MACHINE LEARNING MODELING FOR STUDENTS' STRESS DETECTION

BY

MAHIM HASAN

191-15-12442

SHERIN SULTANA

191-15-12934

AND

MD MEHEDI HASSAN MIM

191-15-12829

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

Supervised By

DR. MD. TAREK HABIB
Associate Professor
Department of CSE
Daffodil International University

Co-Supervised By

NARAYAN RANJAN CHAKRABORTY
Associate Professor
Department of CSE
Daffodil International University

DAFFODIL INTERNATIONAL UNIVERSITY
DHAKA, BANGLADESH
JANUARY 2023

APPROVAL

This Project/internship titled **"Machine Learning Modeling for Students' Stress Detection"**, submitted by Mahim Hasan, ID: 191-15-12442, Sherin Sultana, ID: 191-15-12934 and Md. Mehedi Hassan Mim, ID: 191-15-12829 to the Department of Computer Science and Engineering, Daffodil International University has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of B.Sc. in Computer Science and Engineering and approved as to its style and contents. The presentation has been held on *24 January 2023*.

BOARD OF EXAMINERS

______________________ **Chairman**

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

______________________ **Internal Examiner**

Abdus Sattar
Assistant Professor
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

24.1.23
______________________ **Internal Examiner**

Fatema Tuj Johra
Senior Lecturer
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

24-01-23
______________________ **External Examiner**

Dr. Dewan Md Farid
Professor
Department of Computer Science and Engineering
United International University

DECLARATION

We hereby declare that, this project has been done by us under the supervision of **Dr. Md. Tarek Habib, Associate Professor, Department of Computer Science and Engineering (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:

Dr. Md. Tarek Habib
Associate Professor
Department of CSE
Daffodil International University

Co-Supervised by:

Narayan Ranjan Chakraborty
Associate Professor
Department of CSE
Daffodil International University

Submitted by:

Mahim

Mahim Hasan
ID: 191-15-12442
Department of CSE
Daffodil International University

শিরিন সুলতানা

Sherin Sultana
ID: 191-15-12934
Department of CSE
Daffodil International University

Mehedi

Mehedi Hassan Mim
ID: 191-15-12829
Department of CSE
Daffodil International University

ACKNOWLEDGEMENT

First, we have to admit that this project could not even be scarcely done without the help of adequate people in respective sectors of our project, proper guidance from our supervisor and of course the grace of the **Almighty Allah**.

We are really grateful and wish our profound indebtedness to **Dr. Md. Tarek Habib**, **Associate 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 thesis. His endless patience, scholarly guidance, continual encouragement, constant and energetic supervision, constructive criticism, valuable advice, reading many inferior drafts and correcting them at all stages 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, Daffodil International University, for his kind help to finish our project and also to other faculty members 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 discussion while completing the course work.

Finally, our deepest gratitude goes towards our parents for their continuous support and utter belief in us which ultimately led us towards the completion of this project.

ABSTRACT

Mental stress is an increased mental state of the human body that creates unavoidable situations by which people become mentally and physically weak. Factors which trigger stress are called stressors. If a person continues to take stress all at once, both his mind and body will be affected by the flood. It is very harmful to a human being so it should be prevented as soon as possible. Mental stress as usual happens with every person but if we observe deeply, it is very common among university students. University is the toughest part of life which give us future life problem and solutions. A university student is a grown man who has to think about future earnings. They feel very insecure about their future life which is cause a big amount of mental stress. So, we decided to check the amount of stress over various types of students. We categorized them by gender and age. Then we created a questions sheet where we put twenty-four common questions. We sorted out the questions by talking with a senior psychologist and reading big amount of resources. By answering them we can measure the amount of mental stress among the categories and also, we can find out the reason behind them which will help us to reduce the number of students suffering from mental stress in educational life.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

1.1 Introduction

There are 264 million people who experience mental stress, and one in four persons experience neurological diseases [1]. Mentally stressed individuals frequently exhibit poor performance at job, in the classroom or in the home, which in the worst scenario might result in suicidal thoughts. Suicide is the second most common cause of mortality for people between the ages of 15 and 29 after mental stress [1]. 60.8% of people worldwide report moderate to severe mental stress, 73% report anxiety and 62% report stress [2]. The biggest issue nowadays is drug addiction, which either directly or indirectly resulted from mental stress. About 25 lac persons in Bangladesh are drug addicts, of whom 80% are children and young adults [3].

Mental health issues are an increasing hazard in low and middle-income nations like Bangladesh, despite the fact that they are rarely mentioned. In Bangladesh, there are around 7 million persons who have anxiety and depression illnesses. In 2012, an estimated 10,167 people died by suicide, with 4% of boys and 6% of girls between the ages of 13 and 17 thinking about trying to end their lives [3]. In Bangladesh, levels of mental stress, anxiety and stress have been observed to range from 54.3% to 64.8% to 59.0%, respectively [4].

Worldwide pandemic COVID-19 is also the biggest issue for mental stress by many ways. From them, one of the issues is lack of work, which create some situation where people are interested in non-profitable work like using social media. According to a statistic approximate 4.55 billion people are active in social media since October 2021, which means 57.6% people of worldwide are using social media during pandemic [5]. Especially in Bangladesh, there is a huge user for Facebook. An estimation says 96.3% Bangladeshi use Facebook [6]. Social media addiction is also the biggest cause for mental stress.

Numerous causes, both academic and extracurricular, including socioeconomic, ecological, cultural and psychological ones, can lead to stress in students. Students' feelings of stress vary based on their anxiety symptoms, particularly around exam time.

"Functionally Impairing Test Anxiety" affects college students at a rate of 10% to 35% of the total population. Test anxiety increases a student's likelihood of delaying and dropping out of college, which might result in suicidal behavior and significant financial expenditures. It can go both ways because most students perform poorly in school and have low self-confidence while under pressure.

The literature has identified a number of factors that affect students' levels of stress, anxiety and mental stress, including gender, strained relationships, peer and family pressure, high parental expectations, a lack of financial support and difficulties, sleep deprivation, worry about the future, loneliness, increased screen time, toxic psychological environments, academic pressure environments, workload, academic curriculum volume and demanding exam schedules. Exams, time commitment, competition and the classroom atmosphere are some of the top academic stressors for students, but the top personal stressors are close relationships, money and parental disputes.

Exam stress and student mental health are related, and students who experience exam stress and anxiety suffer. Academic achievement, physical development and living standards. Test anxiety may result from social stigma, which may make rural students feel less capable than students who grew up in urban areas, combined with a worry that they do not have competitive English skills. The average level of higher education in Bangladesh can also be demoralizing before and during exams, which can result in anxiety, stress and frustration.

Higher education today comes with a heavy financial load. Cross-sectional research suggests that financial difficulties frequently result in mental health problems for students, including depression. It also suggests that college students are more vulnerable to depression as a result of financial difficulties. While pre-access tension and depression were managed, out of all the negative reports evaluated, financial difficulty was the only one that was easily able to reveal an unbiased relationship with depression. This indicates that the route of causality is much more likely to be from financial issues to depression than vice versa. Furthermore, excessive parental control is supposed to prevent children from developing their autonomy, leading to perceptions of their environment as unpredictable and a limited sense of their own competence or mastery. In return, those elements are important to contribute to the triggering stress in children.

In 2020, a survey was conducted in Dhaka, Bangladesh, with the intention of identifying factors—particularly sociodemographic and mental factors—that have an impact on college students' ability to think critically.

1.2 Motivation

Mental stress is a sickness with a mood-structure. When someone is depressed, he locks himself in a room by himself. His daily routines include avoiding social meetings, neglecting his family and losing focus on his artwork. These days, despair has become a dangerous reality.

A man or woman may experience depression in any way that is dependent on his life. Disease brought on by mental stress is not unusual in humans. But the biggest issue is that people in Bangladesh are not aware of issues with intellectual capacity. People are unable to recognize their depression. They do not see a psychiatrist or psychologist.

Therefore, it seeks to determine whether or not someone is depressed or no longer based solely on their daily activities or way of life. Because of this, we plan to develop a technique or tool to assess someone's level of mental fitness, in which case the system for learning can accomplish a great lot.

As we evaluate the literature, we are no longer able to find many additional studies on this topic. In this area, numerous investigations are desperately needed. Therefore, we are using a well-known and widely-used approach to learn how to accomplish this despair detecting painting.

1.3 Fundamental Principle of the Study

As we previously mentioned, there are significantly less full-size paintings from the Bangladeshi perspective that were done in the past with desperation detected. This is why we strongly advise using gadget learning techniques and running with despair. As a subset of artificial intelligence, machine learning provides probabilistic, statistical and optimization techniques that let computers "learn" from patterns and similarities found in large, noisy, challenging and complicated datasets. Techniques for examining modern devices are employed in the disciplines of detection, recognition and classification. Systems for analysis, disease recognition, stock market analysis and traffic prediction all use machine learning. The usage of device learning is used to solve a wide range of detection and classification problems. Machine learning plays a significant role in this field and is widely used by academics, producing some spectacular results. Thus, we made the decision to apply device learning for the sake of our research.

1.4 Research Questions

- How do we recognize mentally stressed individuals?
- How do we keep the feature for our data up to date?
- How much data do we gather?
- Where and how the data are gathered?
- How many pieces make up our model and test dataset?
- Are our dataset and machine learning methods reliable and appropriate?
- Should we have to create new models or employ well-known and widely-used machine learning techniques?

1.5 Expected Output

The aim of our paper is to make human behavior distinguishable from desperation. One of our major issues is that we are unable to recognize dejection when it strikes. By employing our method, someone can quickly recall his intellectual realm. The population as a whole might no longer be confined to their intellectual world as a result. We ought to address a range of social groups. We keep noticing that a number of our pals have tied the knot. They are not taking part in any of society's artistic endeavors. However, there is no longer a discernible decrease in the diversity of these individuals within society. This essay will be useful in identifying concepts that express the intellectual side of humanity.

1.6 Report Layout

The following are the contents of this research project:

- The research is summarized in Chapter 1 along with its goal, guiding principles, open-ended research questions and expected outcomes.
- Chapter 2 includes information on the issue's scope, difficulties we face, a summary of the research and relevant earlier publications.

- The flowchart in Chapter 3 includes descriptions of the data collection process, data preprocessing, statistical analysis and feature implementation.
- The experimental analysis in Chapter 4 contains a summary of accuracy, findings from the research and other related investigations.
- Chapter 5 discusses the implications of this research for society.
- The results, limitations and future work of this research are all explored in Chapter 6.

CHAPTER 2

BACKGROUND STUDY

2.1 Introduction

We will go over similar research that has already been conducted by other researchers, a summary of their findings, the scope of the problem and the challenges we ran into in this section. We include a few research, related works, their applied methods, classifiers and accuracy levels that are pertinent to our work in the section on similar work. In the research summary section, we compile a summary of all studies, which we then display as a table to facilitate understanding. We discuss how we may help or advance this effort when it comes to the scope of the problem. The challenges section concludes by describing the kinds of problems and challenges we ran into throughout our research effort and how we overcome them.

