

**PREDICTING STRESS LEVEL IN ONLINE EDUCATION DUE TO  
CORONAVIRUS PANDEMIC: A CASE STUDY OF BANGLADESHI STUDENTS**

**BY**

**SUNZIDA SIDDIQUE  
ID: 172-15-9667  
AND**

**SAJAL BAIDYA  
ID: 172-15-9656**

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

Supervised By

**Md. Sadekur Rahman**  
Assistant Professor  
Department of CSE  
Daffodil International University

Co-Supervised By

**Md. Azizul Hakim**  
Lecturer (Senior Scale)  
Department of CSE  
Daffodil International University



**DAFFODIL INTERNATIONAL UNIVERSITY**

**DHAKA, BANGLADESH**

**JUNE 2021**

## **APPROVAL**

This Research Based Project titled "Predicting Stress Level in Online Education Due to Coronavirus Pandemic: A Case Study of Bangladeshi Students", submitted by Sunzida Siddique 172-15-9667, Sajal Baidya 172-15-9656 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.

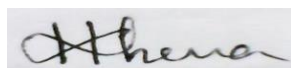
### **BOARD OF EXAMINERS**



---

**Dr. Touhid Bhuiyan**  
**Professor and Head**  
Department of CSE  
Faculty of Science & Information Technology  
Daffodil International University

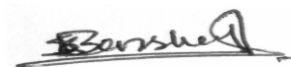
**Chairman**



---

**Most. Hasna Hena**  
**Assistant Professor**  
Department of CSE  
Faculty of Science & Information Technology  
Daffodil International University

**Internal Examiner**



---

**Sumit Kumar Banshal**  
**Senior Lecturer**  
Department of CSE  
Faculty of Science & Information Technology  
Daffodil International University

**Internal Examiner**



---

**Shah Md. Imran**  
**Industry Promotion Expert**  
LICT Project, ICT Division, Bangladesh

**External Examiner**

## DECLARATION

This is to certify that this project entitled " **Predicting Stress Level in Online Education Due to Coronavirus Pandemic: A Case Study of Bangladeshi Students** " has been done by us under the supervision of **Md. Sadekur Rahman, Assistant Professor**, 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:



---

**Md. Sadekur Rahman**  
Assistant Professor  
Department of CSE  
Daffodil International University

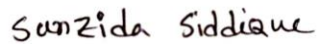
### Co-Supervised by:



---

**Md. Azizul Hakim**  
Lecturer (Senior Scale)  
Department of CSE  
Daffodil International University

### Submitted by:



---

**Sunzida Siddique**  
ID: 172-15-9667  
Department of CSE  
Daffodil International University



---

**Sajal Baidya**  
ID: 172-15-9656  
Department of CSE  
Daffodil International University

## ACKNOWLEDGEMENT

First, we disclose our heartiest thanks and gratefulness to almighty God for His divine blessing which has made us possible to complete the final year project successfully.

We are really thankful and wish our profound our indebtedness to **Md. Sadekur Rahman, Assistant Professor**, 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 drafts and correcting them at all stage have made it possible to complete this project.

We would like to express our heartiest gratitude to **Dr. Touhid Bhuiyan, Professor and Head**, Department of CSE 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 patience of our parents.

## **ABSTRACT**

Education is one of the basic humans needs it is referred to as a country's backbone. Nowadays, Smart education technology has made the entire world smarter by the day. In just a few months, The Novel coronavirus (covid-19) has changed the world. As a result, the entire world is no longer relevant. In the context of Bangladesh, continuing one's studies while dealing with the pandemic is a new challenge. The role of education in our daily life is immense and in this pandemic time, there has been an unprecedented change in the field of education. The online education system has made education a novelty when the whole world has come to a standstill. This paper aims to find out the types of issues that a student may encounter in an online class during the covid-19 situation in Bangladesh. In this covid-19 situation in Bangladesh, students continue their activities online and a student needs to analyze the types of issues that a student may encounter in an online class. In this paper, we used a machine-learning algorithm to and Data have been analyzed by basic questionnaires. To predict this problem, the accuracy was 74 percent given by Random forest by selecting all feature and 66 percent given by Xgb Boost by selecting Limited Feature. The motive of physical class and online class are almost similar to each other and it is very difficult to compare each class value. We used seven algorithms to run our model and the algorithm is KNN, Naive bias, decision tree, random forest, Xgb boost, SVM, and neural network. We have made use of Random forest classification algorithms, correlation heatmap and univariant selection to select necessary data and to remove the unnecessary data that are not related to predicting students' stress level in online classes that have been discussed in a separate section of our paper.

## TABLE OF CONTENTS

<b>CONTENTS</b>	<b>PAGE</b>
Board of examiners	i
Declaration	ii
Acknowledgements	iii
Abstract	iv
<b>CHAPTER</b>	
<b>CHAPTER 1: INTRODUCTION</b>	<b>1-5</b>
1.1 Introduction	1-2
1.2 Motivation	2-3
1.3 Rationale of the Study	3
1.4 Research Questions	3
1.5 Expected Output	3-4
1.6 Project Management and Finance	4
<b>CHAPTER 2: BACKGROUND</b>	<b>6-11</b>
2.1 Preliminaries/Terminologies	6-7
2.2 Related Works	8-9
2.3 Comparative Analysis and Summary	9-10
2.4 Scope of the Problem	10
2.5 Challenges	10-11
<b>CHAPTER 3: RESEARCH METHODOLOGY</b>	<b>12-33</b>
3.1 Research Subject and Instrumentation	12
3.2 Data Collection Procedure/Dataset Utilized	12-15

3.3 Statistical Analysis	16-20
3.4 Proposed Methodology/Applied Mechanism	20-32
3.5 Implementation Requirements	33
<b>CHAPTER 4: EXPERIMENTAL RESULTS AND DISCUSSION</b>	<b>34-39</b>
4.1 Experimental Setup	34
4.2 Experimental Results & Analysis	35-39
4.3 Discussion	39
<b>CHAPTER 5: IMPACT ON SOCIETY, ENVIRONMENT AND SUSTAINABILITY</b>	<b>40-41</b>
5.1 Impact on Society	40
5.2 Impact on Environment	40
5.3 Ethical Aspects	40-41
5.4 Sustainability Plan	41
<b>CHAPTER 6: SUMMARY, CONCLUSION, RECOMMENDATION AND IMPLICATION FOR FUTURE RESEARCH</b>	<b>42-43</b>
6.1 Summary of the Study	42
6.2 Conclusions	42-43
6.3 Implication for Further Study	43
<b>REFERENCES</b>	<b>44-45</b>

## LIST OF FIGURES

<b>FIGURES</b>	<b>PAGE NO</b>
Figure 3.1: Student Stress Measurement Ratio on Online	14
Figure 3.2: Data clean	15
Figure 3.3: Data Description of Attribute	15
Figure 3.4: Step of Data Preprocessing	15
Figure 3.5: Data visualization for all Feature	16
Figure 3.6: Correlation Heatmap	17
Figure 3.7: Feature Correlation with label	18
Figure 3.8: Univariate Selection Score	19
Figure 3.9: Random Forest Ranking	19
Figure 3.10: Final Feature Selection	20
Figure 3.11: Illustrate our proposed Methodology	21
Figure 3.12: Error Rate for KNN Algorithm for Selecting All Feature	22
Figure 3.13: Confusion matrix for KNN Algorithm for selecting all Feature	22
Figure 3.14: Error Rate for KNN Algorithm for limited feature	23
Figure 3.15: Confusion matrix for KNN Algorithm for Limited Feature	23
Figure 3.16: Confusion matrix for Naïve Bias Algorithm for all Feature	25
Figure 3.17: Confusion matrix for Naïve bias Algorithm for Limited Feature	25
Figure 3.18: Error Rate for Decision Tree Algorithm for all Feature	26
Figure 3.19: Confusion matrix for Decision Tree Algorithm for all Feature	26
Figure 3.20: Error Rate for Decision Tree Algorithm for limited Feature	26
Figure 3.21: Confusion Matrix for Decision Tree for Limited Feature	27
Figure 3.22: Error Rate for Random Forest Algorithm for All Feature	27
Figure 3.23: Confusion Matrix for Random Forest using for All Feature	28
Figure 3.24: Error Rate for Random Forest for using Limited Feature	28



Figure 3.25: Confusion matrix for Random Forest for (Limited Feature)	28
Figure 3.26: Confusion Matrix for Xgb Boost Algorithm for All Feature	29
Figure 3.27: Confusion matrix for Xgb Boost for (Limited Feature)	29
Figure 3.28: Error Rate for SVM Algorithm for selecting all feature	30
Figure 3.29: Confusion matrix for SVM Algorithm for using All Feature	30
Figure 3.30: Error rate for SVM Algorithm for Limited feature	30
Figure 3.31: Confusion matrix for SVM Algorithm for Limited Feature	31
Figure 3.32: Error rate for Neural network algorithm for all feature	31
Figure 3.33: Error rate for Neural network algorithm for Limited feature	32
Figure 3.34: Confusion matrix for NN Algorithm for Limited Feature	32

## LIST OF TABLES

<b>TABLES</b>	<b>PAGE NO</b>
Table 1.1: Project Management and Timeline	4
Table 2.1: Comparative Analysis and Summary of different approaches made to predict student Stress in online class	9
Table 3.1: Summary of different attribute name to predict online student stress	13
Table 4.1: Performance table for KNN Algorithm for all feature selection	35
Table 4.2: Performance table for Random Forest Algorithm all feature selection	36
Table 4.3: Performance table for Xgb Boost Algorithm for all feature selection	36
Table 4.4: Accuracy Table For Different Algorithm (All Feature)	37
Table 4.5: Performance table KNN Algorithm for Limited feature selection	37
Table 4.6: Performance table for Random Forest Algorithm for Limited feature selection	38
Table 4.7: Performance table for Xgb Boost Algorithm for Limited feature selection	38
Table 4.8: Accuracy table for different algorithm (Limited Feature)	39

# CHAPTER 1

## INTRODUCTION

### 1.1 Introduction

During the Corona period, when the whole world was locked up, the education system of Bangladesh started online activities. Which is a new chapter in the context of Bangladesh. The emergence of the coronavirus has forced the world to face difficult situations. The world has seen how deadly the covid-19 virus can be. People around the world have been affected by an epidemic of coronavirus. It is a zoonotic disease and it is caused by a coronavirus, it has not been found before in humans. The coronavirus was first identified in late December in the city of Wuhan, Hubei Province, China, and WHO March 11 declared it an epidemic. Since then, the whole world has been affected by the virus. The virus first spread to China, then to Europe, and then to Italy, where it spread to other European countries, including Spain and the UK. It then took a deadly turn in the USA, Brazil, and India. The city of Wuhan in China was declared a lockdown early in the coronavirus outbreak. Then there was a worldwide lockdown when the Corona epidemic broke out and millions of people began to be affected. Then people face serious problems. All kinds of business and trade are stopped, transportation is stopped, coming and going from one country to another is stopped. People are facing a kind of catastrophe. With the ban on travel and the closure of business, all kinds of educational institutions were closed. No medicine or method to deal with this disease was known and no discovery was made. Only by maintaining social distance could the spread of the disease be stopped and the incidence of the disease be reduced. So, the academy was declared closed. 1.5 billion students worldwide are affected by the epidemic coronavirus. After that, some educational institutions started online classes to keep the education system running temporarily.

Educational institutions thought that the introduction of online education would enable students to continue their studies in a state of epidemics like this epidemic and students would be able to continue their studies. Technology has advanced a lot at the present time.

