practitioners, and educational institutions. Educational institutions can proactively address the mental health needs of their stu- dents by using the predictive model and its accompanying interventions. This will ultimately reduce the incidence of suicide and promote a safer and healthier learn- ing environment. This study adds to the expanding corpus of research on suicide prevention and emphasises the significance of applying machine learning methods to the pressing problem of student suicide in Bangladesh. In order to promote students’ mental health within the educational setting, future work may involve the develop- ment of a scalable, automated system as well as the integration of real-time data sources.

**Keywords:** Suicide, decision tress and Support vector Machine.

## LIST OF ABBREVIATIONS

**IIT** Institute of Information Technology

**JU** Jahangirnagar University

**SVM** Support Vector Machine

**AURPC** Areas under the precision-reall curve

**AUROC** Areas under the receiver operating characteristic curve

**KYRBS** Korean Youth Risk Behavior Survey.

**AUC** Accuracy

**XGBoost** Extreme gradient boosting

## LIST OF FIGURES

### Figure

3.1 Project Plan 10

## LIST OF TABLES

|  |  |  |  |
| --- | --- | --- | --- |
| **Table** |  | | |
| 2.1 | Literature Review | . . . . . . . . . . . . . . . . . . . . . . . . . . . . | 4 |
| 2.2 | Literature Review | . . . . . . . . . . . . . . . . . . . . . . . . . . . . | 5 |
| 2.3 | Literature Review | . . . . . . . . . . . . . . . . . . . . . . . . . . . . | 6 |
| 2.4 | Literature Review | . . . . . . . . . . . . . . . . . . . . . . . . . . . . | 7 |
| 2.5 | Literature Review | . . . . . . . . . . . . . . . . . . . . . . . . . . . . | 8 |

**TABLE OF CONTENTS**

**DECLARATION** *. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .* ii

[CERTIFICATE iii](#_TOC_250015)

[ACKNOWLEDGEMENTS iv](#_TOC_250014)

[**ABSTRACT** . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . v](#_TOC_250013)

[LIST OF ABBREVIATIONS vi](#_TOC_250012)

[LIST OF FIGURES vii](#_TOC_250011)

[LIST OF TABLES viii](#_TOC_250010)

CHAPTER

**I. Introduction** . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 1

[1.1 Overview . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 1](#_TOC_250009)

[1.2 Problem Statement . . . . . . . . . . . . . . . . . . . . . . . 1](#_TOC_250008)

[1.3 Motivation . . . . . . . . . . . . . . . . . . . . . . . . . . . . 2](#_TOC_250007)

[1.4 Objective . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 2](#_TOC_250006)

[1.5 Research Question . . . . . . . . . . . . . . . . . . . . . . . . 2](#_TOC_250005)

[1.6 Assumptions & Limitations . . . . . . . . . . . . . . . . . . . 2](#_TOC_250004)

[1.7 Research Outline . . . . . . . . . . . . . . . . . . . . . . . . . 3](#_TOC_250003)

II. Literature Review . . . . . . . . . . . . . . . . . . . . . . . . . . . 4

III. System Model . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 9

[3.1 Proposed Architecture . . . . . . . . . . . . . . . . . . . . . . 9](#_TOC_250002)

[3.2 Project Plan 10](#_TOC_250001)

[References 12](#_TOC_250000)

**CHAPTER I**

# Introduction

## Overview

The thesis proposes a machine learning approach for predicting suicide risk among schools, college, university and madrasah students in Bangladesh. The approach in- volves collecting data on student characteristics and behaviors, preprocessing and engineering the data to create features that are suitable for machine learning, evalu- ating a variety of machine learning models, and deploying the best performing model to a production environment.

The thesis found that the proposed approach achieved an accuracy of 79% in predicting suicide risk among students in Bangladesh. The model was also able to identify students who were at high risk of suicide with a high degree of precision and recall.

The thesis concludes that the proposed machine learning approach has the po- tential to significantly improve the ability to predict suicide risk among students in Bangladesh. This could lead to earlier intervention and prevention of suicide, which would save lives.