2.2 Similar Works

Studies on the importance of mental strain and its calculations have been conducted. The authors of [7] have suggested using the suggested system learning framework to calculate mental strain from confused members' electroencephalograms (EEGs). Five capacities are retrieved from the EEG spectrum, and 94.6% category accuracy is attained after function selection and the use of the three classifiers support vector machine (SVM), naïve Bayes (NB) and logistic regression (LR). The authors of [8] have created a portable EEG sensor that can be used in daily life. It is suggested that a set of guidelines for calibrating the sensor be used, which can be customized by customers and have the electrode settings altered. It has been demonstrated that continuous strain can be identified using nonlinear EEG capabilities. 90% category accuracy is attained using the discrete wavelet transform (DWT) and adaptive noise cancellation (ANC)-based entire set of rules. Evaluation of coronary heart rate variability (HRV) is used in the procedures in [9] and [10]to determine intellectual pressure. According to this study [11], Parkinson condition (PD) can be detected by analyzing radio indicators that show a version when a healthy person is walking as opposed to someone who has the disease. The segment and amplitude capabilities of the radio indications, which can be categorized using the assist vector system are captured using a leaky wave cable. 90% accuracy is shown by the category effects during closed-loop monitoring. Similar to [12], different sensors are employed to store facial and posture data inside the proposed method to calculate excellent strain.

The authors of [13] used omnidirectional antennas and a few wireless devices to diagnose roaming behavior in dementia patients. Phase and amplitude indicator versions that fall within the S band are those that are recorded. SVM, which achieves 90% accuracy for the three patterns, is used to classify the proposed device. It is advised that early discovery of dementia will aid in prompt treatment of the patients and may limit the horrible effects of the illness. It has been suggested in [14] that electrodermal activity (EDA) can help distinguish pressure from cognitive load in enclosed office settings. The height top and height price are used to compute a person's stress levels based on EDA indications. Pass validation and an assist vector machine are used to achieve a performance accuracy of 80% (SVM).

The authors of [15] have presented a non-intrusive respiratory monitoring system that is entirely based on C-band sensing techniques. A microwave sensor platform is used to measure the chest's expansion and contraction. With the aid of a height detection algorithm, variations in respiration rates are used to identify common breathing types. Similar to [16], a framework with a community interface card, omnidirectional antenna and a router that statistics the variation of amplitude and section statistics with Wi-Fi channel interface is used to identify aberrant gait and hand tremors. According to [17], a wearable sensor that measures electrocardiogram (ECG), EEG and strain levels. The suggested device's objective changed to correlating variations in salivary cortisol with variations in strain intensities. This study [18], diagnoses the distinction between stability and tremor circumstances. Time area alerts and their spatial proximity to the Wi-Fi spectrum band, which provided 90% accuracy, are used to extract a variety of tasks. The authors of [19] suggested that near infrared spectroscopy (NIRS) and physiological data may be combined to improve the quality and accuracy of measuring intellectual strain.

Heart Rate Monitors (HRMs) were employed in a study [20] to detect variations in heart rate that are proportional to mental stress. The wavelength of the coronary heart and its pace both decrease as pressure rises. When utilized to capture coronary heart rate signals, HRMs have shown to offer evaluation results that are just as good as an ECG. Pulse Density Modulation (PDM) technique is employed for classification with pronounced 83% accuracy to extract functions and decide accuracy impacts. Coronary heart rate variability and its spectral components were used by the authors in [21]. Spectral frequency bands' left-to-right facets are employed as ratios. With increased pressure, the ratio will rise. Although the experiments produced satisfactory results, categorization accuracy could yet be improved. In [22], authors employed a combination of the Heart Rate (HR) and HRV functions to determine pressure. 28 people made up a tiny sample used in the research. Pix has been employed for recognition in a short amount of time to create pressure. It became clear that short-term HRV is correlated with intellectual pressure, although further research is required to confirm the results.

Laboratory tests had been utilized in [23] to cause mental strain. Recall tests for entering six digits at the screen are part of the cognitive experiments. The frequency ratio between the left and right bands is predicted to rise in proportion to how severe the strain ranges are. Additionally, blood pressure was tracked, and it was noted that under pressure, blood pressure stayed high. In [24], to conduct HRV analysis, a fuzzy clustering method was applied. To extract capabilities from fluctuations in heart rate, wavelet remodel was utilized. Simulations based on air traffic control were employed in the trials. Because of the enormous pressure and the uncontrollable environment, the examination has few practical applications. According to [25], strain ranges are correlated with lower HR rates, lower arm oxygen saturation levels and higher frame temperatures. The examination pattern includes 25 female subjects and is deemed modest because to the outcomes' hipster reputation. The authors in [26] applied deterministic fractals to the ECGs of all 26 participants as well as some basic chaotic structure features. Fractal algorithms were applied to the waveforms of HRV signals in a test device that was based on the ECG. The FD values that were recorded in the various stages of the examination, such as resting, early strain and late strain, confirmed an increase in the value ranges from the resting to the strain phase. It has been proposed that strain detection using fractal analysis may be useful for determining the severity of the strain. The authors of [27] employed morphologic variability (MV) in conjunction with other quantifiable characteristics, including multiple coronary HRV measures from ECG signs in both the temporal and frequency domains. Using HRV analysis over a very short time period, [28] calculated mental strain. Different HRV measurements, such as the sympathovagal stability index (SVI), the mean RR intervals (mRR), the mean heart rate (mHR), the low frequency (LF), the very low frequency (VLF) and the high frequency (HF) strength spectrum, were utilized to determine the stress levels.

To determine degrees of mental strain, the authors in [29] used a microwave reflectometric cardiac sensing device. They developed methods to record HRV by using dynamic movement markers at the surface of the frame. At first, a cross-correlation characteristic was employed to look for similarities between the recorded signals and a template signal that had been produced by using a waveform's periodic additions and regional averaging. Secondly, an entropy technique based entirely on the most probable feature was used to reconstruct the time-variant of the pulse frequencies inside the recorded HRV. To calculate continuous stress levels, a time-based variation evaluation of HRV capabilities was conducted in [30]. Three distinct time periods have been chosen within a day to record the HRV capabilities. LR technique was utilized for categorization, and it produced accuracy levels of 63.2%. The authors of [31] employed fuzzy clustering in conjunction with machine learning classifiers to assess mental fatigue. The collected indicators were smoothed online using continuous wavelet rework capabilities for the evaluation of heart rate variants. Fuzzy clustering techniques have been used to model the experimental records. Regression techniques and fuzzy reasoning have been used to remove a variety of anomalies and uncertainties from the accumulated records.

In [32], the authors study about many papers related to mental stress. From the papers they found various reason and result to cause for mental stress. They discuss about using NB, SVM, random forest (RF), decision tree (DT), and *k*-nearest neighbor (*k*-NN). Their basic work topics are psychological, biological, behavioral, prevalence of stress worldwide, risk factors of parenting stress, and soft-computing techniques.

In [33], the authors say many studies are using ultra-sort-term HRV but no ever make it automatic. So, they do the job. By using their model, we can able to detect mental stress automatically and it has 90% accuracy.

In [7], the authors used to perceive stress scale (PSS) to measure mental stress. Their subject was total forty-two from them eleven was female. They mention that their all subject was healthy and no one had previous medical report. After test they divided them into four level. They said the relative power produce the best performance with the t-test and NB in level 1. The accuracy was 94%.

In [34], it was a medical case study over mental stress. The authors used ECGs for detecting mental stress. Their model was created by using ultra-short HRV. Their accuracy was above 88%.

2.3 Comparative Analysis and Summary

With the aid of machine learning algorithms and data mining techniques, some work has already been done on the detection and attribution of mental stress. With the detection of mental stress, the identification of alcohol users and the detection of different ailments, the use of gadgets for learning has expanded in the modern period. This stage denotes an evaluation of those related works. An evaluation of several research' topics, methodologies and results is provided in Table 2.1 below.

Table 2.1: An Overview of Related Research Works

Serial No.	Authors name	Methodologies	Descriptions	Results
1	A. R. Subhani et al.	EEG	Using EEG spectrum, five features are extracted	94.6% accuracy achieved by NB and LR
2	B. Hu et al.	By a lightweight EEG sensor	Users can alter the setting of the electrodes and a user-tuned algorithm for calibrating the sensor is proposed	90% classification accuracy is attained by the DWT and ANC based technique
3	X. Yang et al.	The phase and amplitude characteristics of the radio signals that are identified by the SVM are captured using a leaky wave wire SVM	Radio waves that differ between normal and PD patients are processed to help diagnose PD. Unlikely person is walking	The categorization findings show a 90% accuracy under closely supervised conditions
4	S. Koldijk et al.	By various sensors	Incorporate facial and posture information	In closed controlled conditions, the categorization results show a 90% accuracy rate
5	J. Choi et al.	HRV	Analyzing HRV to assess mental stress	Result demonstrates 87% accuracy in various algorithms
6	X. Yang et al.	Omnidirectional antennas and a few wireless devices were used to identify the roaming behavior of dementia sufferers	SVM is used to conduct the suggested system's classification	90% accuracy is achieved for the three input patterns
7	C. Setz et al.	EDA	By assessing the peak height and peak rate of EDA signals, stress levels of an individual are determined	80% accuracy is attained
8	X. Yang et al.	A non-intrusive breathing monitoring device based on a C-band wavelength sensor method	Using a microwave sensor platform, the chest's expansion and contraction are recorded	An accuracy of 85% is achieved
9	E. Labbe et al.	HRMs were employed to identify variations in heartbeats correlated with mental stress	Heart rate signals are captured using HRMs, and studies have shown that their performance is comparable to that of an ECG	83% accuracy is reported
10	R. Costin et al.	Microwave reflectometric cardiopulmonary sensor device for assessing degrees of stress	Two methods for recording HRV and comparing it to a template signal	84% accuracy is reported

Deep learning and artificial intelligence (AI) are now widely used in all areas of information science for models of expectation, categorization and discovery. The discovery models use calculated relapse, convolutional neural networks (CNN), artificial neural networks (ANN), SVM, *k*-NN and many more well-known calculations. Based on the results of our writing survey, it is clear that the *k*-NN, NB, SVM and DT algorithms have high values and are widely used in forecasting, discovering and acknowledging models. To identify the demotivated people in Bangladesh, we attempted to carry out calculations using *k*-NN, DT, NB, gradient boosting (GB), SVM and RF classifier in our research effort. We received 93.50% accuracy in our calculated relapse.

2.4 Scope of the Problem

Our investigative job essentially consists of putting on a show by analyzing the supplied data and using machine learning computations. Our suggested demonstration is capable to identifying sorrow. This work will have a significant impact on society's citizens. There are so many people in our culture who are struggling with mental clutter. They are disheartened, but they are unaware of it. They isolate themselves and lack the necessary knowledge. They decide to act in an off-base manner because of this. In this situation, they need a framework that will help them identify their problem and determine if they are truly demoralized or not. Finally, let them know what they need to do by providing them with information. As of late, as machine learning and fake insights are being utilized for different question discovery and illness forecasts, the come about is very satisfactory. Hence, we choose to utilize machine learning in order to make a show of discouragement location.

2.5 Challenges

We encountered a few problems when doing our investigation. The most challenging element was gathering the data. It was difficult to distinguish between a healthy individual and a sad individual. The people who are experiencing mental health problems are reluctant to talk and do not want to provide their consent. We spoke with individuals from many social classes, read a lot of publications, and spoke with a lot of doctors, but we were unable to come up with a good response or solution. Nobody can give us unambiguous information regarding mental stress because depressive symptoms vary from person to person and from situation to scenario. As a result, it was difficult to change the research questionnaires. People were hesitant to divulge their personal information, though. We occasionally chatted with people personally while gathering data. However, this was also suspended as a result of the COVID-19 epidemic.

Engineering and some of the machine learning algorithms were unknown to us. It took some time at first to understand it. After that, with the help of our supervisor and practice, we were able to get through that difficulty and finished our assignment.