Education and learning have been maintained even in the lockdown by maintaining social distance through technology. However, due to this epidemic, people face severe mental

problems. When people go to work, there is a loss in business. People began to be severely affected and millions upon millions of people began to die. People become unemployed and start suffering from mental problems.

After that, even though the online class education system was introduced, the students were facing various pressures. The financial system collapsed. And there are many questions about the quality of education in online classes. Not all countries in the world have the same financial system and not all countries are financially well off. And the internet system of the country or all areas is not good. For this, the online education system faces a kind of discrimination. Not everyone can continue the education system equally and many educational institutions do not have the capacity to continue the online education system. Along with all the countries of the world, Bangladesh also faces problems in getting people. The coronavirus has had a devastating effect on the education system of Bangladesh. In this report, we have focused on the Prediction of student stress during online class.

## **1.2 Motivation**

Educational institutions around the world have been closed for more than a year due to the Corona epidemic. Corona has had a huge impact on the education system of Bangladesh along with the rest of the world. Some educational institutions have the ability to continue their online education activities and they have been able to continue their students' studies online. However, most of the countries are poor and not everyone has the opportunity to study in a good educational institution and not all educational institutions have the capacity to continue their online education. The financial system of Bangladesh has suffered a lot and the education system has suffered a lot.

Although online classes have been introduced through the online education system, students have not been able to receive a proper and quality education. Although online classes have been introduced online, students have faced various pressures through online learning activities. The Corona epidemic is responsible for the fact that people are becoming mentally weak and coronary conditions are causing students to take online education but also to break down mentally due to various types of stress. People are

financially indigent and suffer from depression and many are suffering from depression due to failure to take up educational activities. He has faced various problems including mental anxiety, stress, fear, depression, mental exhaustion.

For this corona situation, we will be able to find out through our work how the corona has affected the Bangladesh education system and what kind of emotional change has taken place among the students as a result of running online classes. The mental state of the students.

### **1.3 Rationale of The Study**

The main purpose of this project is to identify a student's problem during online classes and prove that a student is capable of doing online class or not. By determining these activities, we will be able to measure the student's problems that they face in online education system during covid-19 pandemic in Bangladesh.

### **1.4 Research Question**

In order to conduct the research, we have set the following research questions:

- Can we collect real and authentic data from Bangladeshi Students to study their stress level due to covid-19 pandemic?
- Does Feature selection help improve accuracy?
- Which classifier perform best in predicting stress level of students?

### **1.5 Expected Output**

The expected result of our research is to prepare a calculation where a student will find out what problems he is facing in the online class during covid-19 pandemic time in Bangladesh. The algorithms we apply will look for whether a student is experiencing any mental, physical, or financial or No problems at all. And will try to give incoming new students a prediction of such problems reacts to our data. We discovered which machine learning algorithms had the highest accuracy, which we can use in future applications. Different types of machine learning algorithms will be applied to our dataset to test the model's performance. The goal of the project is to see how precise estimations.

## 1.6 Project Management and Finance

Table 1.1: Project Management and Timeline

<b>Activities</b>	<b>Timeline</b>
Planning and Knowledge Gathering	3 months
Data Collection	3 Months
Data Preprocessing	1 Months
Implementation	1 Months
Report Generation	20 Days
Total	8 Months 20 Days

The members of the project team are solely responsible for funding this project.

## 1.7 Report Layout

In our report we have tried to cover six chapters.

Chapter 1 Provides a brief introduction related to the online class problem of a student in Bangladesh perspective. Then covered Motivation of our project, it's Objectives, It's Expected Outcome, & the Report Layout.

Chapter 2 Provides a description about Terminologies, some Related Works, Comparative Analysis and Summary, Scope of the Problem and Challenges.

Chapter 3 Provides a description about Research Subject and Instrument, Data Collection Procedure, Statistical Analysis, Proposed Methodology and Implementation Requirements.

Chapter 4 provides a description Experimental Setup, Experimental Results and Analysis, Discussion

Chapter 5 provides a description of Society Impact, Environmental impact, Ethical Aspects, Sustainability Plan.

Chapter 6 provides a description of Summary of the Study, Conclusions and Implication for Further Study.

## **CHAPTER 2**

### **BACKGROUND**

#### **2.1 Terminologies**

##### **Machine Learning (ML):**

The abbreviation of ML stands for machine learning. It is Machine learning that is a field of combination in computer sector as well as the statistical sector. It is a model that learns from observed data. After that, the outcomes are predicted by unobserved data. In most cases, by providing more data it will learn better for better performance. In a single word, it's a set of algorithms that aid in the analysis of statistical data [13].

##### **K-Nearest Neighbors (KNN):**

KNN denoted for K Nearest Neighbor. As a machine learning algorithm, this algorithm is used in classification problem and regression problem. When there is a pattern and statistical analysis KNN was very well there. It is a KNN algorithm that is based on feature similarity. It is a very simple algorithm that gives a high accuracy. So, for analysis (classes, real value) this algorithm is very useful [13].

##### **Naïve Bias:**

The naive bias denotes a new name that is simple bias. It is used in probabilistic analysis. It is used to predict of an assumption. This algorithm is used in Machine Learning as a probabilistic analysis. It is analysis in classification problems and multiple class problem very well. To prediction's outcome, it gives the highest probability. So, for analysis different classes with different various attribute this algorithm is very useful [13].

##### **Decision Tree:**

A decision tree is called a support tool of decision. It generates a tree. It can make decisions and give an outcome of possible consequences. We used it (explicitly, visually). By using this it can make solve decision-related problems. It is a decision tree that gives all possible



outcomes by tracing each path.so, A decision tree is used for solving regression and classification analytical problems [13].

### **Random Forest (RF):**

Random decision forests are a method that is used for classification problem analysis, regression problem analysis, and other tasks also. It is a machine learning algorithm that is flexible. It helps to find a good result the majority of the time. It is used for both simplicity and variety. It helps to make multiple decisions and helps to get good accurate predictions. It creates a lot of trees. Its elimination overfitting problem. By doing that it improves its accuracy [13].

### **Xgb Boost:**

Xgb Boost is called gradient boosting algorithm. To determine a range, it gives a better performance. it is a method that is helps to residuals errors of models to predict final outcome. For to minimize loss it helps a lot, it helps to prevent overfitting of the model it built and developed solely for model performance and computational speed. [13].

### **Support Vector Machine (SVM):**

SVM stands for Support Vector Machine. It helps both classification analysis and regression problem analysis. It helps to divide the classes. The line boundary that divides n-dimensional space into classes is created by svm. SVM models can categorize new text after being given sets of labeled training data for each category. [13].

### **Neural Network (NN):**

NN stands for Neural Network. It is NN that tries to mimic the human brain. As we know a human brain builds with a billion of neurons' works by passing one data to another. It is a set of algorithms that uses a process to attempt to recognize underlying relationships in a set of data. It is used to image and classification problem analysis. Because a neural network is made up of many perceptron layers, the name "multi-layer perceptron" was coined. These neurons receive the information from the set of inputs. To produce a single output, these numerical inputs are combined with a bias and a set of weights. [13]

## 2.2 Related Works

There has been a lot of research on education since the Coronation period. Some have researched variously on online and offline education.

Shweta [1] discusses the online classes vs the offline classes which is more important to a student. This paper uses the (DEA) model that means Data Envelopment Analysis Model. The data set had to be assisted by two types of classes where the online class had 26 students offline and 44 students online. This paper focuses on an average score with an average score of 6.00 in the online class and 5.76 in the offline class. They have suggested to improve their samples. Muhammad [2] discussed the legging and issues of a student on online education problems in Pakistan. The survey was conducted with 126 students where 84 were female and 42 were male. As a result of the survey, 10.3% of students gave online education as a positive feedback and 71.45% of students gave negative feedback. They suggested that their sample size was too small and was non-random which they would later in case. A Abidah [3] The paper covers various sectors, including physical and social distance, self-quarantine, and the education sector in Indonesia. Data has been collected from UNESCO where Out of 160 countries, 871 percent have total student enroll. They used conceptual methodology in their paper. Kunal [4] shows a measurement of the negative impact of corona has on education. The survey deal with 1182 data. Kruskal Wallis test, Fisher's exact test and Pearson Chi Square test was used in this paper with statistically value of  $P < 0.05$  was. This paper suggests that sample testing has been used in this paper and work has been done with specific area. Sumitra [5] discussed in about the online learning and teaching system in covid-19 period. It has been suggested to focus on different levels of classes. Effat [6] Discuss the deterioration of the education system in Bangladesh. The number of data sets in the paper press was 2460. Sampling method has been used in this paper. The paper suggests increasing the scale of statistical analysis. Pan [7] This paper discusses a student's performance and satisfaction in a student online class. The survey deal with 397 student's data from choice and 433 student's data from no choice. The satisfaction level was 11% in online performance. Two-Factor Analysis of Variance (ANOVA) was used as their methodology. They suggest to focus on the noise of the data

set. William [8] discusses the improve of online education system. By using technology. This is the module which they take as their platform and that is LMS.

### 2.3 Comparative Analysis and Summary

There are wide ranges of approaches made to predict student stress on online class. Experiments used different algorithms, sensors, and different accuracy was:

Table 2.1: Comparative Analysis and Summary of different approaches made to predict students stress in online class

Author	year	Number of datasets with participation	Model	Score
Shweta [1]	2012	26 -offline 44 - online	Data Envelopment Analysis Model	average score 6.00 - online class 5.76 - offline class
Muhammad [2]	2020	126 students 84 -female 42 - male	Sampling method	10.3%-positive feedback 71.45%-negative feedback online class
A Abidah [3]	2020	871	conceptual methodology	Not specified
Kunal[4]	2021	1182	Kruskal Wallis test,Fisher's exact test , Pearson Chi Square test	statistically value of $P < 0.05$
Effat[6]	2020	2460	Sampling method	Not specified
Pan [7]	2002	397-choice,433-No choice	(ANOVA)	Satisfaction 11%

Form the comparative analysis in the table 2, we see that a lot of research have been survey on covid-19 impact on education and online education but performance can be achieved better if there used algorithms together.

#### **2.4 Scope of The Problem**

All educational institutions in Bangladesh, including the rest of the world, were closed during the operation. Then some educational institutions started teaching students online. Many students started learning online, many students did not accept taking classes online and many students did not have the opportunity to take classes online. In the context of Bangladesh, students suffer from many kinds of mental problems during the coronation. Many people start to suffer financially, especially when people lose their jobs. Online students face a lot of stress for learning.

Our paper deals with the mental state of students during their online classes in Bangladesh, their opinions, and the emotional stress that students face when it comes to online classes.

In the context of Bangladesh, work has been done on the mental state of the students for online classes during the Corona period and the mental condition of the students due to the closure of educational institutions in Corona. No such secrets have been worked out and even if such things have been worked out, there is not much research on online classes and the pressure to study at this time. Some studies have been done but they have used data sets from the internet and may not be valid. In our study, authentic data sets of Bangladeshi students have been used. Opinions have been taken from the students of Bangladesh and their opinions have been taken. Many good applications can be created from our research and from datasheets.