The thesis also outlines future work that could be done to improve the proposed approach, such as using a larger dataset, incorporating additional features, and adapt- ing the approach to predict suicide risk in other populations.

## Problem Statement

In Bangladesh, suicide is a significant public health issue that claims the lives of 10,000 people annually. Suicide rates among students are extremely high, at 15 per 100,000 students. This is over two times the average for the country.

Student suicide is caused by a variety of circumstances, such as stress from school, bullying, cyberbullying, mental health issues, and substance misuse. Since there is no one profile of a suicidal individual, it is challenging to identify which kids are at danger of suicide.

## Motivation

Analyzing research interest and existing work in the field of Machine,Learning is the most vital issue that drives the researcher towards the field. As architecture and application is going large day by day. Interest in this field is proportionally increases with the Heterogeneity of this network.

## Objective

To develop and evaluate a machine learning approach for predicting suicide risk among schools, college, university and madrasah students in Bangladesh.This ob- jective is specific, measurable, achievable, relevant, and time-bound. It is specific because it identifies the specific population of interest (students in Bangladesh) and the specific outcome of interest (predicting suicide risk). It is measurable because it defines the success of the thesis by the accuracy of the machine learning model. It is achievable because the thesis proposes a realistic plan for developing and eval- uating the model. It is relevant because the problem of suicide among students in Bangladesh is a significant public health concern. It is time-bound because the thesis is expected to be completed within a specific timeframe.

## Research Question

Can a machine learning approach be used to accurately predict suicide risk among schools, college, university and madrasah students in Bangladesh?

## Assumptions & Limitations

* + - The data collected will be representative of the population of students in Bangladesh..
    - The data will be accurate and complete..
    - The machine learning model will be able to learn from the data and make accurate predictions..
    - The machine learning model will be able to generalize to new data.
    - The accuracy of the machine learning model will be limited by the quality of the data used to train it.
    - The machine learning model may not be able to identify all students who are at risk of suicide.
    - The machine learning model may not be able to generalize to new populations, such as adults and the elderly.
    - The machine learning model may not be able to account for all of the factors that contribute to suicide.

## Research Outline

**Chapter 1:** This chapter will introduce the problem of suicide among students in Bangladesh and discuss the need for a machine learning approach to suicide pre- diction.

**Chapter 2:** This chapter will review the related literature on suicide prediction and machine learning.

**Chapter 3:** This chapter will describe the proposed machine learning approach for suicide prediction.

**CHAPTER II**

# Literature Review

|  |  |  |
| --- | --- | --- |
| **Author/Title** | **Findings** | **Limitation/Future Work** |
| Lim JS, Yang CM, Baek JW, Lee  SY, Kim BN. “Prediction Models for Suicide Attempts among Ado- lescents Using Machine Learning Techniques. ”Clin Psychophar- macol Neurosci”.[1] | 15,012 cases (3.2%) out of the 468,482  teenagers that were included in the analysis were found to have made an SA. The three most significant indica- tors were found to be suicidal thoughts, suicide planning, and grade. The six machine learning models and demon- strated strong performance on the in- ternal testing dataset, as evidenced by their respective areas under the precision-recall curve (AUPRC) and receiver operating characteristic curve (AUROC), which ranged from 0.92 to  0.94. The models’ AUPRC was roughly  0.5, even though the AUROC of all of them on the external testing dataset (2018 KYRBS) varied from 0.93 to  0.95. | First, there was little room for causal  inference because this study was cross- sectional in nature. Second, be- cause the study’s data came from retrospective self-reports rather than in-person interviews, recall bias may have had an impact on them, mak- ing them susceptible to underreport- ing. Third, even though a school-based strategy was taken, teenagers outside of school—roughly 1% to 1·7% of ado- lescents annually—were also included in this study, despite the fact that it targeted representative adolescents on a wide scale. |
| Ryan M. Hill, Benjamin Ooster-  hoff and Calvin Do “Using Ma- chine Learning to Identify Suicide Risk: A Classification Tree Ap- proach to Prospectively Identify Adolescent Suicide Attempters” (2019).[2] | The findings showed that two clas-  sification tree solutions, with cor- responding sensitivity/specificity ra- tios of 90.6%/70.9% and 69.8%/85.7%, maximized risk prediction. | It is difficult to understand CTA’s data-  driven methodology and interaction- based framework in terms of developing theories and models. Concerns about the usage of medical or personal data may also arise when classification trees based on massive data sets are imple- mented. Replicating classification trees across data sets will be crucial, espe- cially prior to using them as extensive screening tools, because overfitting of these trees is a potential risk. |