CHAPTER 3

RESEARCH METHODOLOGY

3.1 Introduction

The development of a model for identifying mental stress is the aim of this study. This model is created and studied in light of typical activities people engage in and their reactions. To build this model, several machine learning methods are employed. A number of techniques were used including LR, RF, SVM, k-NN, GB, Extreme Gradient Boosting (XGBoost), DT and NB. People are classified as Stressed, Mild stressed or Unstressed using algorithms. There are identified to be twenty-four traits that are either directly or indirectly associated to mental stress. Prior to the final implementation, we processed our dataset as needed. Additionally, we computed the sensitivity, specificity, precision, recall and F_1-score and exhibited receiver operating characteristic (ROC) curves for LR, k-NN and NB in order to determine the optimal accuracy.

3.2 Data Collection

The set of facts consists of a wide range of aspects or components that are either directly or indirectly related to mental stress. We were unable to collect the required information as we traveled to the clinic since the government had warned them that disclosing patient information could violate patients' privacy, have an impact on their policies, and occasionally not be available as a prepared dataset. Because of this, we made the decision to develop our own dataset using information gathered via in-person surveys, online government documents and a paper with a list of questions. Hopefully, we were successful in gathering 830 human facts based entirely on 24 factors that are typical human activities. After obtaining all the information, the main challenge became apparent: the information can be leveled into Stressed, Mild Stressed and Unstressed. We consult a variety of documents as we look for assistance and identify fashions to figure things out. The facts had finally been leveled independently with the help of one doctor, one physiatrist and one psychology student and integrating their three choices into the final leveling outputs, placing a focus on the consensus. There are 248 stressed, 335 mild stressed and 247 unstressed individuals among the one fact. The information we have gathered through an online poll differs between secondary and higher secondary schools, colleges and universities.

We collected the data based on the below topics:

1. Gender
2. Age
3. Last two-week mood
4. Self-motivation
5. Loneliness
6. Guiltiness
7. Insomnia
8. Suicidal interest
9. Relation with family members
10. Financial status
11. Relation with classmates and teachers
12. Ragging history
13. Career status
14. Decision making
15. Inattention or forgetfulness
16. Insecure by family
17. Educational life happiness
18. Hopeless or disappointed
19. Corona affection
20. Family Corona affection
21. Dead by Corona in family
22. Jobless for Corona
23. Financial damage by Corona
24. Interruption of study by Corona

We made the questions by talking with physiatrists, doctors, reading papers and articles. In present situation, Corona can be the main cause for mental stress, so we take it as serious topic. We divided mental stress into three categories to see the stress difference. Process details are included below.

3.3 Research Subject and Instrumentation

System mastery algorithms, data mining and deep learning techniques are extremely appropriate and popular right now for any kind of prediction, recognition and detection. We made an effort to test a variety of system-mastering algorithms on our acquired dataset in order to determine which algorithms would best serve our needs and perform well. LR, RF, SVM, *k*-NN, GB, XGBoost, DT and NB algorithms were just a few of the system mastering algorithms we performed. Recently, "Python" has become one of the most popular and widely utilized programming languages that is frequently used for research functions by researchers. Therefore, we employed python as our programming language, google colab, google forms, jupyter notebook, Microsoft Excel as our dataset and Django framework for the implementation on this research project.

3.3.1 Proposed Methodology

In Figure 3.1 our steps of the proposed methodology is shown:

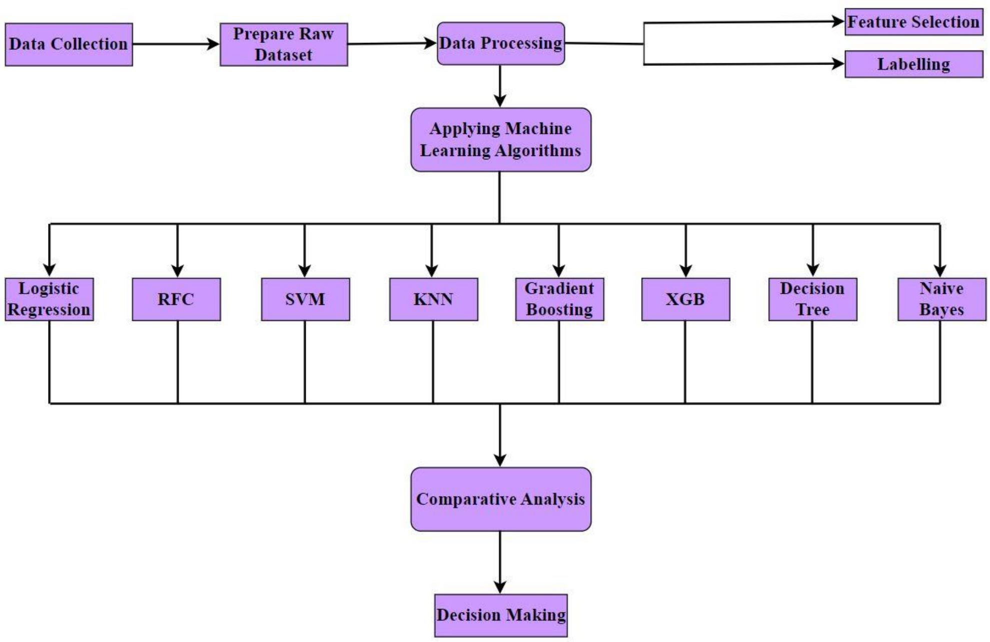


Figure 3.1: Steps of Proposed Methodology

3.3.2 Data Preprocessing Process

When we successfully aggregate a surplus of information, we frequently discover that some of the data has some missing values and that there are several types of information, such as categorical data and numerical information. Machine learning techniques are not suited for this kind of knowledge. As a result, we often decide to organize our knowledge in a way that suits our preferences and makes it compatible with algorithms. After knowledge collection, processing has the ability to transform knowledge into acceptable formats. Information that has been more specifically processed to assist produce the best results quickly.

Data preprocessing process is shown below here in Figure 3.2.

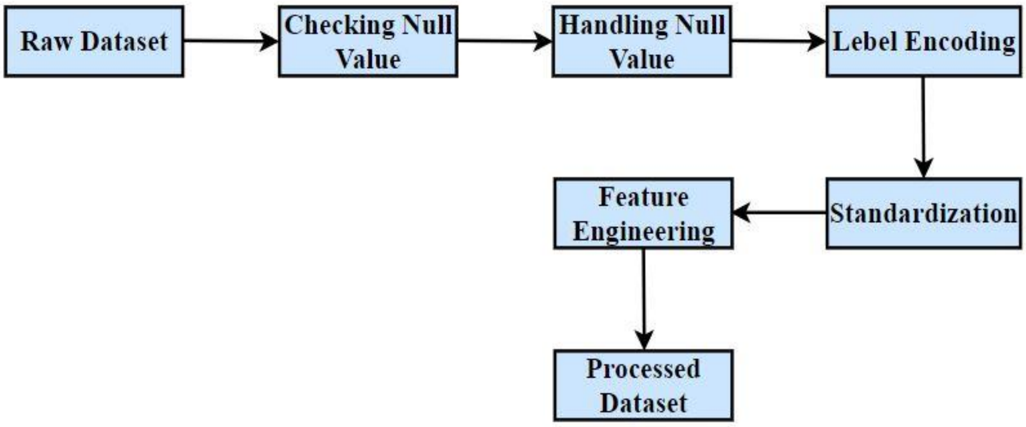


Figure 3.2: Data Preprocessing Process

First of all, we collected and created our raw dataset as a way to start our work. Then we focused on cleaning up the statistics. We looked to see if the statistics set contained any missing or null values. Then, we encoded the portion of the textual material or specialized statistical data that converts it all to the relevant numerical statistics. The statistical transformation was then completed using standardization. After that, for the goal of function engineering, we identified and examined the correlation between multiple features. Then we identified the 16 best features among 22 features keeping the other two features i.e., gender and age as implied. As a result, we ultimately obtained the final processed dataset of 18 features including both the gender and the age features. The google colab was used for the entirety of the work involved in processing the data.

3.4 Statistical Exploration

From our survey we collected 830 students' data from different levels and categories. Students from schools, colleges and universities were included in the data we gathered. In Figure 3.3 we can see how much people are stressed, mild stressed and unstressed. We worked and built our project and fulfilled all further processes. Based on our work we found 248 stressed, 335 mild stressed and 247 unstressed people. In percentage which is about 29.9% stressed, 40.4% mild stressed and 29.8% unstressed.

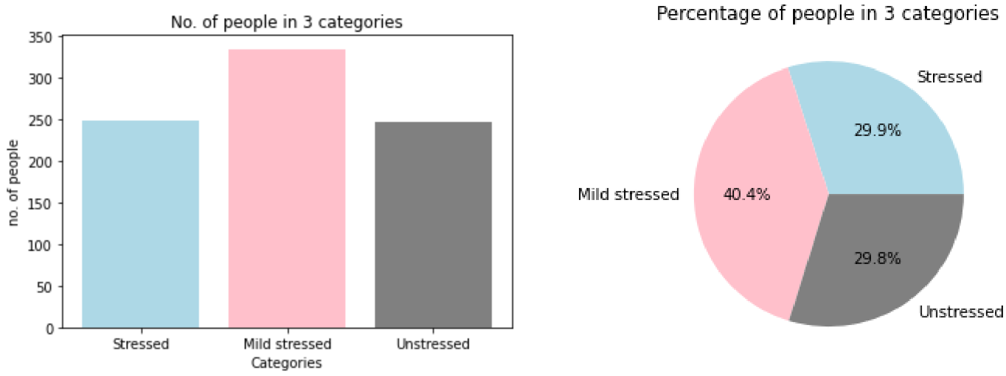


Figure 3.3: Stressed, Mild stressed and Unstressed Cases

Figure 3.4 shows the number of males in three categories. There were total 423 males. From them 117 were stressed, 186 were mild stressed and 120 were unstressed. In percentage 27.7% were stressed, 44% were mild stressed and 28.4% were unstressed. Figure 3.4 is shown below.

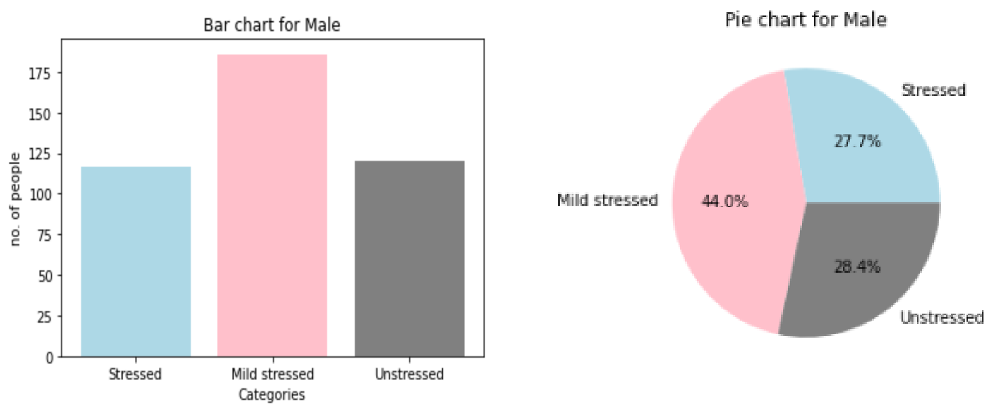


Figure 3.4: Male Mental Stress

Figure 3.5 shows the number of females in three categories. There were total 326 females. From them 103 were stressed, 115 were mild stressed and 108 were unstressed. In percentage 31.6% were stressed, 35.3% were mild stressed and 33.1% were unstressed. Figure 3.5 is shown below.