#### **2.5 Challenges**

Our research has had to face a variety of challenges, there have been a variety of problems, they have overcome these problems. Collecting data from ordinary students in Bangladesh was a challenge. Most of the data taken from the people of Bangladesh are students. All kinds of students and everyone's opinion has been taken. Here are the opinions of others including students who gave their opinions from their online class experience. Those who

provided our data, many have taken online classes here and many have not taken online classes or have not had the opportunity to take online classes. Everyone here has commented on the state of mind they have through online classes. We collect our data survey from the Google Forms site and share it on various social media. Different people have been collecting data from them personally through social media. We collect data through Google Forms through online. It was very difficult to do because people did not want to give data easily and did not want to fill the form. Even then it was more difficult to fix it by pre-processing some missing values and data. Also feature selection, algorithm selection, the first stage was much tougher. Algorithm implementation was really a lot harder. Despite so many problems, this research work has been completed. Our research has uncovered information on how much stress students have to take online learning while using machine learning.

## CHAPTER 3

### RESEARCH METHODOLOGY

#### 3.1 Research Subject and Instrument

As we are a human being, we have the ability to think anything and to recreate anything. When we make any decisions at any circumstances by using our brain then it is called Intelligence. But in the meanwhile, when we implemented any intelligence by using a machine then it is called artificial intelligence. It is machine learning that is a field of artificial intelligence, by train machines from different kind of example and experience. It is one kind of system maker Machines that is as like as human beings to make the decisions. It helps to teach computer to learn from data and act without programs. It's given a data to the generating various algorithm of the Machines and the machine that will be learn by the data automatically. Doing such a thing into machine these are used in such a place where taking human is work is difficult to do. In our daily life it is a blessing period to us.

#### 3.2 Data Collection Procedure

Data is that is the main part and efficient way to conduct any kind of research in a specific field. Data is, without a doubt, the center of the machine-learning process. Furthermore, we choose a limited number of data sources for our research. As a result, it has become the most difficult task in our research. We gathered 1,264 real-time data from Bangladeshis via an online survey that utilized a variety of social media platforms. Due to the horrors of Covid-19, a lockdown system has been issued in Bangladesh, the beginning of which we have to take the data through an online survey. For Data collection survey we used google form.

After data collection, we have preprocessed the raw data and we make a it for a suitable dataset for implement different algorithms. First of all, we have removed unnecessary strings to irrelevant string of the text file. After that we clean our dataset and make it suitable for data preprocessing.

Table 3.1: Summary of different attribute name to predict online student stress

No	Attribute Name in short form	Attribute name in Full Form	Unit
1	gen	gender	years
2	edu	education	N/A
3	fiwc	Is family infected with Corona	N/A
4	iedfc	is education disrupted for corona	N/A
5	uaop	use any online platform	N/A
6	wopu	which online platform use	N/A
7	ocbss	online class become student skillful	N/A
8	saiem	student assess in exam	N/A
9	sfsiocl	student face any stress in online class	N/A
10	oibtpcls	online is better than physical class	N/A
11	op	Opinion on online class	N/A
12	osph	online spent hour	hour
13	ostdh	online study hour	hour
14	pclsr	physical class rate	N/A
15	oclsr	online class rate	N/A
16	label	Problem of online class	N/A

Then we have converted each text file into comma-separated values (csv) file. Here, we have 15 input columns and 1 output column where we have labeled the activity name. After that we have converted it into csv files into one to get our complete dataset. In our dataset we have got a total 1,264 samples. The ratio of csv file stress measurement in Figure-3.1”

Mental=528

No problem=351

Financial=247

Physical=137

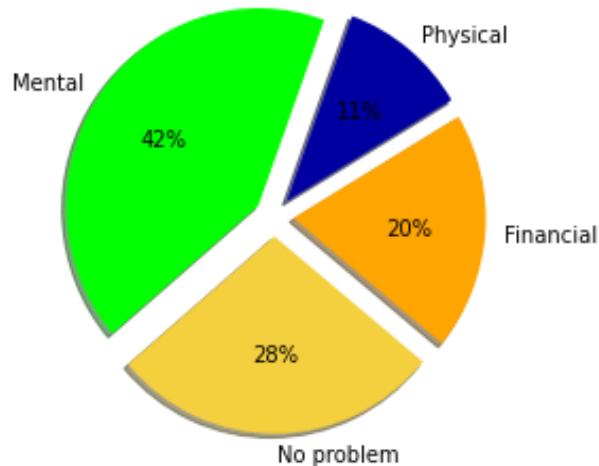


Figure 3.1: Student Stress Measurement Ratio on Online

splitting the dataset, we have implemented feature scaling on our dataset to bring all the data points into

After that we had done Label encoding as our two types of data in our data set, a numerical and a categorical. We have converted our categorical data to numerical. After that we have split our dataset (train set and test set) at the ratio of 8:2.

After splitting the dataset, we have implemented feature scaling. We use it to bring all the data points into a specific range. We used Standard Scalar to scale the dataset. After that we implement Machine learning algorithms using all feature in Method-1. After calculating all feature, we use selected Feature which is find by random forest classifier, correlation and univariant feature selection in Method-2. By the mixture and selection of all these common 7 feature we have again implement Machine Learning Algorithm using Limited Feature.



	gen	edu	fiwc	iedfc	uaop	wopu	ocbss	saiem	sfsiocls	oibtpcsls	op	osph	ostdh	pclsr	oclsr	label
792	1.0	1.0	0.0	0.0	1.0	3.0	0.0	2.0	1.0	0.0	2.0	6.0	4.0	9.0	8.0	2.0
554	1.0	0.0	0.0	2.0	1.0	3.0	1.0	1.0	2.0	1.0	0.0	9.0	5.0	9.0	3.0	1.0
592	1.0	1.0	0.0	2.0	1.0	2.0	0.0	1.0	1.0	1.0	2.0	5.0	3.0	9.0	8.0	2.0
186	1.0	1.0	1.0	2.0	1.0	2.0	0.0	1.0	2.0	1.0	1.0	8.0	4.0	9.0	5.0	1.0
219	1.0	1.0	1.0	2.0	1.0	2.0	0.0	1.0	1.0	1.0	1.0	11.0	1.0	1.0	7.0	1.0

Figure 3.2: Data Clean

After a brief description of dataset as shown in the Figure-3.3 below

	gen	edu	fiwc	iedfc	uaop	wopu	ocbss	saiem	sfsiocls	oibtpcsls	op	osph	ostdh	pclsr	oclsr	label
count	1263.000000	1263.000000	1263.000000	1263.000000	1263.000000	1263.000000	1263.000000	1263.000000	1263.000000	1263.000000	1263.000000	1263.000000	1263.000000	1263.000000	1263.000000	1263.000000
mean	0.605701	0.890736	0.444972	1.816310	0.895487	2.330958	0.736342	0.973872	1.435471	0.899446	0.752969	5.067300	3.285036	5.548258	4.515439	1.299287
std	0.488893	0.513619	0.497160	0.527669	0.311181	0.776874	0.593949	0.471147	0.695520	0.484490	0.888796	3.131593	1.945145	3.091334	2.407883	0.904676
min	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
25%	0.000000	1.000000	0.000000	2.000000	1.000000	2.000000	0.000000	1.000000	1.000000	1.000000	0.000000	1.000000	2.000000	2.000000	3.000000	1.000000
50%	1.000000	1.000000	0.000000	2.000000	1.000000	2.000000	1.000000	1.000000	2.000000	1.000000	0.000000	5.000000	3.000000	7.000000	4.000000	1.000000
75%	1.000000	1.000000	1.000000	2.000000	1.000000	3.000000	1.000000	1.000000	2.000000	1.000000	2.000000	7.000000	4.000000	8.000000	6.000000	2.000000
max	1.000000	2.000000	1.000000	2.000000	2.000000	3.000000	2.000000	2.000000	2.000000	2.000000	2.000000	14.000000	9.000000	10.000000	10.000000	3.000000

Figure 3.3: Data Description of Attribute

The complete procedure of our data collection, data storing, and data preprocessing are given below in Figure 3.3.

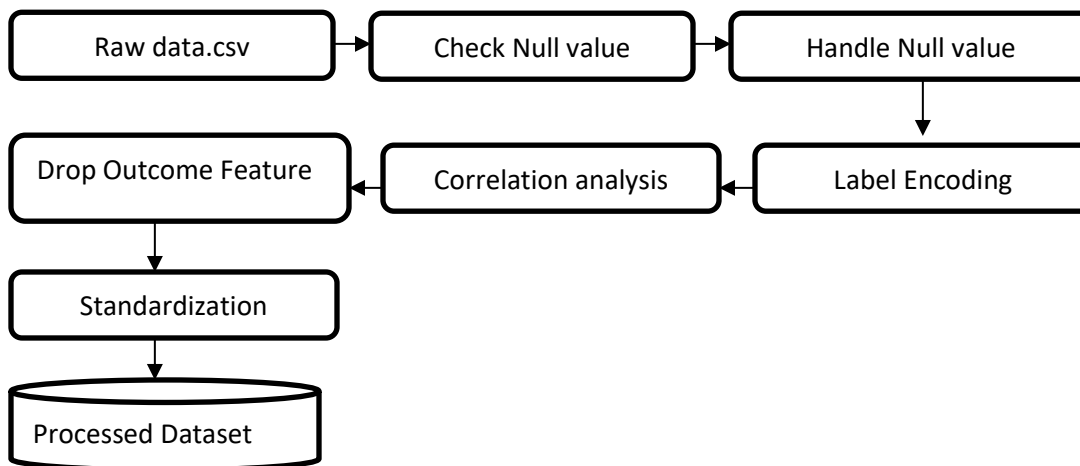


Figure 3.4: Step of Data Preprocessing

This is the complete procedure of our data collection, data storing, and data preprocessing.

### 3.3 Statistical Analysis

Different kind of feature gives different measurement. But often two or more feature can provide similar kinds of data. This can be seen from the Fig 5 visualization of all feature.