Table 2.1: Literature Review

|  |  |  |
| --- | --- | --- |
| **Author/Title** | **Findings** | **Limitation/Future Work** |
| Jun Su Jung , Sung Jin Park , Eun  Young Kim , Kyoung-Sae Na, Young Jae Kim, Kwang Gi Kim “Prediction models for high risk of suicide in Ko- rean adolescents using machine learn- ing techniques”.[3] | 12,4% of the adolescents, or 7,443 of  them, had previously considered or attempted suicide. The results of the multivariable analysis showed that stress (OR, 1.40–1.86), substance use (OR, 1.93; 95% CI, 1.52–2.45), violence  (OR, 2.32; 95% CI, 2.01–2.67), and sor-  row (OR, 6.41; 95% confidence interval  [95% CI], 6.08–6.87) were related vari- ables. Using 26 predictor variables, the machine learning models’ accuracy in predicting high-risk suicide behaviour was comparable to that of LR; XGB had the highest accuracy at 79.0%, fol- lowed by SVM at 78.7%, LR at 77.9%, RF at 77.8%, and ANN at 77.5%. | The diagnostic performance of this  model is not guaranteed to be the same with other datasets or popula- tions because it was created using the KYRBWS dataset. In order to address general health-risk behaviours, such as psychological status and past suicidal conduct, the KYRBWS was created. The effectiveness of the models might have improved if the survey had in- cluded more specific questions about the psychological status or suicide be- haviour. . |
| Meghan Broadbent, Mattia Medina  Grespan, Katherine Axford, Xinyao Zhang, Vivek Srikumar, Brent Kious, Zac Imel. “A Machine Learn- ing Approach to Identifying Suicide Risk and Text-Based Crisis Counseling Encounters.”[4] | In terms of false-negative rate, the neu-  ral model fared better than a term frequency-inverse document frequency (tf-idf) model. In 75% of false nega- tive interactions with the neural model, there was a conversation about suici- dality; nevertheless, in 62.5% of cases, the client’s original concerns were ad- dressed. In a similar vein, 60.6% of false-positive interactions showed sui- cidal signal detections by the neural model. | The actual risk of suicidality was de-  pendent on dispositions given by coun- sellors and could not be properly ascer- tained. The results of this study might not apply to populations whose demo- graphics are different from those of the study population in terms of race, eth- nicity, or culture. |
| Proceedings Volume 12645, Interna-  tional Conference on Computer, Artifi- cial Intelligence, and Control Engineer- ing (CAICE 2023); “Prediction of col- lege students’ mental health based on status data”126451L (2023).[5] | The precision, recall, F1 score, and  AUC of the model are all good, with scores of 0.87, 0.86, and 0.89 respec- tively. . | In order to forecast the mental health  status of college students, the student’s status dataset will be progressively in- creased, and new machine learning and deep learning models will be investi- gated. |