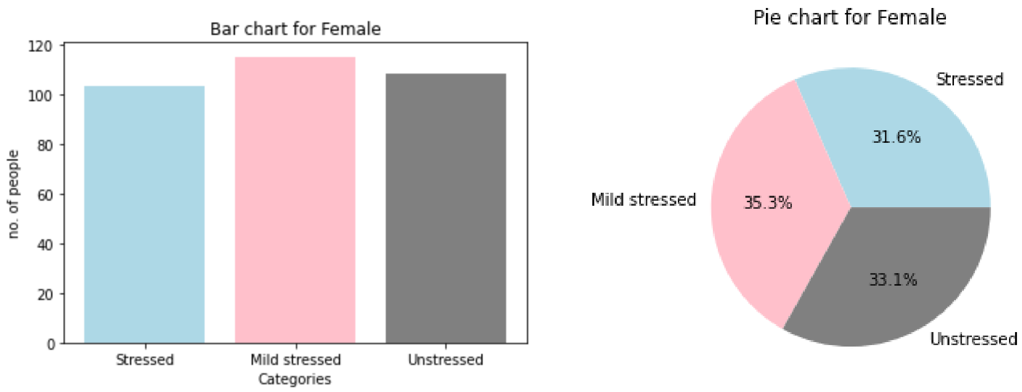


Figure 3.5: Female Mental Stress

Figure 3.6 shows the number of prefer not to say in three categories. There were total 81 prefer not to say. From them 28 were stressed, 34 were mild stressed and 19 were unstressed. In percentage 34.6% were stressed, 42% were mild stressed and 23.5% were unstressed. Figure 3.6 is shown below.

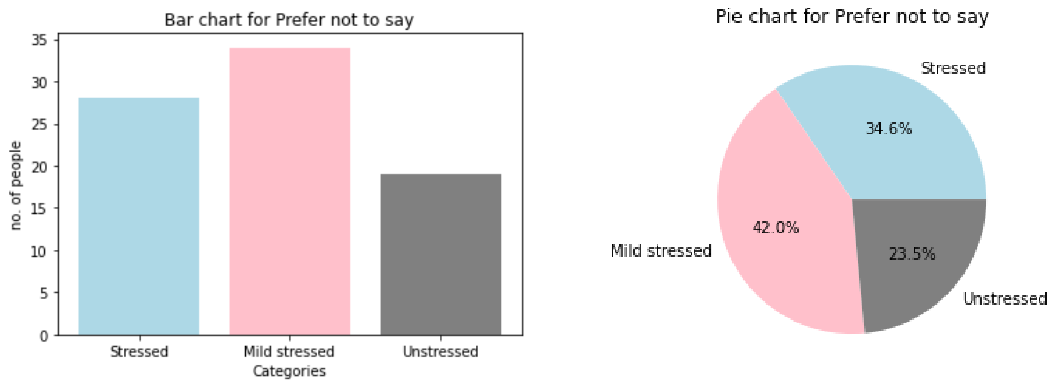


Figure 3.6: Prefer not to say Mental Stress

Figure 3.7 shows the number of less than 18 in three categories. There was total 149 less than 18. From them 49 were stressed, 69 were mild stressed and 31 were unstressed. In percentage 32.9% were stressed, 46.3% were mild stressed and 20.8% were unstressed. Figure 3.7 is shown below.

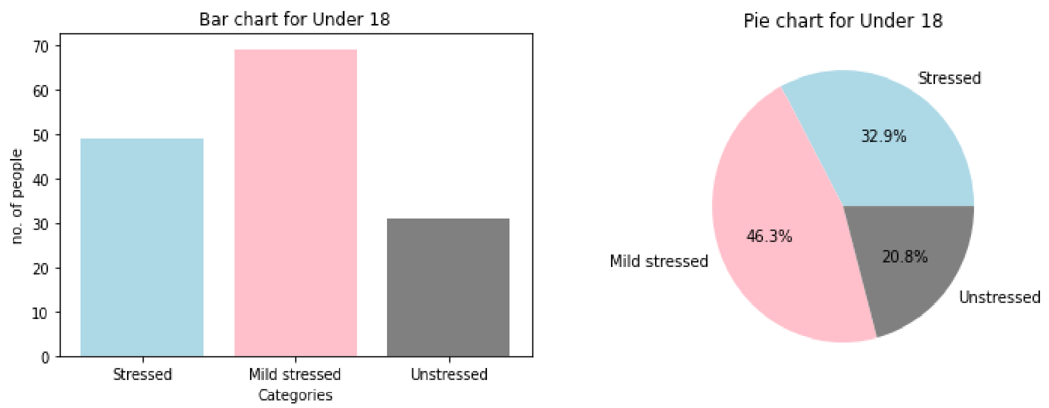


Figure 3.7: Less than 18 Mental Stress

Figure 3.8 shows the number of ages 18-30 in three categories. There was total 524 ages 18-30. From them 157 were stressed, 198 were mild stressed and 169 were unstressed. In percentage 30% were stressed, 37.8% were mild stressed and 32.3% were unstressed. Figure 3.8 is shown below.

Figure 3.8: Ages 18-30 Mental Stress

Figure 3.9 shows the number of more than 30 in three categories. There was total 157 more than 30. From them 42 were stressed, 68 were mild stressed and 47 were unstressed. In percentage 26.8% were stressed, 43.3% were mild stressed and 29.9% were unstressed. Figure 3.9 is shown below.

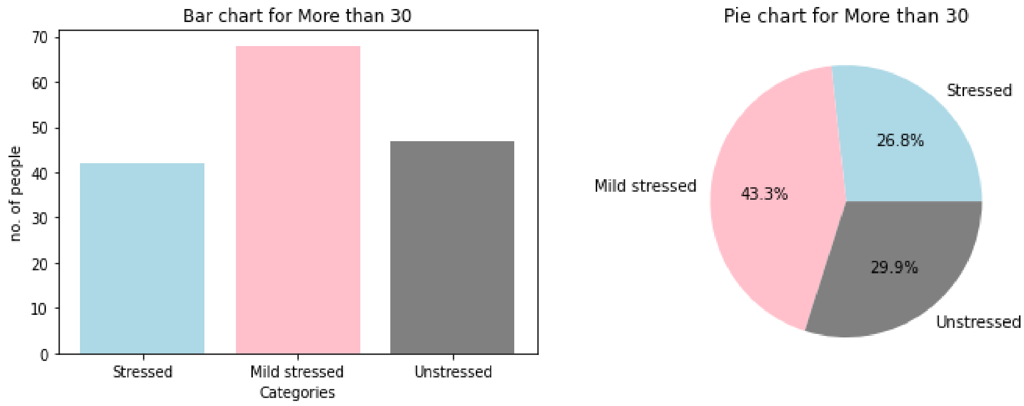


Figure 3.9: More than 30 Mental Stress

Figure 3.10 shows the number of stressed persons according to gender. There was total 248 stressed people. From them 117 were males, 103 were females and 28 were prefer not to say. In percentage 47.2% were males, 41.5% were females and 11.3% were prefer not to say. Figure 3.10 is shown below.

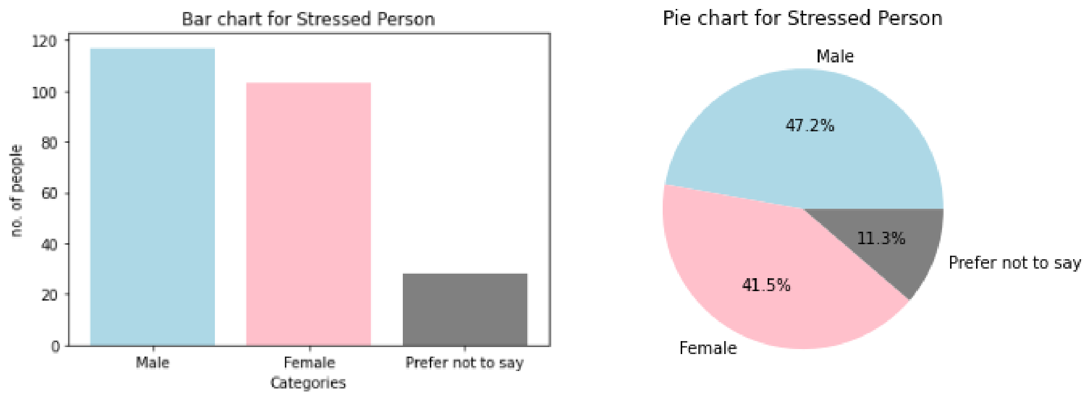


Figure 3.10: Stressed According to Gender

Figure 3.11 shows the number of mild stressed person according to gender. There was total 335 mild stressed people. From them 186 were males, 115 were females and 34 were prefer not to say. In percentage 55.5% were males, 34.3% were females and 10.1% were prefer not to say. Figure 3.11 is shown below.

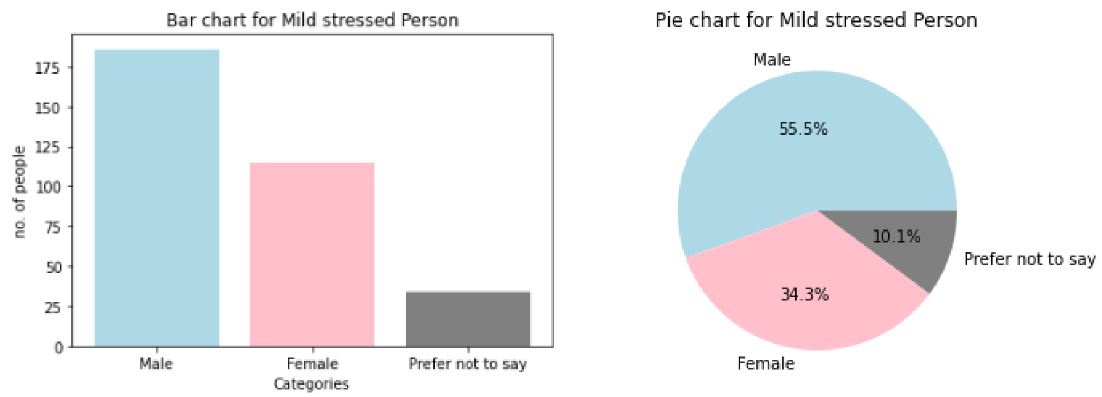


Figure 3.11: Mild stressed According to Gender

Figure 3.12 shows the number of unstressed persons according to gender. There was total 247 unstressed people. From them 120 were males, 108 were females and 19 were prefer not to say. In percentage 48.6% were males, 43.7% were females and 7.7% were prefer not to say. Figure 3.12 is shown below.