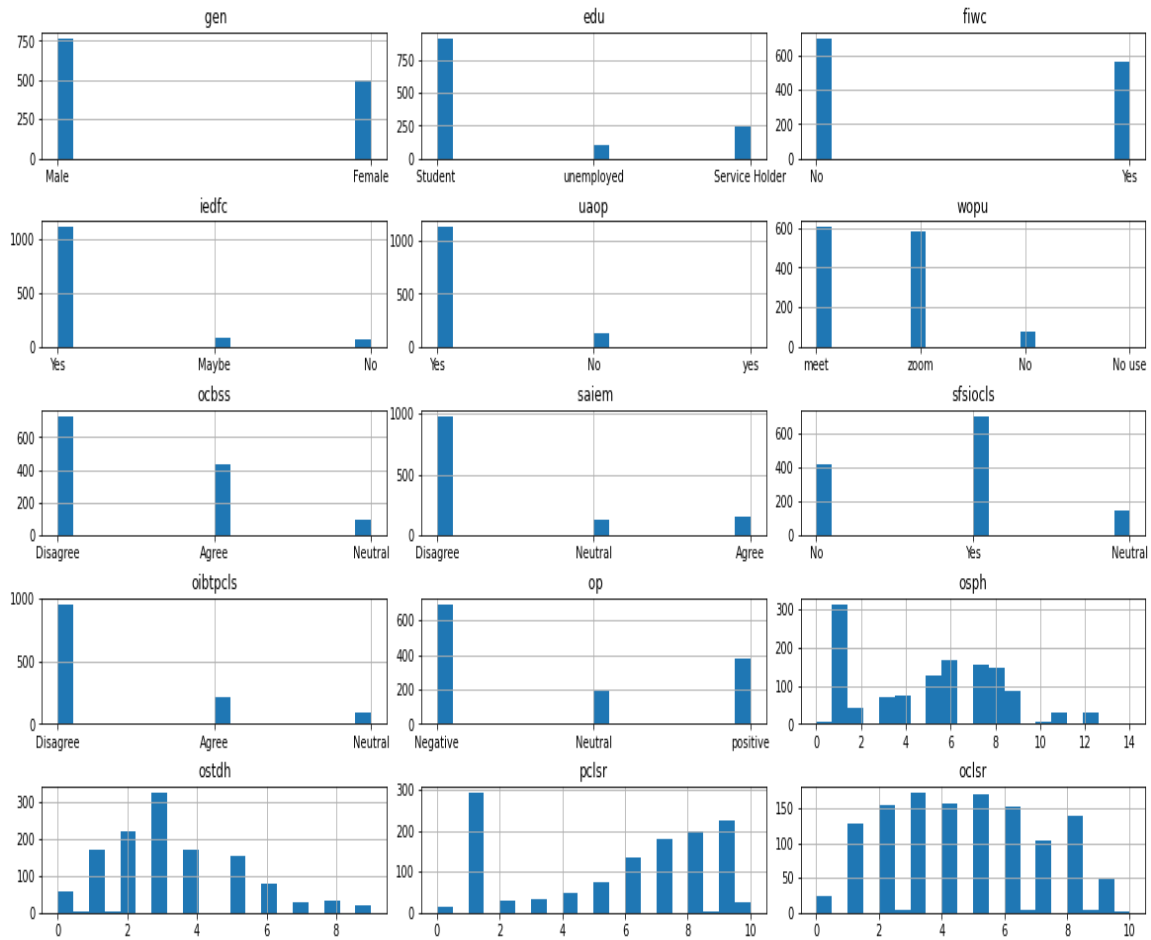


Figure 3.5: Data Visualization for All Feature

The “class” attribute is the dependent attribute in our dataset, while all other attributes are independent attribute that is self-contained, we investigate the relationship between "class" and 15 other characteristics. Using the corr() method is simple.

The step we followed in (Method-2). They are:

- Pearson Correlation technique, Univariate Selection, Random forest classifier
- Feature Elimination technique has been used (The similarities feature chosen).

- Used Algorithm

Pearson Correlation: Calculating the covariance of two variables divided by the product of their standard deviations yields the Pearson Correlation. The test statistic Pearson's correlation coefficient measures the statistical relationship, or association, between two continuous variables. It provides data on the magnitude of the association, or correlation, as well as the relationship's direction [10]. The formula of Pearson correlation coefficient is given below in Equation 1

$$\text{Pearson correlation coefficient, } r = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2 \sum(y_i - \bar{y})^2}} \quad (1)$$

Figure 3.6 shows us the heatmap of correlation between “class” and 15 other attributes shows as given below.

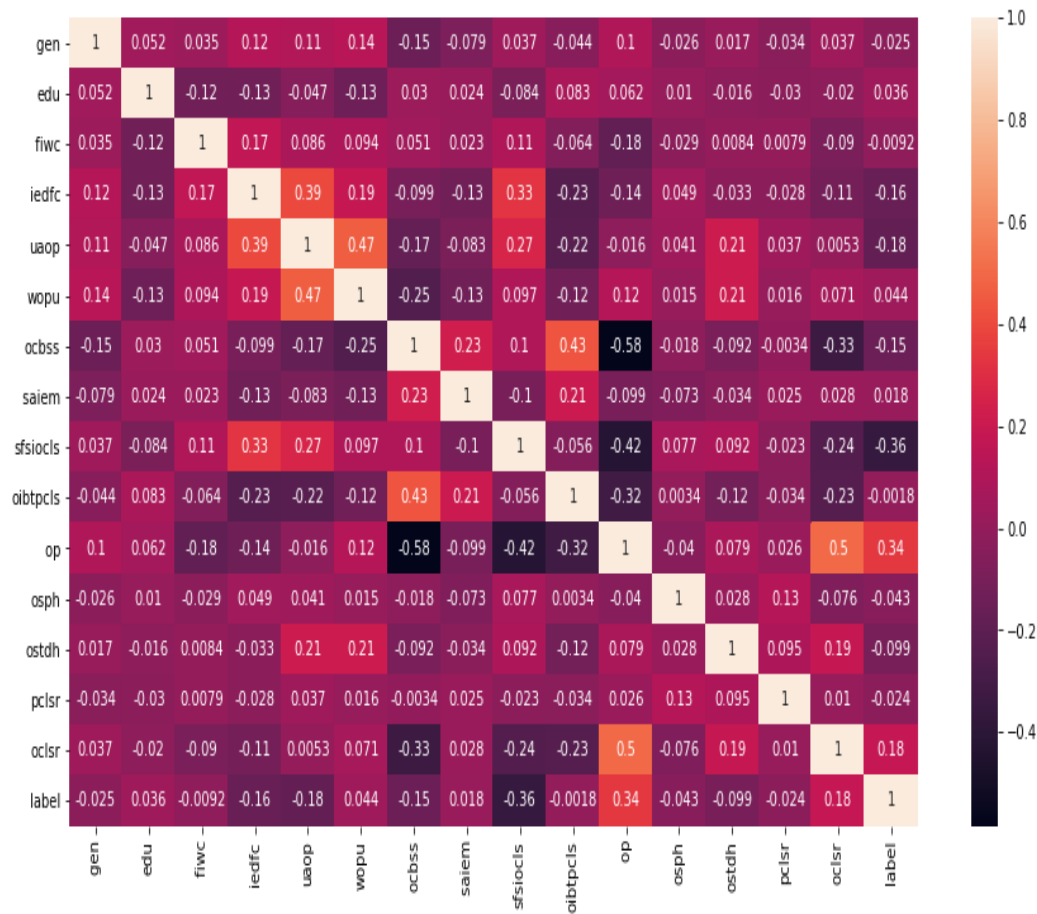


Figure 3.6: Correlation Heatmap

The top 15 features chosen by the Pearson algorithm are shown in Figure 3.7. [10]

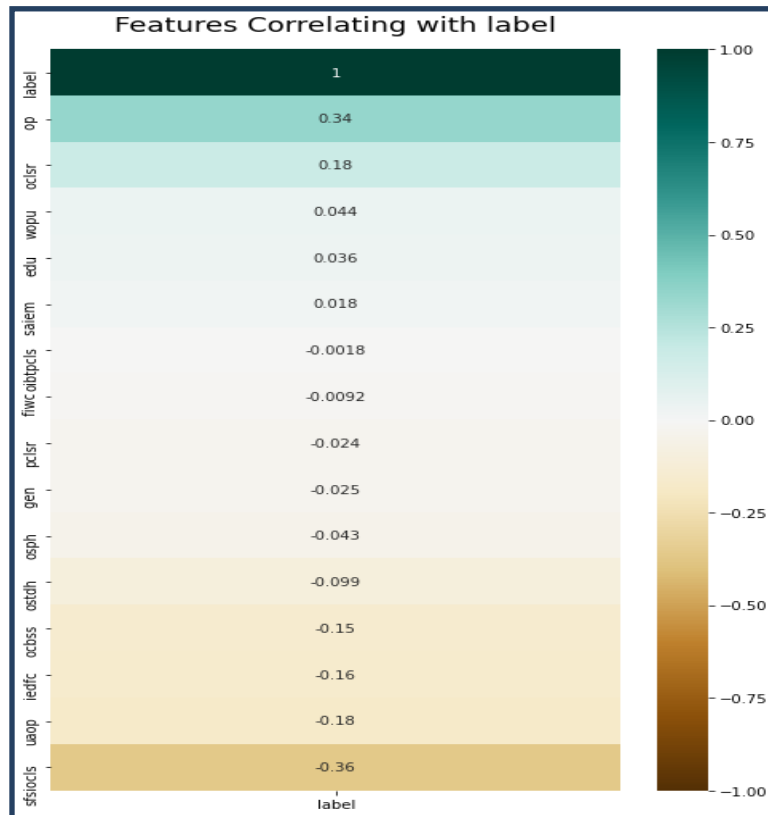


Figure 3.7: Feature Correlation with Label

### Univariate feature selection

The best features are chosen using univariate statistical tests in univariate feature selection. It can be thought of as a preprocessing step in the estimation process. Feature selection routines are exposed by Scikit-learn as objects that implement the transform method: All but the highest-scoring features are removed by SelectKBest [11]. We use `f_classif` technique for our proposal.

`f_classif`: A statistical test based on the Analysis of Variance (ANOVA) that is only used for categorical targets.

Figure 3.8 shows the top 7 features selected by univariate selection.

	Top Features	Score
10	op	121.431747
8	sfsiocls	68.915565
14	oclsr	46.974982
3	iedfc	19.170046
6	ocbss	18.882673
4	uaop	18.416600
2	fiwc	12.454671

Figure 3.8: Univariate Selection Score

### Random Forest classifier

Random Forests are frequently used for feature selection. The reason for this is that random forests' tree-based strategies are naturally ranked by how well they improve node purity. Gini impurity is the average decrease in impurity across all trees [12].

Figure 3.9 shows the top 15 features selected by Random Forest classifier.

```

Feature ranking:
1. Feature ostdh (0.121853)
2. Feature oclsr (0.121386)
3. Feature osph (0.101255)
4. Feature pclsr (0.097607)
5. Feature op (0.079909)
6. Feature sfsiocls (0.075239)
7. Feature wopu (0.067352)
8. Feature ocbss (0.059780)
9. Feature oibtpcls (0.047649)
10. Feature gen (0.045683)
11. Feature edu (0.043252)
12. Feature fiwc (0.041120)
13. Feature saiem (0.040213)
14. Feature iedfc (0.030712)
15. Feature uaop (0.026989)

```

Figure 3.9: Random Forest Ranking

From the above comparative visualization each activity We've worked on among the three most common features. The final features top 6 features will be 'edu', 'wopu', 'sfsiocls', 'op', 'osph', 'oclsr'

Figure 3.10 shows the combined feature selection.

```
['edu', 'wopu', 'sfsiocls', ' op', 'osph', 'oclsr', 'label']
```

Figure 3.10: Final Feature Selection

So, we can clearly say that is our final feature which we select from the similarities of three feature.