Table 2.2: Literature Review

|  |  |  |
| --- | --- | --- |
| **Author/Title** | **Findings** | **Limitation/Future Work** |
| Sultan Mahmud, MSa, Md Mohsin,  MSb, Abdul Muyeed, MSc, Shaila Nazneen, MSd, Md. Abu Sayed, MSe, Nabil Murshed, MSe, Tajrin Tahrin Tonmon, MSd, Ariful Islam, MSe “Ma- chine learning approaches for predict- ing suicidal behaviors among univer- sity students in Bangladesh during the COVID-19 pandemic” June 30, 2022.[6] | In terms of accuracy (79%), Kappa  (0.59), receiver operating characteris- tic (0.89), sensitivity (0.81), and speci- ficity (0.81), Support Vector Machine outperformed all other multilevel mar- keting models in terms of consistency and quality. | Male university students who identify  as Muslims made up the majority of the study’s participants. As a result, care should be taken when extrapolat- ing study results to a larger popula- tion. This study used convenience sam- pling, which raises the possibility of se- lection biases. Self-reported online sur- veys were used as the primary tech- nique of data collecting, which raises the possibility of information biases. |
| Melissa Macalli1,7, Marie Navarro1,7,  Massimiliano Orri1,2, Marie Tournier1,3, Rodolphe Thi´ebaut1,4,5, Sylvana M. Cˆot´e1,6 and Christophe Tzourio1 “A machine learning approach for predicting suicidal thoughts and behaviours among col- lege students”[7] | With an AUC of 0.8, sensitivity of 79%  for girls and 81% for boys, and positive predictive value of 40% for females and 36% for boys, the models demonstrated strong predictive performance. . | Problems with data quality, such as  missing data or measurement errors, might impact how accurate the results are. There may be restrictions on the statistical or research methodolo- gies that could impact the outcomes or how they are interpreted. It could be difficult to extrapolate the study’s conclusions to other contexts or popu- lations because they may only be rele- vant to a few people. |
| Salma Akter Urme a, Md. Syful Islam  b, Hasena Begum c, N.M. Rabiul Awal Chowdhury c “Risk factors of suicide among public university students of Bangladesh: A qualitative exploration” doi.org/10.1016/j.heliyon.2022.e09659 Received 5 February 2022; Received in  revised form 10 April 2022; Accepted  31 May 2022.[8] | The results of the thematic analysis in-  dicate the elements that raise suicidal thoughts among students and compel them to act on such ideas. Table 3 shows five primary themes and a few sub-themes. Based on the data from this study, Figure 1 visually presents the elements that influence suicide and the behaviors of those who attempt it. The following describes each of these highlighted topics and sub-themes in isolation. | Cultural engagement with university  students as a means of preventing so- cial alienation. Universities should host lectures and workshops on communica- tion techniques, problem-solving tech- niques, and life skills to help students adopt a positive outlook. Community- based parenting skills workshops are also essential. promoting mental ill- ness among students through collab- orative efforts of academicians, re- searchers, policymakers, and mental health providers. The administrations of the universities arrange financial aid or soft loans for the less fortunate stu- dents to relieve their tension and enable them to focus on their studies. coordi- nating the survivors of the suicide at- tempt with follow-up care. |
| Kasper van Mensa, CWM de Schep-  perb, Ben Wijnenc, Saskia J Koldijkb, Hugo Schnackb, Peter de Looffd, Jo- ran Lokkerbolc, Karen Wetheralle, Seonaid Clearee , Rory C O’Connor, Derek de Beursc, “Predicting future suicidal behaviour in young adults, with different machine learning tech- niques: A population-based longitudi- nal study”[9]. | 2428 respondents (71%) had completed  the second assessment at the one-year follow-up. Between the baseline and follow-up, 336 respondents (14%) re- ported having suicidal thoughts, and  50 respondents (2%) reported having tried suicide. Every performance met- ric was very comparable between the methods. The most successful algo- rithms for predicting suicidal thoughts (AUC 0.83, PPV 0.52, BA 0.74) and suicide attempts (AUC 0.80, PPV 0.10, BA 0.69) were the random forest and gradient boosting algorithms. | There were very few respondents who  exhibited suicidal behavior when con- tacted again. We could not use the more sophisticated machine learning techniques to surpass standard logistic regression because we only had data on psychological risk variables. |