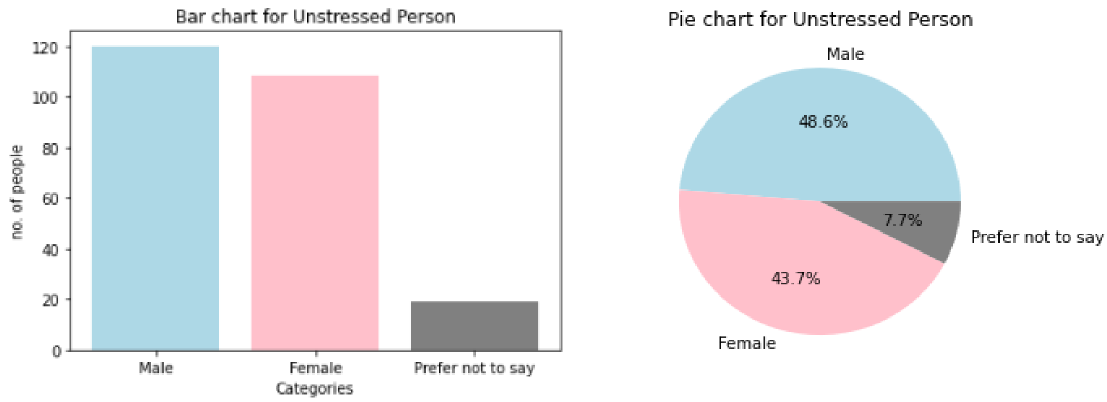


Figure 3.12: Unstressed According to Gender

Figure 3.13 shows the number of stressed persons according to age. There was total 248 stressed people. From them 49 were less than 18, 157 were 18-30 and 42 were more than 30. In percentage 19.8% were less than 18, 63.3% were 18-30 and 16.9% were more than 30. Figure 3.13 is shown below.

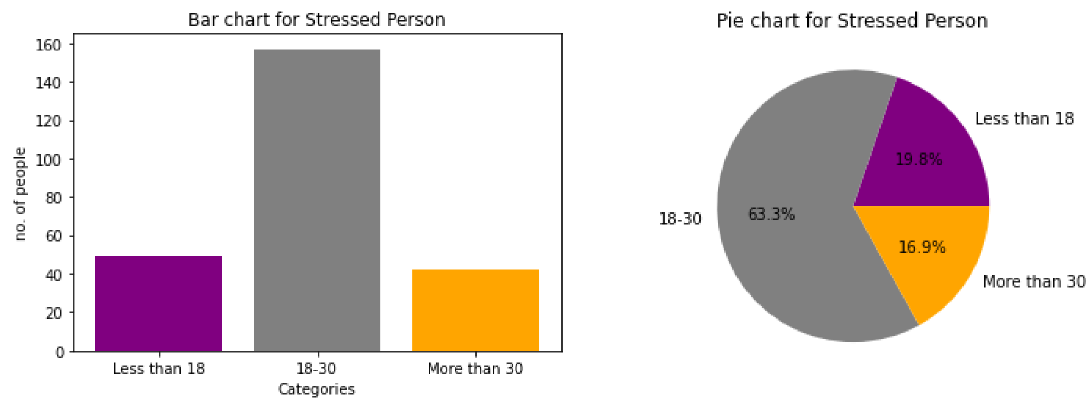


Figure 3.13: Stressed According to Age

Figure 3.14 shows the number of mild stressed person according to age. There was total 335 mild stressed people. From them 69 were less than 18, 198 were 18-30 and 68 were more than 30. In percentage 20.6% were less than 18, 59.1% were 18-30 and 20.3% were more than 30. Figure 3.14 is shown below.

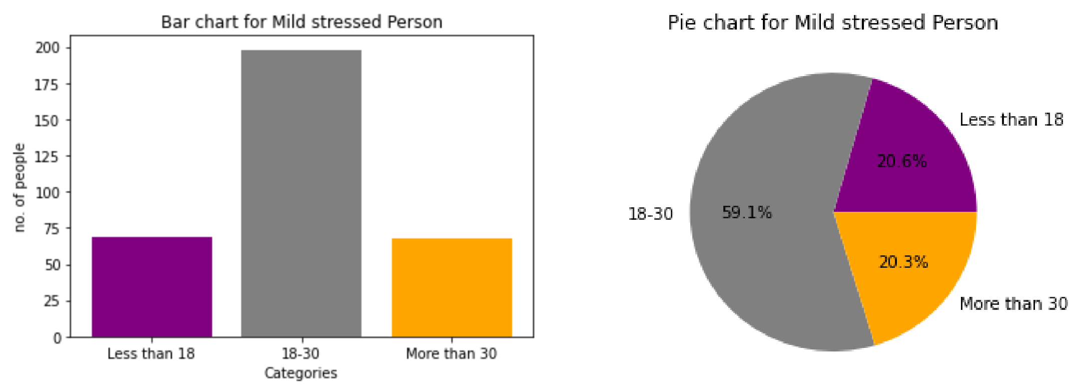


Figure 3.14: Mild stressed According to Age

Figure 3.15 shows the number of unstressed persons according to age. There was total 247 unstressed people. From them 31 were less than 18, 169 were 18-30 and 47 were more than 30. In percentage 12.6% were less than 18, 68.4% were 18-30 and 19% were more than 30. Figure 3.15 is shown below.

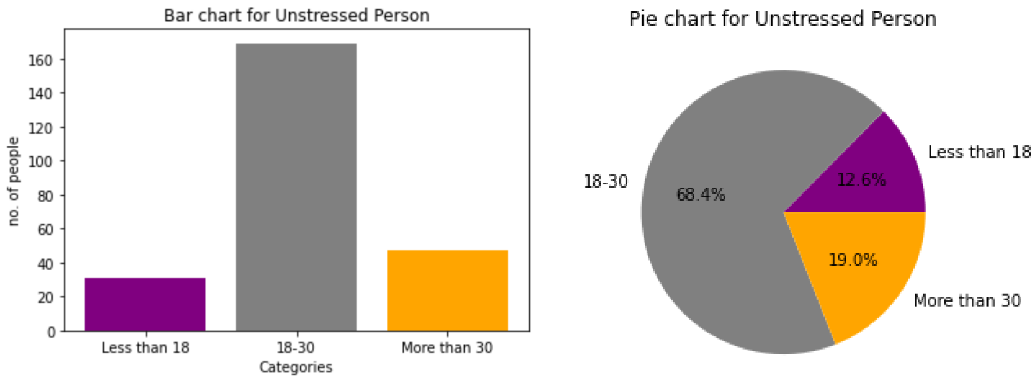


Figure 3.15: Unstressed According to Age

Figure 3.16 is for showing the correlation of multiple features. Which is used to find out the similarity rates among them. Figure 3.16 is shown below.

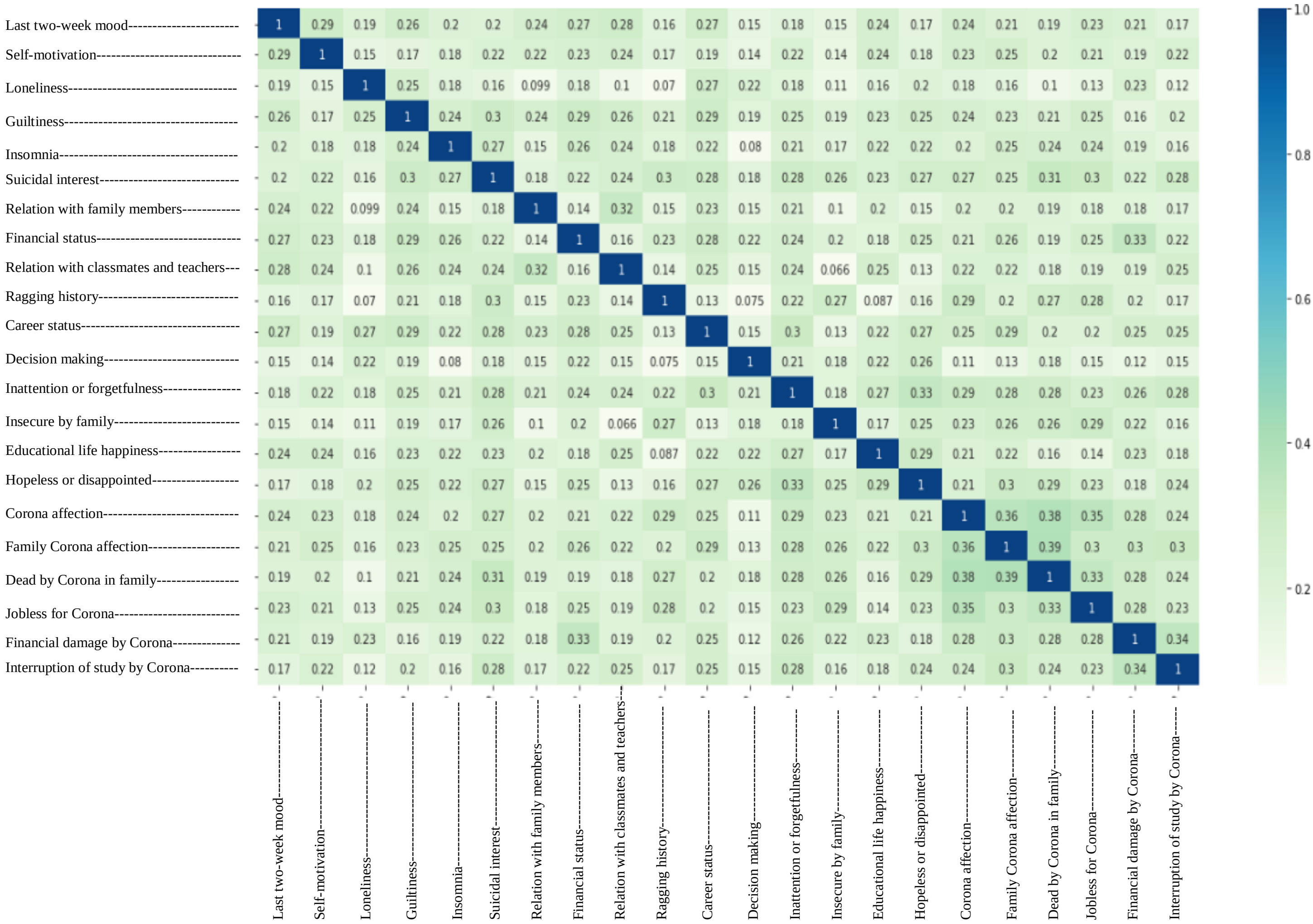


Figure 3.16: Correlation Matrix of Multiple Features

Table 3.1: Feature for Mental Stress Detection

Feature Name	Evidence Based-on	Feature Name	Evidence Based-on
Gender	[35]	Ragging history	[36]
Age	[36]	Career status	[37]
Last two-week mood	[38]	Decision making	[39]
Self-motivation	[38]	Inattention or forgetfulness	[37]
Loneliness	[40]	Insecure by family	[39]
Guiltiness	[38]	Educational life happiness	[37]
Insomnia	[38]	Hopeless or disappointed	[38]
Suicidal interest	[38]	Corona affection	[41]
Relation with family members	[38]	Family Corona affection	[41]
Financial status	[35]	Financial damage by Corona	[41]
Relation with classmates and teachers	[36]	Interruption of study by Corona	[41]
Dead by Corona in family	[41]	Jobless for Corona	[41]

Figure 3.17 shows the scatter plot matrix of four features. The features are decision making, inattention or forgetfulness, insecure by family and educational life happiness. Figure 3.17 is shown below.