### 3.4 Proposed Methodology

Machine learning are used to predict the pressure of a student from the dataset. Algorithms are used to train these datasets. After that we have achieved a different variety of accuracy for different algorithms on the testing dataset. These algorithms we have implemented are following:

- a. K-Nearest Neighbors
- b. Naive Bayes
- c. Decision Tree
- d. Random Forest
- e. Xgb Boost
- f. Support Vector Machine
- g. Neural Networks

Our proposed methodology is given below in “Figure 3.11”

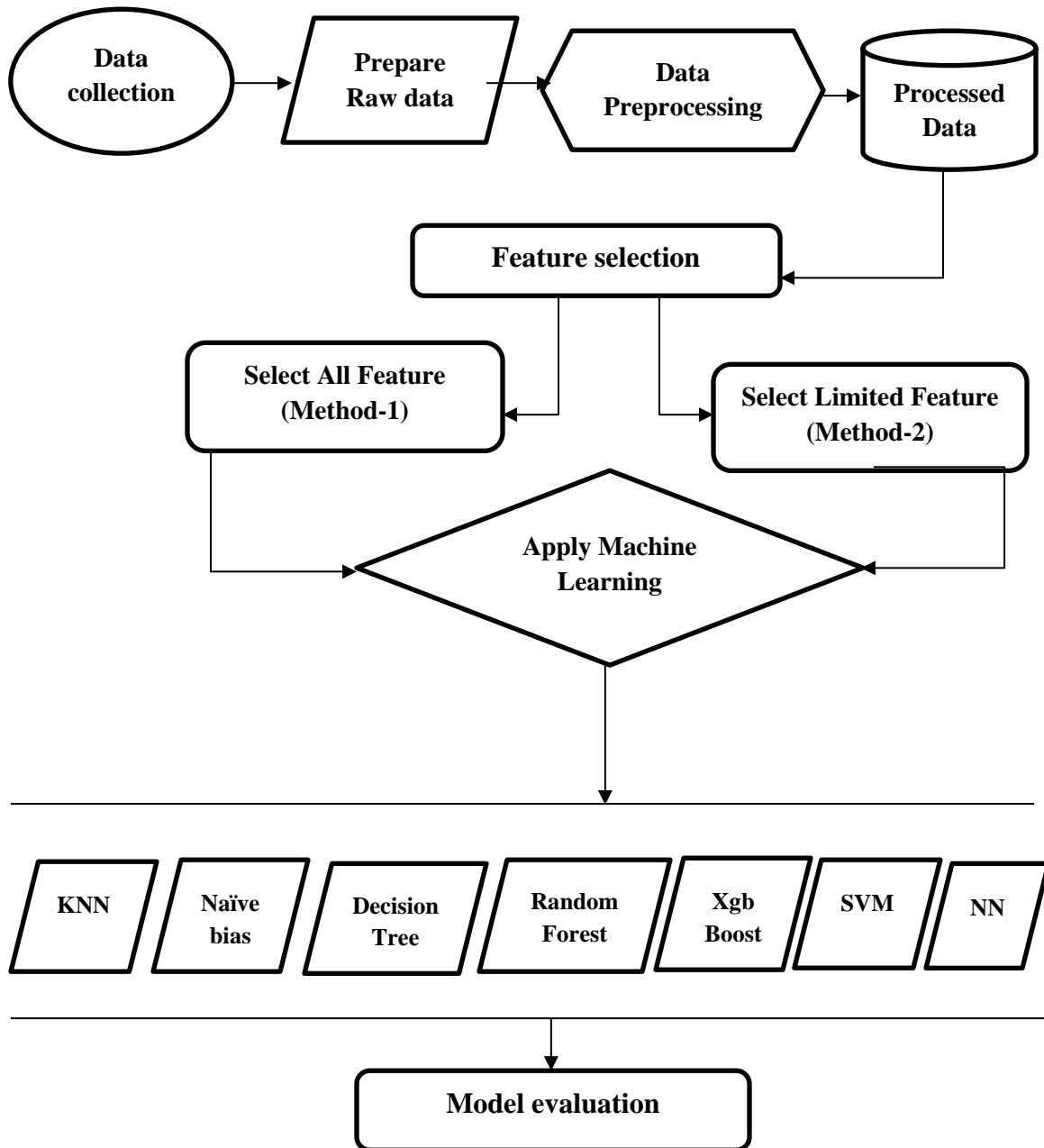


Figure 3.11: Illustrate Our Proposed Methodology

a. K-Nearest Neighbors (KNN):

The algorithm Worked in the first phase of our project is KNN. We used this algorithm for supervised machine learning. This is a very simple algorithm and helps to give a very good shape. This algorithm is mainly used to measure distances. To make a KNN algorithm successful, you have to go through three stages[13]. The steps are given below-

- Determine the Kth number value.
- Determine the Euclidean distance.
- Select the K nearest neighbors.

For the KNN algorithm we first need to determine the value of K, which will find the nearest neighbor. So, we can find K's lowest error rate as shown in the “Figure-3.12,3.14” below:

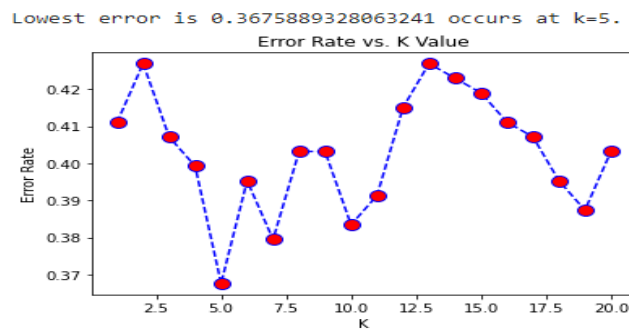


Figure 3.12: Error Rate for KNN Algorithm for All Feature

Here This algorithm gave us 63 percent accuracy results for all feature.

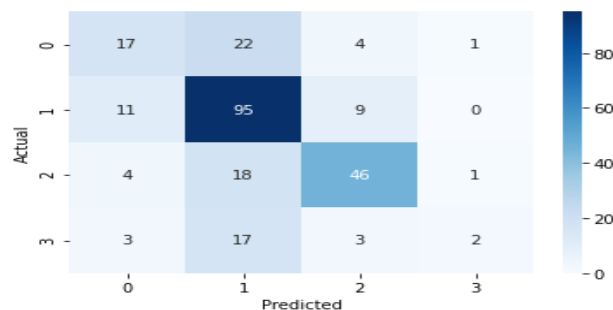


Figure 3.13: Confusion Matrix for KNN Algorithm for Selecting All Feature



Here This is our have confusion matrix for selecting all Feature to measure the performance of our trained model on the tasing dataset.

In this case the error rate of KNN is 0.36 that means if we determine the value of k in form of error rate than the value of k = 5

Here error rate for limited feature,

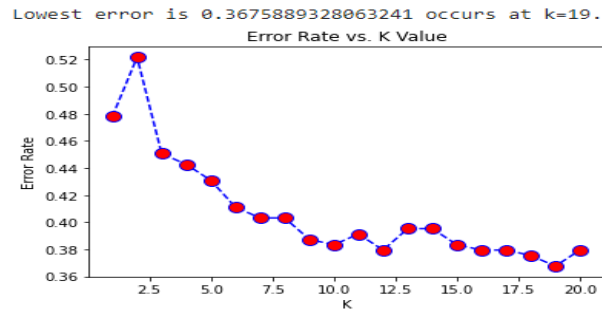


Figure 3.14: Error Rate for KNN Algorithm for Limited Feature

In this case the error rate of KNN is 0.36 that means if we determine the value of k in form of error rate than the value of k = 19

This algorithm gave us 63 percent accuracy results for limited feature.

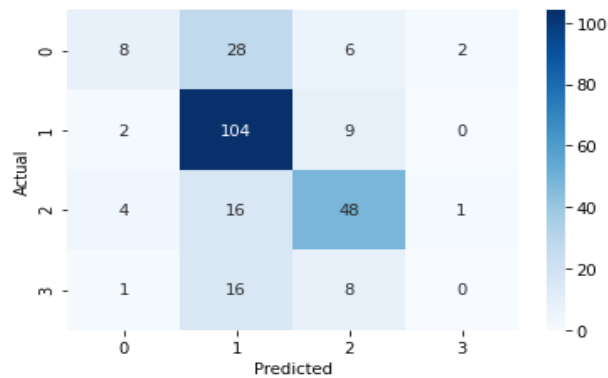


Figure 3.15: Confusion Matrix for KNN Algorithm for Selecting Limited Feature

Here This is our confusion matrix for selecting all Feature to measure the performance of our trained model on the tasing dataset.

KNN Can be diagnosed through another method Which is called Square Root. The distance used to calculate the straight line between two vectors is the Euclidean distance.

Equation-1 given below:

$$\text{Euclidean Distance } d(x, y) = \sqrt{\sum_{i=1}^m (x_i - y_i)^2} \text{ --- (2)}$$

We have to installed its library to use the KNN algorithm and the library is

**from sklearn.neighbors import KNeighborsClassifier**

b. Naive Bayes: The algorithm used in the second phase of our project is Naive bias. This algorithm is mainly used to probability analysis. To make a Naïve bias algorithm successful, you have to go through three stages[13]. The steps are given below-

- Measure probability for class labels.
- calculate probability attribute for each class.
- Use Bayes Formula.

Probability used to calculate this formulation-

Naïve bias in Equation-2 is given below:

$$P(A|B) = \frac{P(A|B).P(A)}{P(B)} \text{ --- (3)}$$

We have to installed its library to use the Naïve bias algorithm and the library is

**from sklearn import naive\_bayes.**

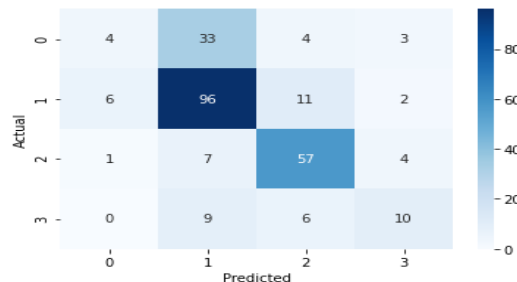


Figure 3.16: Confusion Matrix for Naïve Bias Algorithm for Selecting All Feature

This algorithm gave us 66 percent accuracy results for all feature.

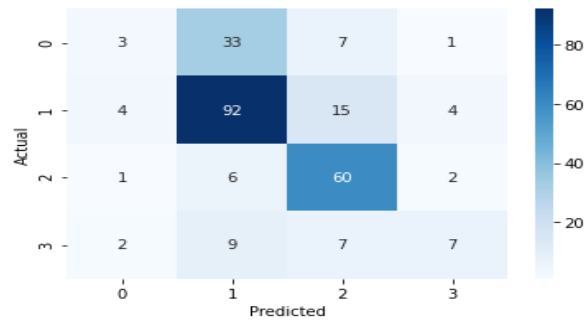


Figure 3.17: Confusion Matrix for Naïve Bias Algorithm for Selecting Limited Feature

This algorithm gave us 64 percent accuracy results for limited feature.

c. Decision tree: The algorithm used in the third phase of our project is Decision Tree. This algorithm is mainly used for decision making. To make a Decision Tree algorithm successful, you have to go through several stages[13]. The steps are given below-

- problem defines.
- Model process.
- Apply probability values.
- Solve the decision tree.
- Perform analysis.
- List assumptions

The error rate of Decision tree is given below “Figure-3.18,3.20” given below:

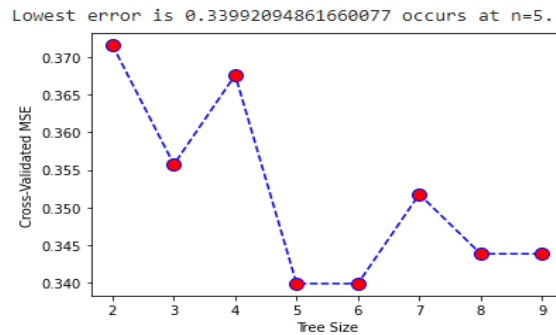


Figure 3.18: Error Rate for Decision Tree Algorithm for All Feature

In this case the error rate of decision tree is 0.33 that means if we determine the value of n in form of error rate than the value of n = 5.

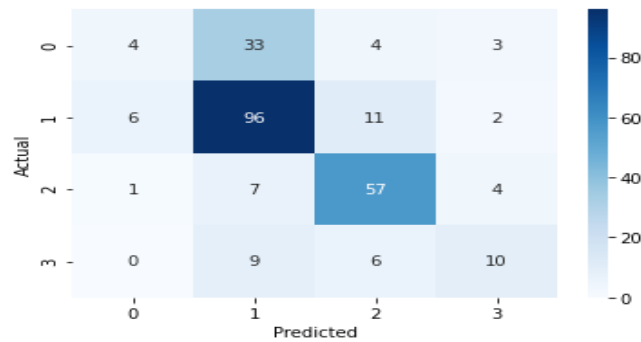


Figure 3.19: Confusion Matrix for Decision Tree Algorithm for Selecting All Feature

This algorithm gave us 66 percent accuracy results for all feature.