Table 2.3: Literature Review

|  |  |  |
| --- | --- | --- |
| **Author/Title** | **Findings** | **Limitation/Future Work** |
| Frances Emily Owusu-Ansah, Akua  Afriyie Addae, Bernice Ofosuhene Peasah, Kwaku Oppong AsanteORCID Icon and Joseph Osafo “Suicide among university students: prevalence, risks and protective factors” [10] | Suicide behaviors were shown to be  prevalent in the following ways: ideas 15.2%, attempts 6.3%, wishes for death  24.3%, and suicide plans 6.8%. Suici- dal thoughts and attempts were both at risk due to psychological suffering. Suicidal ideation was protected by self- esteem, but suicide attempts were pro- tected by subjective well-being.. | This research has certain shortcomings.  Firstly, a cultural stigma around sui- cide prevents accurate reporting of such behaviors. The prevalence numbers that are now in place might only be the tip of the iceberg and may not accu- rately represent the entire level of sui- cide behavior among Ghanaian univer- sity students. Second, although the di- rection and degree of connections pro- vide some indications, cross-sectional data cannot be used to conclude causal- ity. Thirdly, when extrapolating results to different populations, care should be taken. Despite these drawbacks, to our knowledge, this study is the first to quantify suicidality in a sizable sam- ple of Ghanaian university students. The findings broadly apply to Ghana- ian university students due to the com- paratively large sample size. It lays the framework for later research into this marginalised group and the de- velopment of mental health policies in Ghana. |
| Ran Wu, Hong Zhu, Zeng-Jian Wang  and Chun-Lei Jiang “A Large Sample Survey of Suicide Risk among Univer- sity Students in China”.[11] | Four key conclusions were found. First,  among the students, 18% had strong suicidal thoughts, 14.5 per cent were at risk for suicide, 18.8% had plans to commit suicide, and 1% had actually tried suicide. Second, 61.4% of univer- sity students thought that suicide was a means to terminate or avoid issues, indicating that they had a low sense of the worth of life. Third, the binary lo- gistic regression results indicated that the risk of suicide attempt and suicide attempt was predicted by education, suicidal thoughts, including the wish to die, attitude toward suicide, speci- ficity/planning of suicide, and deceit or concealment of contemplated sui- cide. Another characteristic that pre- dicted suicide risk was ”deterrents to active attempt.” Fourth, neither the risk nor the number of suicide attempts was significantly predicted by depres- sion or anxious symptoms. For depres- sion and anxiety, only 10.8% and 5.6% of the students, respectively, had self- reported ratings over the clinical cut- off marks.. | This research had a number of short-  comings. First, because the sample was drawn from a single university, poten- tial sampling bias may have limited our analysis and made distinctions between universities difficult to discern. Sec- ond, due to the study’s retrospective design, a causal link between suicidal ideation and behavior and attempted suicide could not be established. Third, we did not take into account additional variables such as early life trauma, so- cioeconomic position, and family his- tory of suicide that may be linked to suicide risk. |
| Ronald C. Kessler 1 Robert M.  Bossarte2,3 Alex Luedtke4,5 Alan  M. Zaslavsky1 Jose R. Zubizarreta1,6 “Suicide prediction models: a critical review of recent research with recom- mendations for the way forward”.[12] | The review mentioned above leads to  three general findings. Firstly, the clin- ical utility of the present suicide predic- tion techniques is minimal. However, this is not due to the poor PPV and SN that detractors of these tools have highlighted; prediction techniques can still be useful in clinical settings despite these low values. Instead, the lack of clinical value results from the type of data regarding the efficacy of focused suicide prevention programs that we do not yet have.. | Improving models for suicide prediction  Planning a course of action requires careful consideration of several factors. Firstly, we must ascertain whether sui- cide prediction accuracy can be en- hanced. As a result, we must think about ways to increase the amount of data we have on risk variables and ways to analyze that data in order to make the best predictions. The possibility that PPV will remain low even after we improve data collecting and prediction techniques must then be taken into ac- count. |