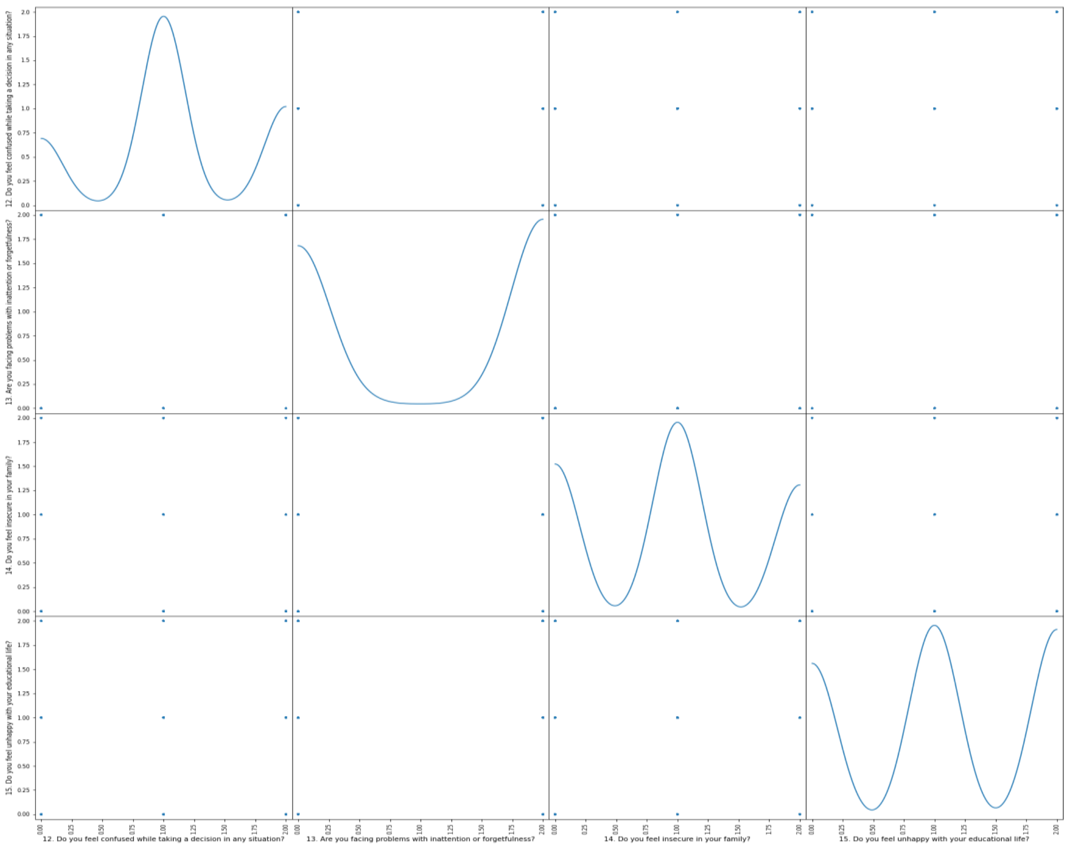


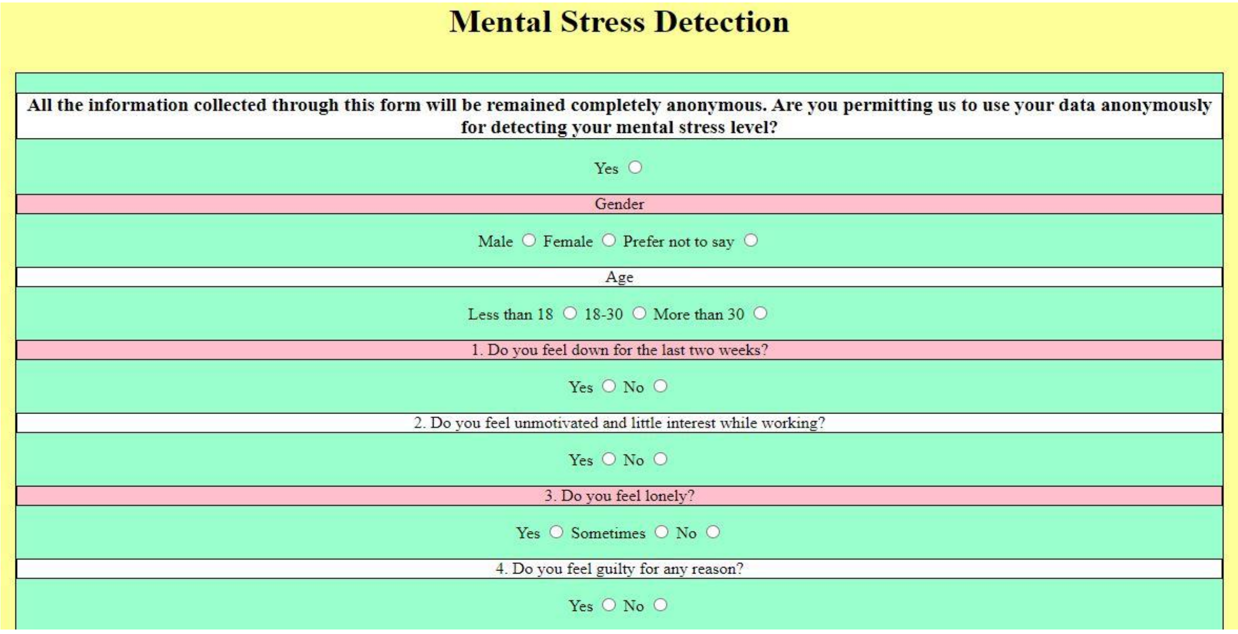
Figure 3.17: Scatter Plot Matrix

3.5 Implementation Equipment

To do this effort, we will need platforms for data mining, platforms for data storing, platforms for data processing and platforms for data presentation. We used Microsoft Excel to store data since it is gathered using handwritten forms and google forms. Google docs and exploratory data analysis are sometimes used to understand data trends for data visualization. We primarily utilized google colab and occasionally jupyter notebook for the data pretreatment and for model implementation we used Django framework.

3.6 Implementation Work

Figure 3.18 shows the form-based user interface of our model. First of all, we decided to deploy our model as a web-based application using the Django framework. Then, we trained our model (*k*-NN) with the dataset. Then using Django and HTML, CSS we successfully done the implementation of our model.



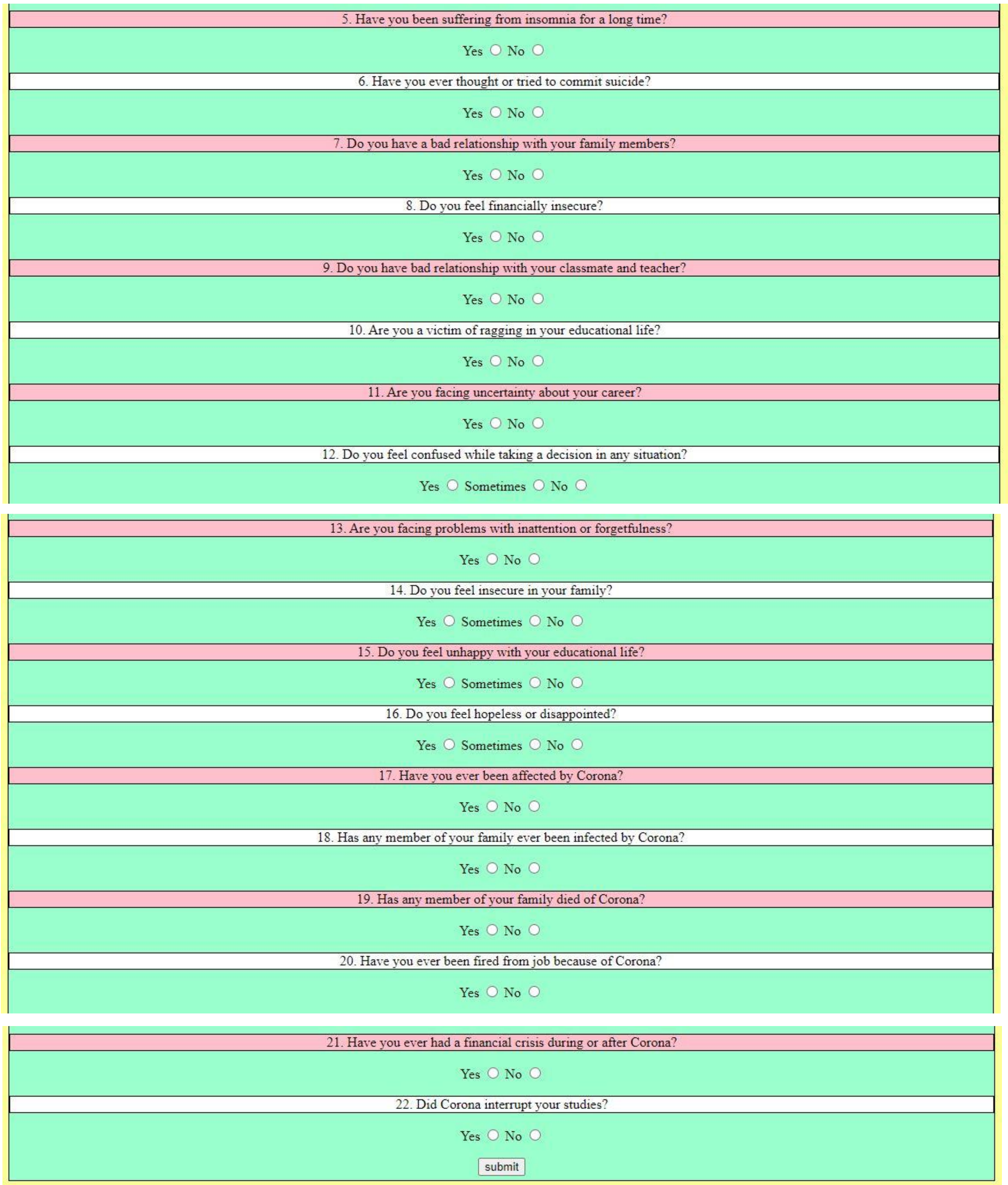


Figure 3.18: Form-based User Interface

Figure 3.19 shows the prediction result of our model. We tested our model a lot of time and Alhamdulillah! We have got the accurate predictions every time.