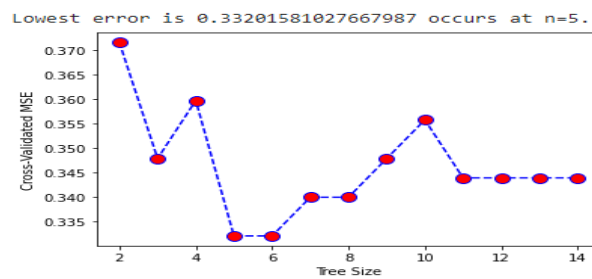


Figure 3.20: Error Rate for Decision Tree Algorithm for Limited Feature

In this case the error rate of decision tree is 0.33 that means if we determine the value of n in form of error rate than the value of n = 5.

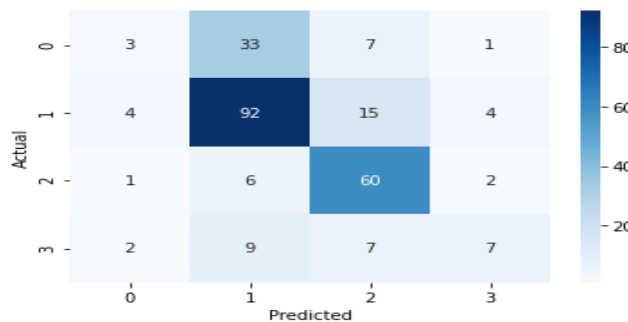


Figure 3.21: Confusion Matrix for Decision Tree Algorithm for Selecting Limited Feature

This algorithm gave us 64 percent accuracy results for limited feature.

Here This is our have confusion matrix for selecting Limited Feature to measure the performance of our trained model on the tasing dataset.

We have to installed its library to use the Decision Tree algorithm and the library is

**from sklearn import tree**

d. Random Forest (RF)

The algorithm used in the fourth phase of our project is Random Forest. We used this algorithm for classification and regression analysis in machine learning. This is a very simple algorithm and helps to give a very good shape. This algorithm is mainly generating a tree and make its decision. It helps to make multiple decisions and helps to get good accurate predictions. It creates a lot of trees. Its elimination overfitting problem. By doing that it improves its accuracy.[13] .The error rate of Random forest is given below “Figure-3.22,3.24” given below:

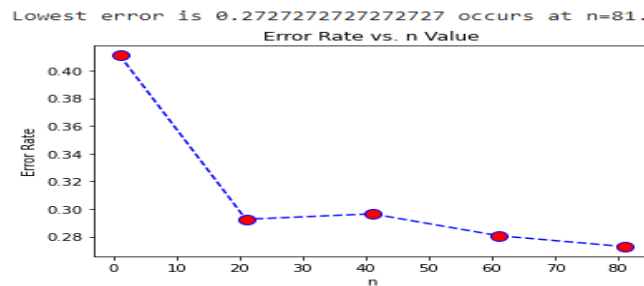


Figure 3.22: Error Rate for Random Forest Algorithm for Using All Feature

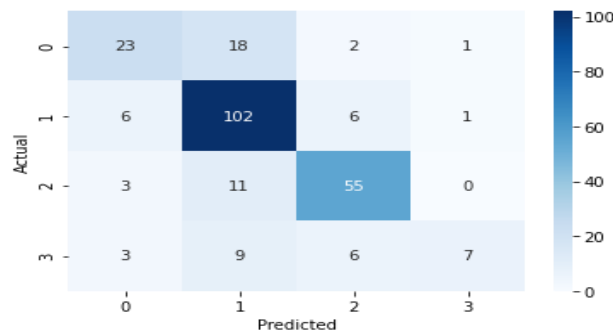


Figure 3.23: Confusion Matrix Map for Random Forest Algorithm for Using All Feature

Here This is our have confusion matrix for selecting all Feature to measure the performance of our trained model on the tasing dataset.

This algorithm gave us 74 percent accuracy results for all feature.

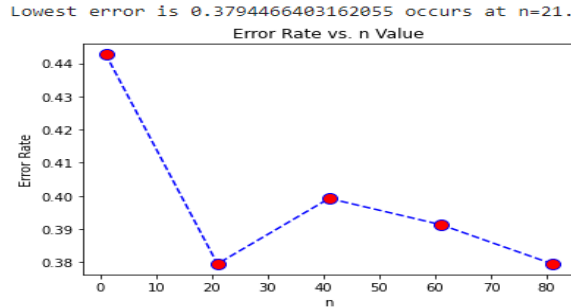


Figure 3.24: Error Rate for Random Forest Algorithm for Using Limited Feature

In this case the error rate of Random forest is 0.37 that means if we determine the value of n in form of error rate than the value of n = 21.

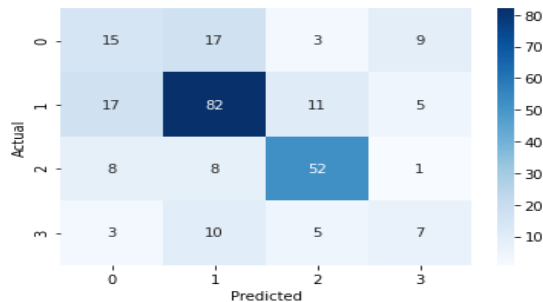


Figure 3.25: Confusion matrix for Random Forest Algorithm for selecting Limited Feature

Here This is our have confusion matrix for selecting Limited Feature to measure the performance of our trained model on the tasing dataset.

This algorithm gave us 62 percent accuracy results for Limited feature.

We have to installed its library to use the Random Forest algorithm and the library is from sklearn import ensemble

e. Xgb Boost

The algorithm that we used in the fifth phase of our project is boosting algorithm. This algorithm is mainly used for performance making[13]. The Confusion matrix of Xgb Boost is given below “Figure-3.26,3.27” given below:

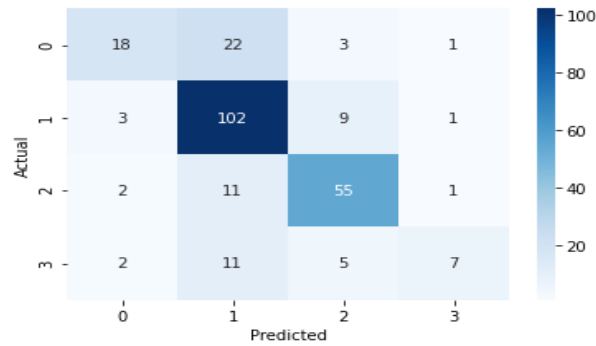


Figure 3.26: Confusion Matrix Map for Xgb Boost Algorithm for Using All Feature

This algorithm gave us 72 percent accuracy results for all feature.

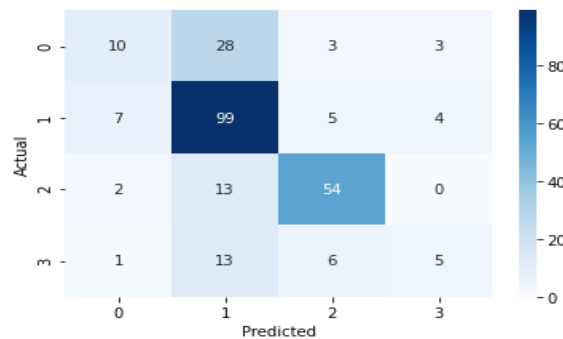


Figure 3.27: Confusion Matrix For Xgb Boost Algorithm For Selecting Limited Feature

This algorithm gave us 66 percent accuracy results for limited feature.

We have to installed its library to use the xgboost algorithm and the library is

**import xgboost**

Support Vector Machine (SVM)

The algorithm that we used in the sixth phase of our project is Support Vector Machine. This algorithm is mainly used divides n-dimensional space[13].

The error rate of svm is given below “Figure-3.28,3.30” given below:

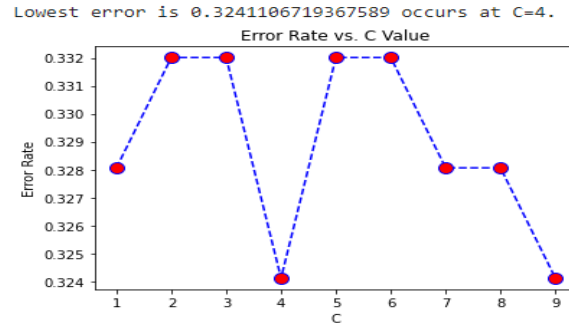


Figure 3.28: Error Rate for SVM Algorithm for All Feature

In this case the error rate of svm is 0.32 that means if we determine the value of c in form of error rate than the value of  $c = 4$ .

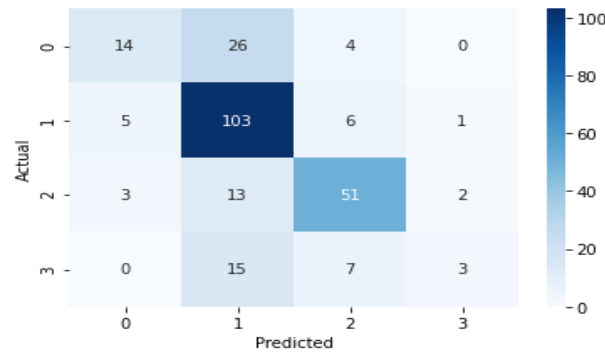


Figure 3.29: Confusion Matrix for SVM Algorithm for Selecting All Feature

This algorithm gave us 68 percent accuracy results for all feature.

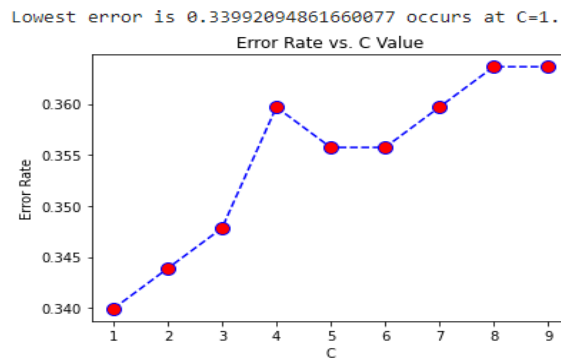


Figure 3.30: Error Rate for SVM Algorithm for Limited Feature



In this case the error rate of svm is 0.33 that means if we determine the value of c in form of error rate than the value of  $c = 1$ .

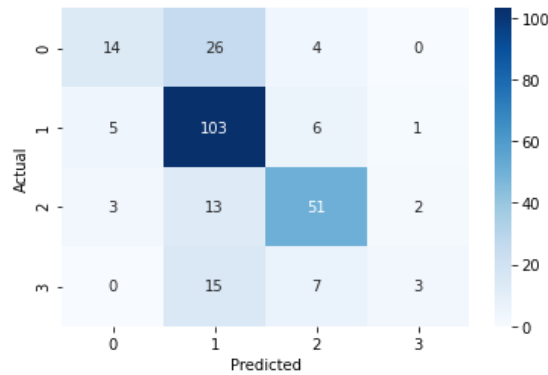


Figure 3.31: Confusion Matrix for SVM Algorithm for Selecting Limited Feature

This algorithm gave us 66 percent accuracy results for limited feature.