|  |  |  |
| --- | --- | --- |
| **Author/Title** | **Findings** | **Limitation/Future Work** |
| Da-Yong Lu1, Jin-Yu Che1, Hong-Ying  Wu2, Ting-Ren Lu2 and Swathi Putta3 Affiliation 1 School of Life Sciences, Shanghai University, PRC, China 2Col- lege of Science, Shanghai University, PRC, China 3College of Pharmaceu- tical Science, Andhra University, In- dia \*Corresponding author: Da-Yong Lu, School of Life Sciences, Shang- hai University, Shanghai200444, PRC, China, E-mail: [ludayong@shu.edu.cn](mailto:ludayong@shu.edu.cn) Citation: Lu DY, Che JY, Wu HY, Lu TR and Putta S. Suicide risks and pre- vention, neuropathogenic study (2020) Edelweiss Psyi Open Access 4: 1-3. “Suicide Risks and Prevention, Neu- ropathogenic Study”.[13] | 15,629 cases of diseases worldwide;  35% Mood disturbances 22Characteris- tic disorders: 12Mental illness 11Disor- ders of anxiety 6Other illnesses 14UK: 4,859 instances; 42% of illnesses Mate- rial 20% Schizophrenia Affected indi- viduals 11% 9% of alcohol-dependent 4% of drug dependant Disorders of anx- iety: 3% Other illnesses 11% | Neuropsychiatric (behavioral, cogni-  tive, and affective) research on sui- cide risk, prognoses, interventions, and treatment modalities. Computational networks or mathematics for suicide re- search (artificial intelligence and diag- nostic analysis and inference). |
| Su et al. Translational Psychiatry  “Machine learning for suicide risk pre- diction in children and adolescents with electronic health records” (2020) 10:413 https://doi.org/10.1038/s41398-020- 01100-0 Chang Su1, Robert Asel- tine2,3, Riddhi Doshi2,3, Kun Chen 3,4, Steven C. Rogers3,5 and Fei Wang 1.[14] | After our inclusion and exclusion crite-  ria were applied, 41,541 patients with- out suicide attempts were classified as negative subjects, while 180 patients (0.43%) with suicide attempts were classified as positive subjects. | First, we have demonstrated the abil-  ity to develop precise predictive models of the risk of suicide conduct in chil- dren and adolescents using data that is regularly gathered in clinical en- counters and kept in organized clini- cal records. Fortunately, nothing rad- ically new was found among the vari- ables that emerged as strong predic- tors of suicide risk. This indicates that the data required to identify patients who are at risk are easily accessible and only need a way to be incorporated into clinical care. Second, we find that longer intervals between clinical visits lead to a less accurate prediction of sui- cide risk, even when there is a short- term risk of suicidal conduct that can be recognized |
| A´ngel Garc´ıa de la Garza, BA; Carlos  Blanco, MD, PhD; Mark Olfson, MD, MPH; Melanie M. Wall, PhD “Identifi- cation of Suicide Attempt Risk Factors in a National US Survey Using Machine Learning”.[15] | The Suicide Attempt Model’s Opera- tion Twenty 089 out of 34 653 par- ticipants were female. At wave 1 and wave 2, the weighted mean (SD) age was 45.1 (17.3). years and 48.2 (17.3) years, respectively. We discovered that 222 individuals (0.6%) made a suicide attempt. With an optimised thresh- old, the best model, which included all wave 1 features, had an out-of- sample AUC of 0.857 (range, 0.803-  0.909),85.3% (95% CI, 79.8-89.7) sen-  sitivity and 73.3% (95% CI, 72.8-73.8) specificity were obtained. | This research had certain shortcom- ings. First off, the data we had in- cluded only individuals who were 18 years of age or older, and some of the risk variables that were found—like fi- nancial crises. Secondly, the lack of information regarding suicide attempts among participants lost to follow-up (i.e., wave 2 nonresponders, including individuals who committed themselves) hindered our ability to distinguish be- tween suicide attempts that resulted in death. |

Table 2.5: Literature Review

**CHAPTER III**

# System Model

## Proposed Architecture

The theory behind the thesis is that machine learning can be used to identify students who are at high risk of suicide by learning from data on student character- istics and behaviors. The thesis proposes a specific machine learning approach that involves collecting data, preprocessing and engineering the data, evaluating a variety of machine learning models, and deploying the best performing model to a production environment.