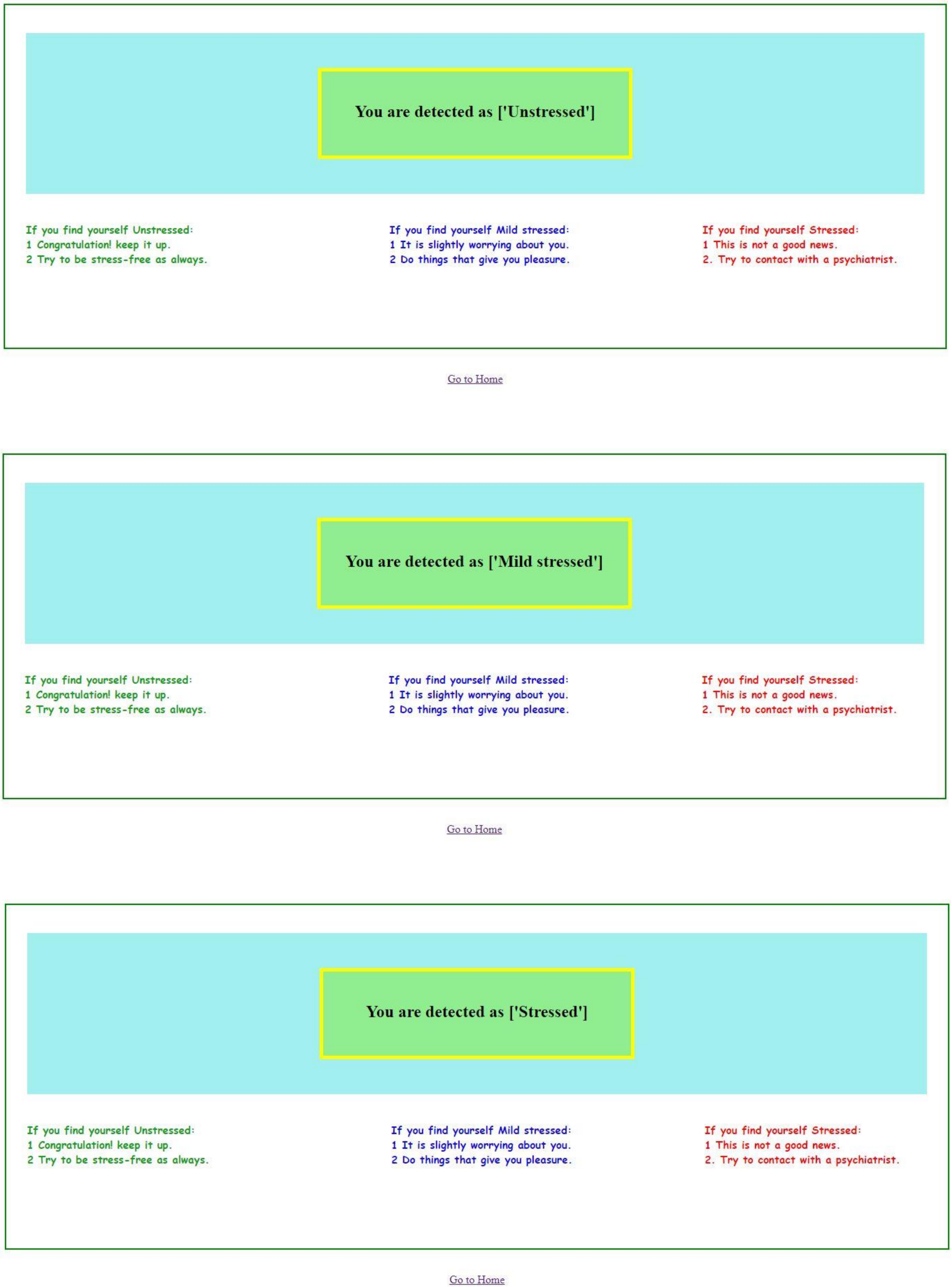


Figure 3.19: Prediction Result of the Model

CHAPTER 4

RESULTS AND DISCUSSION

4.1 Introduction

In our dataset, we applied some machine learning techniques, and we obtained some findings. We will talk about the procedures and the outcomes in this part. A number of techniques were utilized, including LR, RF, SVM, *k*-NN, GB, XGBoost, DT and NB. We obtained some findings after applying them and we were able to determine which method provided us with greater accuracy. The accuracy is computed after preprocessing and feature engineering techniques are applied to processed data. A total of 830 data points from stressed, mild stressed and unstressed people were collected. The following steps were taken in order to calculate data with the highest accuracy: Initially, we divided the data into training and testing, using 80% of the data for training and rest of the 20% for testing. After dividing the total data into training and testing sets, we determined the algorithms' accuracy.

4.2 Experimental Analysis

Eight machine learning algorithms were applied. Then, using confusion matrices a few classifications were calculated such as sensitivity, specificity, precision, recall, F_1-score and accuracy, we compared them with one another. The entire dataset was divided into two categories for calculation: training set and testing set.

4.2.1 Experiment for Evaluation

Twenty-four features make up our dataset. We used eight machine learning methods with feature numbers of 18 for the processed dataset.

The accuracy of the eight algorithms on our dataset, which had 18 features, is shown in Figure 4.1, with the best accuracy of 87.3% being achieved by SVM.

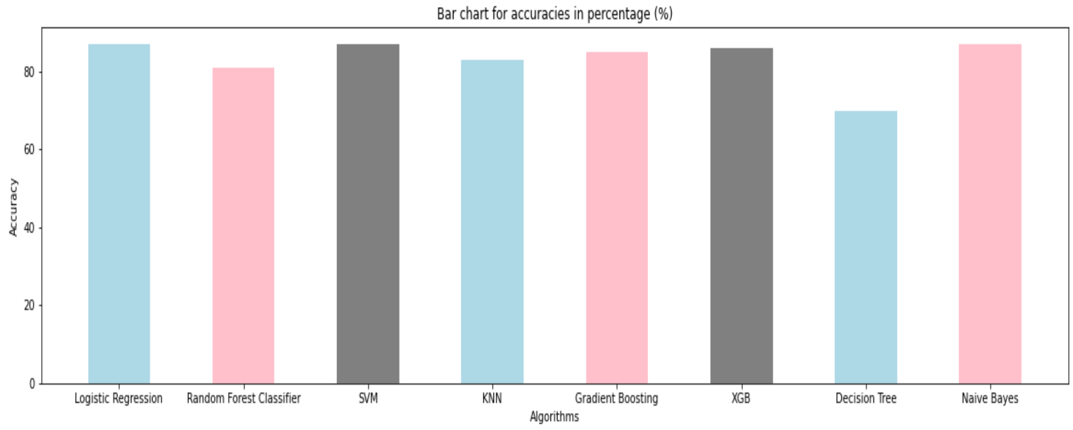


Figure 4.1: Accuracy for Dataset

Logistic Regression is a machine learning algorithm. It is used to address classification problems similar to linear regression. It is a predictive analysis-based approach with probability serving as its main tenet.

In classification and regression problems, supervised machine learning methods like Random Forest classifiers are often used. It builds decision trees from different samples using their average for categorization and majority vote for regression.

Support Vector Machine is an algorithm for supervised machine learning. It can resolve both linear and nonlinear classification and regression problems. However, SVM are commonly used for classification tasks because it uses less computation to provide results that are remarkably accurate. It is advantageous since it generates reliable findings even with scant information.

The *k*-Nearest Neighbor algorithm is one of the simplest categorization methods. It is a supervised machine learning algorithm. *K*-NN saves all of the available cases while classifying new instances based on similarity criteria.

Among other things, classification and regression tasks are accomplished using a machine learning technique called Gradient Boosting. It offers a prediction model in the form of a collection of weak prediction models that resemble decision trees.

The tree-based ensemble machine learning method Extreme Gradient Boosting is a scalable machine learning system for tree boosting. The main reasons of using XGBoost are speed and model performance.

A Decision Tree is a non-parametric supervised learning technique for classification and regression. The goal is to learn simple decision rules based on data attributes in order to develop a model that predicts the value of a target variable.

Naïve Bayes classifiers are a type of classification algorithm based on the Bayes' Theorem. It is more of a family of algorithms than a single approach, and they are all founded on the premise that each pair of features being classified is independent on the other.

The accuracy of all methods is shown in Table 4.1 below.

Table 4.1: Algorithms' Accuracy for Dataset

Algorithm	Accuracy
LR	86.7%
RF	84.3%
SVM	87.3%
k-NN	82.5%
GB	84.9%
XGBoost	85.5%
DT	71.7%
NB	86.7%

4.2.2 Expressive Analysis

Numerous algorithms' accuracy was assessed and we also evaluated each algorithm's precision, recall, F_1-score, accuracy, sensitivity, specificity, confusion matrix and ROC curve. Any model selection demands a model evaluation. Measurements of specific categories are required in the case of model evaluation. For advanced measures, classification measurements are based on the test data sets.

Sensitivity is the ability of a test to correctly identify the true positive rate.

$$\text{Sensitivity} = \frac{TP}{TP+FN} \times 100\% \qquad \text{(i)}$$

Specificity is the ability of a test to correctly identify the true negative rate.

$$\text{Specificity} = \frac{TN}{FP+TN} \times 100\% \qquad \text{(ii)}$$

Recall literally is how many of the true positives were found. That is why it is the ratio of true positive value and predicted negative value.

$$\text{Recall} = \frac{TP}{TP+FN} \times 100\% \qquad \text{(iii)}$$

Precision means how many of the positively classified were relevant. It is all about the ratio of true positive and predicted positive value. A test can maximize this by only returning positive on one result it is most confident in.

$$\text{Precision} = \frac{TP}{TP+FP} \times 100\% \qquad \text{(iv)}$$

F_1-score is the weighted average of precision and recall. Subsequently, this score considers both false positives and false negatives into account.

$$F_1\text{-score} = \frac{2 \times Precision \times Recall}{Precision + Recall} \times 100\% \qquad \text{(v)}$$

For the visual comparison of curve classification models, ROC curves are particularly helpful. True positive and false positive rates were used to create the ROC curve. A guess is represented by the diagonal line. A less accurate model is one whose curve resembles a random estimation. Here are a few ROC curves generated for our method.