Here This is our have confusion matrix for selecting Limited Feature to measure the performance of our trained model on the tasting dataset.

We have to installed its library to use the svm algorithm and the library is

**from sklearn import svm**

f. Neural Network: The algorithm used in the seventh phase of our project is Neural Network. This algorithm is mainly used for connect together[13].

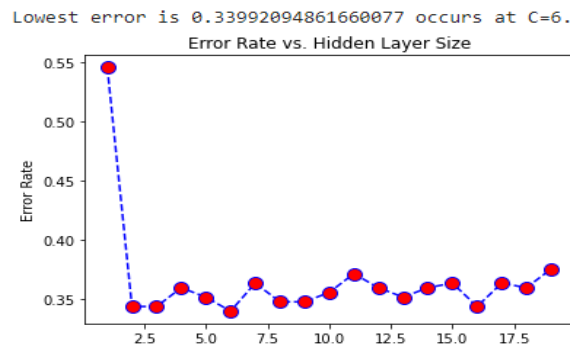


Figure 3.32: Error Rate Neural Network Algorithm All Feature

In this case the error rate of NN is 0.33 that means if we determine the value of c in form of error rate than the value of  $c = 6$ .

This algorithm gave us 63 percent accuracy results for selecting all feature.

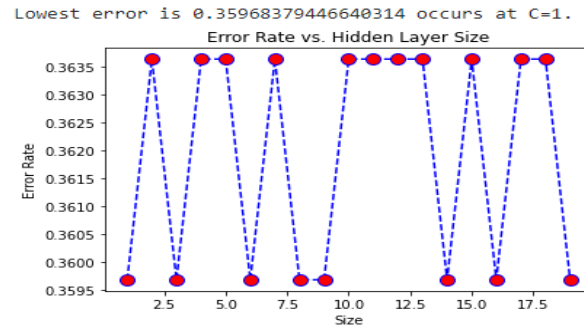


Figure 3.33: Error Rate Neural Network Algorithm Limited Feature

In this case the error rate of NN is 0.35 that means if we determine the value of c in form of error rate than the value of  $c = 1$ .

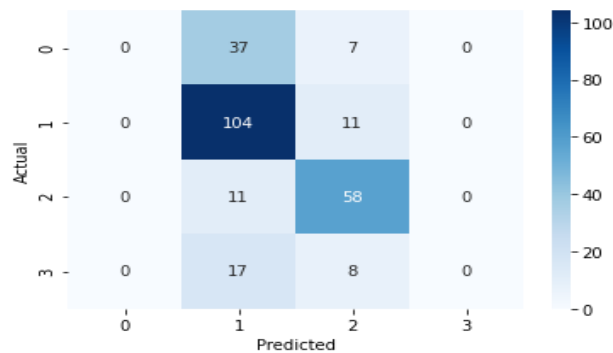


Figure 3.34: Confusion Matrix for NN Algorithm for Selecting Limited Feature

This algorithm gave us 64 percent accuracy results for Limited feature. We have to installed its library to use the NN algorithm and the library is

**from sklearn import neural\_network**

### 3.5 Implementation Requirements

- Operating System - Windows
- Environment -Google Colab with Python version 3.7.1
- Data Preprocessing- Pandas, NumPy
- Data Visualization -Seaborn Matplotlib for graph plotting s
- Score Showing- classification report, accuracy score, confusion Matrix,
- Algorithm -from sklearn
  - a. K-Nearest Neighbors (KNN)
  - b. Naïve Bias
  - c. Decision Tree
  - d. Random Forest Tree
  - e. Xgb Boost
  - f. Support Vector Machine (SVM)
  - g. Neural Network

## **CHAPTER 4**

### **EXPERIMENTAL RESULTS AND DISCUSSION**

#### **4.1 Experimental Setup**

Our main purpose of this experiment was to Predict a student's stress on online class during Covid-19 situation in Bangladesh. Names of Four activities are Mental, No problem, financial, Physical.

To get better results we use three different Algorithm based on Machine Learning.

We used a 1,264-sampling data in for our project. As a result, many studies have been conducted in recent years to examine the stress of an online class for a student. By putting our model to the test to see how well our data predicted the online class problems they faced during the pandemic. Our dataset was analyzed using a machine learning approach. We followed every workflow to get results in order to predict our dataset and implement various algorithms to test our model.

The next part of the test was to process the data and implement the model, we have used Python version 3.7.1 in Google Colab Environment. To implement the models, we used Python's library functions form sklearn: KNeighborsClassifier, Neighbors, naive\_bayes, ensemble, svm, xgb Boost, Decision tree, Neural Network in addition we have used some libraries like Pandas, NumPy, label encoder etc. for data preprocessing purposes.

So, after getting the above experimental setup, we did various experiments to get the desired result.

## 4.2 Experimental Results and Analysis

Now a day, some research has been done their survey as their online pressure Measurement. By testing our model to show how accurately our dataset can Predict students stress in online class. We used seven algorithms to test our model. We are following the above workflow for getting Result which is given by Figure-4.1:

We loaded 80% of the data onto the train and 20% onto the test to see how well the model performed. The best accuracy comes from random forest (using all feature) and Decision Tree algorithms (Limited feature).

The models' performance is summarized in the table below.

### BY SELECTING ALL FEATURE (METHOD-1)

Table 4.1: Performance Table For KNN Algorithm All Feature Selection

	precision	Recall	F1-Score	Support
<b>Mental</b>	0.49	0.39	0.43	44
<b>NO Problem</b>	0.62	0.83	0.71	115
<b>Financial</b>	0.74	0.67	0.7	69
<b>Physical</b>	0.5	0.8	14	25

<b>Accuracy</b>			0.63	253
<b>Macro Avg</b>	0.49	0.49	0.5	253
<b>Weighted Avg</b>	0.62	0.63	0.6	253

Table 4.2: Performance Table For Random Forest Algorithm All Feature Selection

	<b>precision</b>	<b>Recall</b>	<b>F1-Score</b>	<b>Support</b>
Mental	0.66	0.52	0.58	44
NO Problem	0.73	0.89	0.8	115
Financial	0.8	0.8	0.8	69
Physical	0.78	0.28	0.41	25

<b>Accuracy</b>			0.74	253
<b>Macro Avg</b>	0.74	0.62	0.65	253
<b>Weighted Avg</b>	0.74	0.74	0.67	253

Table 4.3: Performance Table For Xgb Boost Algorithm For All Feature Selection

	<b>precision</b>	<b>Recall</b>	<b>F1-Score</b>	<b>Support</b>
Mental	0.72	0.41	0.52	44
NO Problem	0.7	0.89	0.78	115
Financial	0.76	0.8	0.78	69
Physical	0.7	0.28	0.4	25

<b>Accuracy</b>			0.72	253
<b>Macro Avg</b>	0.72	0.59	0.62	253
<b>Weighted Avg</b>	0.72	0.72	0.57	253

From Table-4.1,4.2,4.3 above we can see that Random Forest Algorithm has shown a near-perfect performance achieve 74% accuracy. We have also trained some more

algorithm by selecting all feature to test our data set which is given below by “Table-4.3”

Table 4.4: Accuracy Table for Different Algorithm (All Feature)

By Selecting All Feature														
Algorithm	KNN		Naïve Bias		Decision Tree		Random Forest Tree		Xgb Boost		Svm		NN	
Activities	F1-Score	accuracy	F1-Score	accuracy	F1-Score	accuracy	F1-Score	accuracy	F1-Score	accuracy	F1-Score	accuracy	F1-Score	accuracy
Mental	0.43	63%	0.15	66%	0.27	66%	0.58	74%	0.52	72%	0.42	68%	0	64%
No problem	0.71		0.74		0.74		0.8		0.78		0.78			
Financial	0.7		0.78		0.75		0.8		0.78		0.62			
Physical	0.14		0.45		0		0.41		0.4		0			

From Here we can see our best accuracy given By Random Forest Algorithm Which is 74 percent accuracy.

### USING LIMITED FEATURE (METHOD-2)

Table 4.5: Performance table KNN Algorithm for Limited feature selection

	precision	Recall	F1-Score	Support
<b>Mental</b>	0.53	0.18	0.27	44
<b>NO Problem</b>	0.63	0.9	0.75	115
<b>Financial</b>	0.68	0.7	0.69	69
<b>Physical</b>	0	0	0	25

<b>Accuracy</b>			0.63	253
<b>Macro Avg</b>	0.46	0.45	0.43	253
<b>Weighted Avg</b>	0.57	0.63	0.57	253

Table 4.6: Performance Table for Random Forest Algorithm For Limited Feature Selection

	<b>precision</b>	<b>Recall</b>	<b>F1-Score</b>	<b>Support</b>
Mental	0.35	0.34	0.34	44
NO Problem	0.7	0.71	0.71	115
Financial	0.73	0.75	0.74	69
Physical	0.32	0.28	0.3	25

<b>Accuracy</b>			0.62	253
<b>Macro Avg</b>	0.53	0.52	0.52	253
<b>Weighted Avg</b>	0.61	0.62	0.61	253

Table 4.7: Performance Table For Xgb Boost Algorithm For Limited Feature Selection

	<b>precision</b>	<b>Recall</b>	<b>F1-Score</b>	<b>Support</b>
Mental	0.5	0.23	0.31	44
NO Problem	0.65	0.86	0.74	115
Financial	0.79	0.78	0.79	69
Physical	0.42	0.2	0.27	25

<b>Accuracy</b>			0.66	253
<b>Macro Avg</b>	0.59	0.52	0.53	253
<b>Weighted Avg</b>	0.64	0.66	0.63	253

From Table-4.5,4.6,4.7 above we can see that Xgb Boost Algorithm has shown a near-perfect performance achieve 66% accuracy. We have also trained some more algorithm by selecting limited feature to test our data set which is given below by “Table-4.8”.



Table 4.8: Accuracy Table for Different Algorithm (Limited Feature)

By Selecting Limited Feature														
Algorithm	KNN		Naïve Bias		Decision Tree		Random Forest Tree		Xgb Boost		Svm		NN	
Activities	F1-Score	accuracy	F1-Score	accuracy	F1-Score	accuracy	F1-Score	accuracy	F1-Score	accuracy	F1-Score	accuracy	F1-Score	accuracy
Mental	0.27	63%	0.11	64%	0.27	64%	0.34	62%	0.31	66%	0.27	66%	0	64%
No problem	0.75		0.72		0.74		0.71		0.74		0.73			
Financial	0.69		0.76		0.79		0.74		0.79		0.76			
Physical	0		0.36		0		0.3		0.27		0			

In (Method-2) Xgb Boost Algorithm has achieved 66% accuracy. By using Dimensionality Reduction (Method-2) accuracy will be improved In Future.

### 4.3 Discussion

By examining the table above, we can see that during the Covid-19 period, the online education system produced a variety of results for students. We had a total of 15 features in our dataset, but we used feature ranking to rank them using a random forest classifier. Then we use the feature number seven. For a better result, feature ranking is used. In our experiment, we used a machine learning algorithm. The random forest is given a score of 74 percent when all features are used (Method-1), and the Xgb Boost is given a score of 66 percent accuracy in limited feature (Method-2). We have found our best accuracy from Method-1 that means using all feature we find our best accuracy by Random Forest. By including more data, and dimensionality Reduction (Method-2) accuracy will be improved. We will continue to work on this feature to ensure that its accuracy remains at least 2%.