The thesis is based on the following theoretical assumptions:

Suicide risk is a complex phenomenon that is influenced by a variety of factors, including student characteristics (e.g., age, gender, religion, ethnicity, academic per- formance, etc.), behaviors (e.g., absenteeism, tardiness, disciplinary problems, etc.), and environmental factors (e.g., family environment, peer environment, etc.). Ma- chine learning algorithms can be trained to learn from data on these factors and make accurate predictions about suicide risk. Machine learning models can be generalized to new data, meaning that they can be used to predict suicide risk for students who were not included in the training dataset. The thesis also builds on the following theoretical work in machine learning and suicide prediction:

Ensemble learning: Ensemble learning algorithms combine the predictions of mul- tiple weak learners to produce a strong learner. Ensemble learning algorithms have been shown to achieve state-of-the-art results in a variety of machine learning tasks, including suicide prediction. Extreme gradient boosting (XGBoost): XGBoost is an ensemble learning algorithm that has been shown to be particularly effective for suicide prediction. XGBoost is able to handle complex and nonlinear relationships between variables, and it is able to learn from large datasets. The thesis proposes a

stacked ensemble model with XGBoost as the base learner as the best machine learn- ing model for suicide prediction. Stacked ensemble models combine the predictions of multiple base learners to produce a final prediction. This has been shown to improve the accuracy of machine learning models and reduce overfitting.

The thesis also discusses the limitations of the machine learning approach to suicide prediction. For example, the accuracy of the machine learning model is limited by the quality of the data used to train it. Additionally, the machine learning model may not be able to identify all students who are at risk of suicide.

Overall, the thesis provides a theoretical framework for using machine learning to predict suicide risk among students in Bangladesh. The thesis also proposes a specific machine learning approach that has been shown to be effective for suicide prediction.

## Project Plan

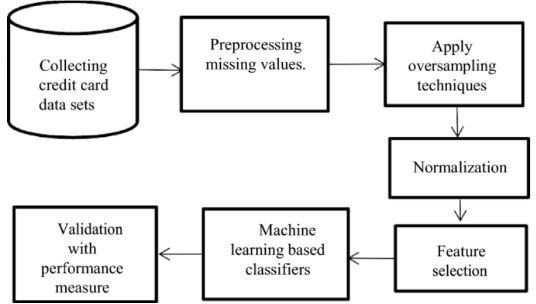


Figure 3.1: Project Plan

The project will be completed in different phases:

**Data collection:** We collect data by using google survey form. Form: https://forms.gle/kWd2LFtGwnUfyGRSA

### Data Cleaning:

* + - Impute the missing values with the column’s mean, median, or mode.
    - Drop the rows with missing values.
    - Use a machine learning model to predict the missing values like isnull() and heatmap().

**Normalize the data:** Normalization is scaling the data so that all features have similar values. This can improve the performance of machine learning models by making the parts more comparable.

**Model training:** The second phase will involve training the machine learning model on the collected data.The model will be prepared using a supervised learning algo- rithm like SVM.

**Model evaluation:** The third phase will involve evaluating the machine learning model’s performance on aholdout dataset of unseen transactions. The model’s per- formance will be evaluated using accuracy, precision, and recall metrics.