## **CHAPTER 5**

### **IMPLICATION ON SOCOTY, ENVIRONMEN AND SUSTAINABILITY**

#### **5.1 Impact on Society**

The impact of this project will be significant on society. This project will work for Society in a variety of ways. If the study is stopped in such a catastrophic situation including Corona, then the information will be available about the impact on the education system of Bangladesh. As a result, people in society can know about their mental condition and mental health in the event of a disaster. It helps in determining the way of life of society. With the help of this project, it will make life easier by solving some problems of society. So obviously this project will have a much bigger impact on society.

#### **5.2 Impact on Environment**

About 80% of all projects in the world have an impact on the environment. And every project goes in a certain direction. When you do a job, everything you use, such as the power supply, has an adverse effect on the environment. However, our project will not have any impact on the environment, but this research will have a positive impact on the environment. No gas, no carbon emissions, no heating, no chemicals, no power supplies. As a result, it does not affect the environment in any way. This project will have zero impact on the environment, without any harm to the students of Bangladesh will find out the impact of the impact on education under the circumstances.

#### **5.3 Ethical Aspects**

For our project, ordinary students in Bangladesh needed to know their information. This is the data of our project. Using this data algorithm, we find out the desired outcome. We need to be aware that we are using general student data. This includes a solution and protection for personal data privacy. We need to let them know how much data we're taking and what kind of data we're taking. Personal information cannot be used without authorization. When we took data from Bangladeshi students, they did

not have any personal information. No emails, personal names, or personal addresses were taken from them. No data was shared with anyone or anyone outside our research team except us. We must strictly maintain the privacy of everyone here

#### **5.4 Sustainability Plan**

Our project will be very useful for human work. Study breaks down, especially during disasters. At present, a lot has advanced in the welfare of technology. So even using technology, you can continue your studies at home during this disaster without going directly to class. However, students suffer from mental problems. As a result, work will be done to solve this later. However, we will be able to find out about the impact of corona time on education. From this project, we will be able to know about the impact on corona time education, about the impact on education. Our project can be made more effective later with more data and using different types of applications. In addition, the work will be done in accordance with the guidelines to make the work more interesting and effective. The model will increase the exposure more and the effectiveness will be explained to the people so that the people will give more information and our project will come closer.

## CHAPTER 6

### SUMMARY, CONCLUSION, RECOMMENDATION, AND IMPLICATION FOR FUTURE RESEARCH

#### 6.1 Summary of The Study

Basically, our goal was to create a model where Corona would know-how and to what extent it has affected the education system in Bangladesh. We have collected data from students in Bangladesh for our project. We pre-processed the data we collected, cleaned the data, deleted the lost data values. We have selected features that had an impact on the education system in Bangladesh. We have often used seven machine learning algorithms, the most notable of which are K-Nearest Neighbors (KNN), Naïve Bias, Random Forest, Xgb boost, Neural Network, etc. Best accuracy 74% by using all feature values on Random forest Algorithm and on the other hand by using limited (7) feature we get 66% for Xgb Boost algorithm accuracy.

#### 6.2 Conclusion

This is our project for COVID-19 which has a big impact on the education sector in Bangladesh. Online classes have been conducted during the coronation period, which has had a kind of effect on the brains of the students. This study covers a large part of everyday life. So it is natural to understand how important it is to calculate this effect. The solution can be found by judging the impact results on education for Corona.

We've made the data we've processed usable. And used about seven algorithms of machine learning. Algorithms It was very important to accurately identify the appropriate algorithm to use the data then apply the data accurately to the algorithm. Many algorithms were used here. Including we use K-Nearest Neighbors (KNN), Naive Bias, Random forest, Xgb boost, Neural Network, Decision Tree, etc. various traditional algorithms have been applied to our dataset. Best accuracy 74% by using all feature values on Random forest the other hand by using limited (7) feature we get 66% for Xgb Boost accuracy. Different types of algorithms were used here but all were done using the same data. We have learned a lot from this data collection in data

processing data sorting and one applied task. I have also learned a lot for the future by working with machine learning. Here we have learned a lot about machine learning and pedicure which will help us to do better in the future.

### **6.3 Implication for Further Study**

We had a limitation in this work especially the limitations is that our data volume was very low because we collected data using Google form. Collecting data from people through Google Chrome is really very challenging so If we could collect more data, our accuracy would increase a lot. Our future plan is that we will collect a lot more data for our project in the future. This will result in a lot of data and Data will actually increase our accuracy level and we will do different types of work to make our project more effective and use a special type of application to use it normally.

## REFERENCE

- [1] S. Singh, D. H. Rylander, and T. C. Mims, "Efficiency of Online vs. Offline Learning: A Comparison of Inputs and Outcomes," *Int. J. Business, Humanit. Technol.*, vol. 2, no. 1, pp. 93–98, 2012, [Online]. Available: [http://ijbhtnet.com/journals/Vol\\_2\\_No\\_1\\_January\\_2012/12.pdf](http://ijbhtnet.com/journals/Vol_2_No_1_January_2012/12.pdf).
- [2] S. Singh, D. H. Rylander, and T. C. Mims, "Efficiency of Online vs. Offline Learning: A Comparison of Inputs and Outcomes," *Int. J. Business, Humanit. Technol.*, vol. 2, no. 1, pp. 93–98, 2012, [Online]. Available: [http://ijbhtnet.com/journals/Vol\\_2\\_No\\_1\\_January\\_2012/12.pdf](http://ijbhtnet.com/journals/Vol_2_No_1_January_2012/12.pdf).
- [3] D. S. Bestiantono, P. Z. R. Agustina, and T.-H. Cheng, "How Students' Perspectives about Online Learning Amid the COVID-19 Pandemic?," *Stud. Learn. Teach.*, vol. 1, no. 3, pp. 133–139, 2020, doi: 10.46627/silet.v1i3.46.
- [4] A. Abidah, H. N. Hidaayatullaah, R. M. Simamora, D. Fehabutar, and L. Mutakinati, "The Impact of Covid-19 to Indonesian Education and Its Relation to the Philosophy of 'Merdeka Belajar,'" *Stud. Philos. Sci. Educ.*, vol. 1, no. 1, pp. 38–49, 2020, doi: 10.46627/sipose.v1i1.9.
- [5] K. Chaturvedi, D. K. Vishwakarma, and N. Singh, "COVID-19 and its impact on education, social life and mental health of students: A survey," *Child. Youth Serv. Rev.*, vol. 121, no. December 2020, p. 105866, 2021, doi: 10.1016/j.chilyouth.2020.105866.
- [6] S. Pokhrel and R. Chhetri, "A Literature Review on Impact of COVID-19 Pandemic on Teaching and Learning," *High. Educ. Futur.*, vol. 8, no. 1, pp. 133–141, 2021, doi: 10.1177/2347631120983481.
- [7] E. K. H. Emon, A. R. Alif, and M. S. Islam, "Impact of COVID-19 on the Institutional Education System and its Associated Students in Bangladesh," *Asian J. Educ. Soc. Stud.*, vol. 11, no. 2, pp. 34–46, 2020, doi: 10.9734/ajess/2020/v11i230288.
- [8] H. Simon and P. Yatrakis, "The Effect of Self-selection on Student Satisfaction and Performance in Online Classes," *Int. Rev. Res. Open Distrib. Learn.*, vol. 3, no. 2, Oct. 2002, doi: 10.19173/irrodl.v3i2.93.
- [9] W. Villegas-ch and X. Palacios-pacheco, "applied sciences Improvement of an Online Education Model with the Integration of Machine Learning and Data Analysis in an LMS," 2020.
- [10] S. Kumar and I. Chong, "Correlation analysis to identify the effective data in machine learning: Prediction of depressive disorder and emotion states," *Int. J. Environ. Res. Public Health*, vol. 15, no. 12, 2018, doi: 10.3390/ijerph15122907.
- [11] Y. Ghfrudwlrq, I. Sulfh, and V. Duudqjhpqhqw, "\$ 8Qlyduldwh ) Hdwxuh 6Hohfwlrq \$ Ssurdfk Iru ) Lqglqj . H \ ) Dfwrub Ri 5Hvwdxudqw % Xvlqhv," vol. 7, pp. 3–8.
- [12] Y. Zhou and G. Qiu, "Random forest for label ranking," *Expert Syst. Appl.*, vol. 112, pp. 99–109, 2018, doi: 10.1016/j.eswa.2018.06.036.
- [13] T. O. Ayodele, "Atherosclerotic Cardiovascular Disease," *Atheroscler. Cardiovasc. Dis.*, 2012, doi:

10.5772/711.

- [14] E. M. Onyema, “Impact of Coronavirus Pandemic on Education,” *J. Educ. Pract.*, vol. 11, no. 13, pp. 108–121, 2020, doi: 10.7176/jep/11-13-12.
- [15] A. Schleicher, “The impact of COVID-19 on education: Insights from education at a glance 2020,” *OECD J. Econ. Stud.*, pp. 1–31, 2020, [Online]. Available: <https://www.oecd.org/education/the-impact-of-covid-19-on-education-insights-education-at-a-glance-2020.pdf>.
- [16] S. Dawadi, R. A. Giri, and P. Simkhada, “Impact of COVID-19 on the Education Sector in Nepal: Challenges and Coping Strategies Saraswati Dawadi 1 , Ram Ashish Giri 2 and Padam Simkhada 3,” pp. 1–16, 2020.
- [17] J.-L. N. Upoalkpajor and C. B. Upoalkpajor, “The Impact of COVID-19 on Education in Ghana,” *Asian J. Educ. Soc. Stud.*, vol. 9, no. 1, pp. 23–33, 2020, doi: 10.9734/ajess/2020/v9i130238.
- [18] G. Di Pietro, F. Biagi, P. Costa, Z. Karpiński, and J. Mazza, *The Likely Impact of COVID-19 on Education: Reflections based on the Existing Literature and Recent International Datasets*, vol. EUR 30275, no. JRC121071. 2020.
- [19] D. Zhang, J. L. Zhao, L. Zhou, and J. F. Nunamaker, “Can e-learning replace classroom learning?,” *Commun. ACM*, vol. 47, no. 5, pp. 75–79, 2004, doi: 10.1145/986213.986216.
- [20] N. J. Ogunode, “Impact of COVID-19 on the Higher Institutions Development in Nigeria,” *Electron. Res. J. Soc. Sci. Humanit.*, vol. 2, no. I, pp. 168–172, 2020.

## predicting stress level in online education due to pandemic: A case study of Bangladeshi students

### ORIGINALITY REPORT

<b>18%</b>	<b>14%</b>	<b>8%</b>	<b>13%</b>
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

### PRIMARY SOURCES

<b>1</b>	<b>dspace.daffodilvarsity.edu.bd:8080</b> Internet Source	<b>5%</b>
<b>2</b>	<b>Submitted to Daffodil International University</b> Student Paper	<b>4%</b>
<b>3</b>	<b>Zülfikar Aslan, Mehmet Akin. "Automatic Detection of Schizophrenia by Applying Deep Learning over Spectrogram Images of EEG Signals", Traitement du Signal, 2020</b> Publication	<b>1%</b>
<b>4</b>	<b>"Proceedings of International Joint Conference on Computational Intelligence", Springer Science and Business Media LLC, 2020</b> Publication	<b>1%</b>
<b>5</b>	<b>Inggar Tri Agustin Mawarni, Novia Ratnasari, Anik Nur Handayani, Muladi Muladi, Eka Purmanta Aji Wibowo, Rahmania Sri Untari. "Effectiveness of Whatsapp in Improving Student Learning Interests During The Covid-19 Pandemic", 2020 4th International</b>	<b>1%</b>