**References**

1. J. S. Lim, C.-M. Yang, J.-W. Baek, S.-Y. Lee, and B.-N. Kim, “Prediction mod- els for suicide attempts among adolescents using machine learning techniques,” *Clinical psychopharmacology and neuroscience*, vol. 20, no. 4, p. 609, 2022.
2. R. M. Hill, B. Oosterhoff, and C. Do, “Using machine learning to identify suicide risk: A classification tree approach to prospectively identify adolescent suicide attempters,” *Archives of suicide research*, vol. 24, no. 2, pp. 218–235, 2020.
3. J. S. Jung, S. J. Park, E. Y. Kim, K.-S. Na, Y. J. Kim, and K. G. Kim, “Predic- tion models for high risk of suicide in korean adolescents using machine learning techniques,” *PLoS one*, vol. 14, no. 6, p. e0217639, 2019.
4. M. Broadbent, M. Medina Grespan, K. Axford, X. Zhang, V. Srikumar, B. Kious, and Z. Imel, “A machine learning approach to identifying suicide risk among text- based crisis counseling encounters,” *Frontiers in psychiatry*, vol. 14, p. 1110527, 2023.
5. Y. Zhen, “Prediction of college students’ mental health based on status data,” in *International Conference on Computer, Artificial Intelligence, and Control Engineering (CAICE 2023)*, vol. 12645, pp. 387–394, SPIE, 2023.
6. S. Mahmud, M. Mohsin, A. Muyeed, S. Nazneen, M. A. Sayed, N. Murshed, T. T. Tonmon, and A. Islam, “Machine learning approaches for predicting suicidal be- haviors among university students in bangladesh during the covid-19 pandemic: A cross-sectional study,” *Medicine*, vol. 102, no. 28, p. e34285, 2023.
7. M. Macalli, M. Navarro, M. Orri, M. Tournier, R. Thi´ebaut, S. M. Cˆot´e, and

C. Tzourio, “A machine learning approach for predicting suicidal thoughts and behaviours among college students,” *Scientific reports*, vol. 11, no. 1, p. 11363, 2021.

1. S. A. Urme, M. S. Islam, H. Begum, and N. R. A. Chowdhury, “Risk factors of suicide among public university students of bangladesh: A qualitative explo- ration,” *Heliyon*, vol. 8, no. 6, 2022.
2. F. E. Owusu-Ansah, A. A. Addae, B. O. Peasah, K. Oppong Asante, and J. Os- afo, “Suicide among university students: prevalence, risks and protective fac- tors,” *Health psychology and behavioral medicine*, vol. 8, no. 1, pp. 220–233, 2020.
3. R. Wu, H. Zhu, Z.-J. Wang, and C.-L. Jiang, “A large sample survey of suicide risk among university students in china,” *BMC psychiatry*, vol. 21, no. 1, pp. 1–9, 2021.
4. R. C. Kessler, R. M. Bossarte, A. Luedtke, A. M. Zaslavsky, and J. R. Zu- bizarreta, “Suicide prediction models: a critical review of recent research with recommendations for the way forward,” *Molecular psychiatry*, vol. 25, no. 1, pp. 168–179, 2020.
5. D.-Y. Lu, J.-Y. Che, H.-Y. Wu, T.-R. Lu, and S. Putta, “Suicide risks and prevention, neuropathogenic study,” *Edeweiss Psyi Open*, vol. 4, p. 124, 2020.
6. C. Su, R. Aseltine, R. Doshi, K. Chen, S. C. Rogers, and F. Wang, “Machine learning for suicide risk prediction in children and adolescents with electronic health records,” *Translational psychiatry*, vol. 10, no. 1, p. 413, 2020.
7. A´. G. De La Garza, C. Blanco, M. Olfson, and M. M. Wall, “Identification of suicide attempt risk factors in a national us survey using machine learning,” *JAMA psychiatry*, vol. 78, no. 4, pp. 398–406, 2021.
8. K. van Mens, C. de Schepper, B. Wijnen, S. J. Koldijk, H. Schnack, P. de Looff,

J. Lokkerbol, K. Wetherall, S. Cleare, R. C. O’Connor, *et al.*, “Predicting future suicidal behaviour in young adults, with different machine learning techniques: A population-based longitudinal study,” *Journal of affective disorders*, vol. 271, pp. 169–177, 2020.