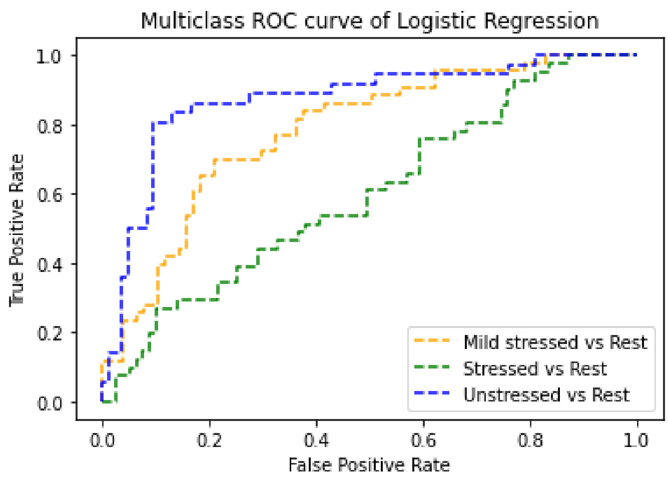


Figure 4.2: ROC Curve of LR

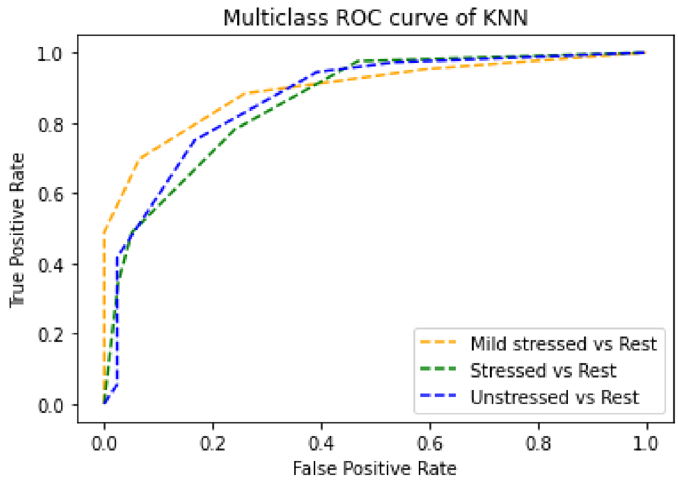


Figure 4.3: ROC Curve of *k*-NN

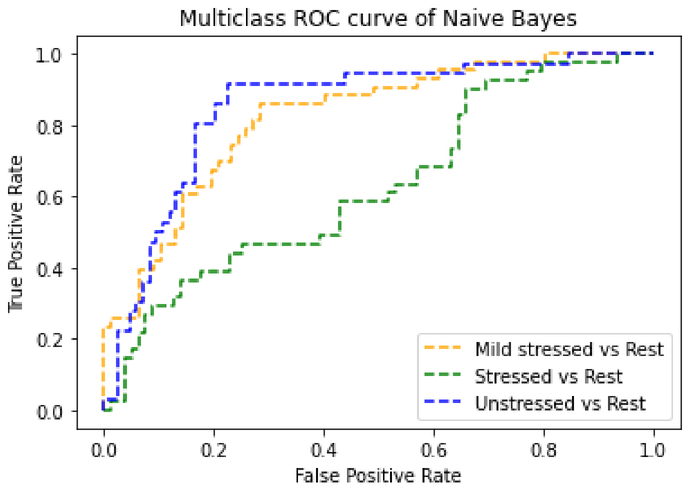
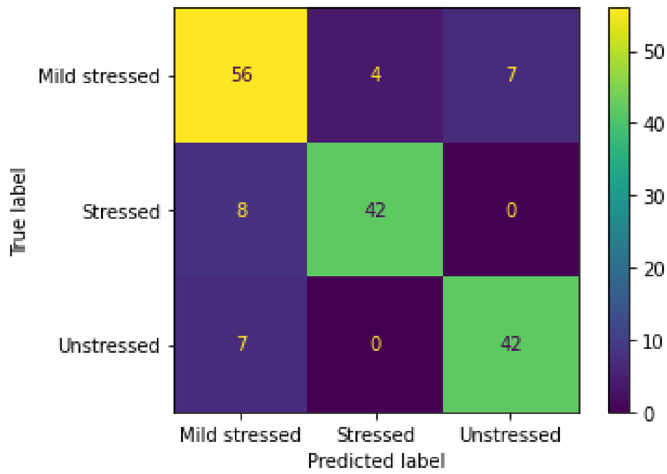
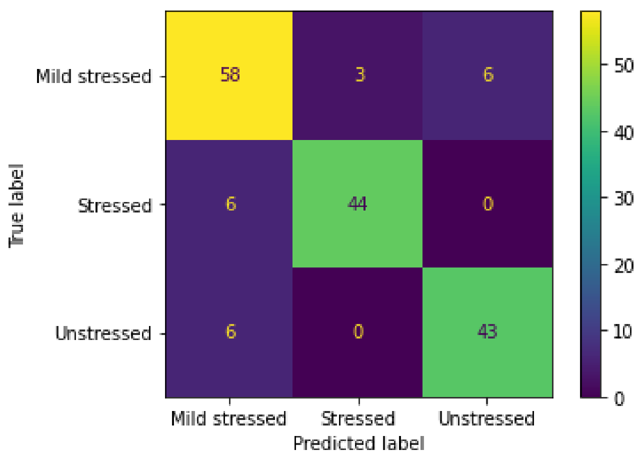


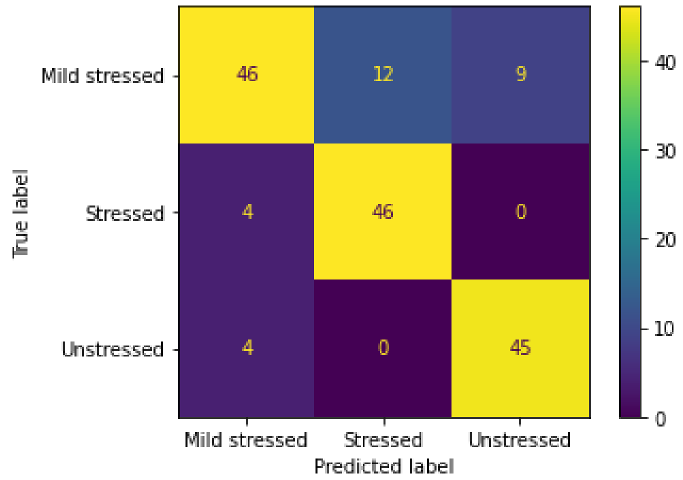
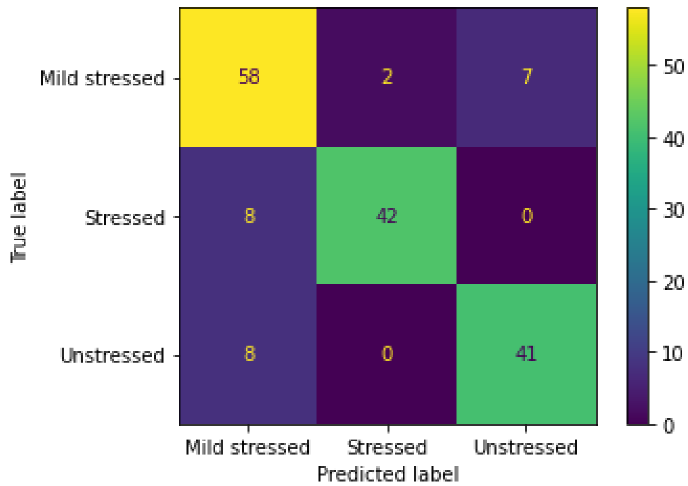
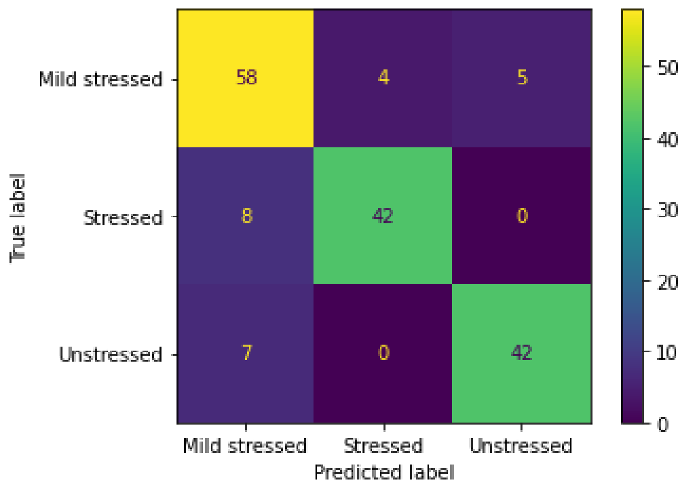
Figure 4.4: ROC Curve of NB

A method of predicting outcomes on a classification problem using machine learning is the confusion matrix. This is compared to the actual goal values that the machine learning model had predicted. We can see at a glance how well our categorization model is performing. We can also determine the types of errors we have. It is crucial in determining how effective a classifier is.

The confusion matrix of the methods we utilized is shown in Table 4.2. The table below provides an accurate description of each classification.

Table 4.2: Confusion Matrix of all Classifiers

<table>
<tr><th>Algorithm</th><th>Confusion Matrix</th></tr>
<tr><td>LR</td><td>
</td></tr>
<tr><td>RF</td><td>
</td></tr>
<tr><td>SVM</td><td>
</td></tr>
</table>

<table>
<tr><td>*k*-NN</td><td>
</td></tr>
<tr><td>GB</td><td>
</td></tr>
<tr><td>XGBoost</td><td>
</td></tr>
</table>

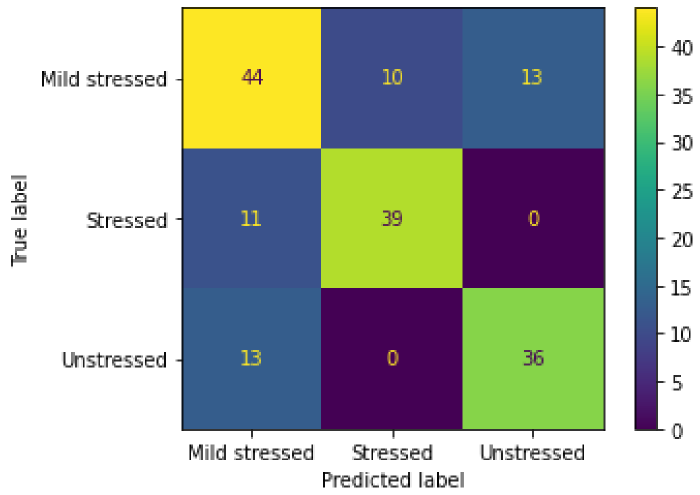
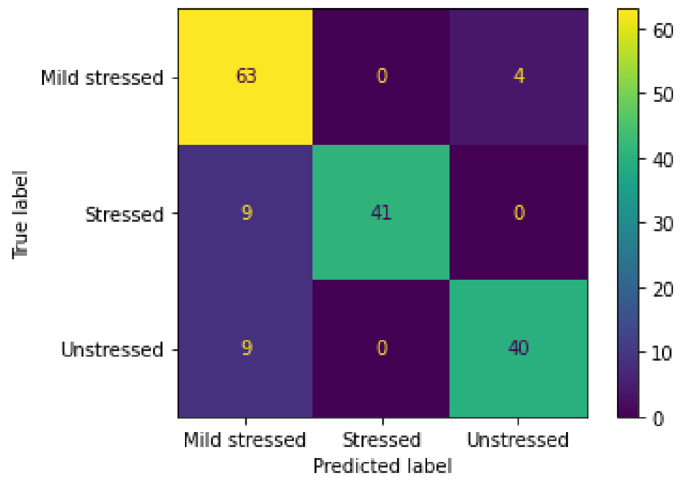
<table>
<tr><td>DT</td><td>
</td></tr>
<tr><td>NB</td><td>
</td></tr>
</table>

Table 4.3 shows the performance of each algorithm. The accuracy and performance of the algorithms will be utilized to determine which method is optimal for our model. LR is clearly the finest in terms of precision, specificity and accuracy.

Table 4.3: Performance of Classifier

Algorithm	Classes	Class-wise Accuracy (%)	Sensitivity (%)	Specificity (%)	Precision (%)	Recall (%)	F_1-score (%)	Macro F_1-score (%)
LR	Mild stressed	86.7	80.6	90.9	86	81	83	87
	Stressed	94	92	94.8	88	92	90	
	Unstressed	92.8	89.8	94	86	90	88	
RF	Mild stressed	84.9	85.1	84.8	79	84	81	84
	Stressed	92.8	86	95.7	91	84	87	
	Unstressed	92.2	83.7	95.7	86	86	86	
SVM	Mild stressed	87.3	86.6	87.9	83	87	85	88
	Stressed	94.6	88	97.4	94	88	91	
	Unstressed	92.8	87.8	94.9	88	88	88	
k-NN	Mild stressed	82.5	68.7	91.9	85	69	76	83
	Stressed	90.4	92	89.7	79	92	85	
	Unstressed	92.2	91.8	92.3	83	92	87	
GB	Mild stressed	84.9	86.6	83.8	78	87	82	85
	Stressed	94	84	98.3	95	84	89	
	Unstressed	91	83.7	94	85	84	85	
XGBoost	Mild stressed	85.5	86.6	84.8	79	87	83	86
	Stressed	92.8	84	96.6	91	84	87	
	Unstressed	92.8	85.7	95.7	89	86	88	
DT	Mild stressed	72.3	68.7	74.7	65	66	65	72
	Stressed	86.7	74	92.2	80	78	79	
	Unstressed	85.5	75.5	89.7	73	73	73	
NB	Mild stressed	86.7	94	81.8	78	84	85	87
	Stressed	94.6	82	100	100	82	90	
	Unstressed	92.2	81.6	96.6	91	82	86	

4.3 Comparative Performance Analysis

Our research aims to detect mental stress in Bangladeshi people. In study [13], 1000 college students' text and emoji data were gathered in order to identify mental stress using six criteria. Mental stress level identification using 35,887 photos is described in paper [16]. Mental stress detection using text and question-based data is discussed in study [17]. Mental stress was identified in papers [18] and [11] using 10,000 tweets. But in research [11], sentiment analysis was carried out employing datasets related to Bengali depression. In the research [20], depression was examined using Bangla-language social media datasets, although the use of algorithms was not mentioned. In paper [21], depression is identified using face photos, although classifiers and accuracy are not mentioned. Using feature engineering, depression was identified in study [22]. In paper [14], verbal cues are used to detect depression. With thirteen characteristics employed, their accuracy was 81.567%. In study [16], CNN and recurrent neural network (RNN) algorithms are used to diagnose depression. In article [26], they develop an emotive chatbot without mentioning features or accuracy. A summary of our work and other works can be found in Table 4.4.

Table 4.4: Comparative Performance Analysis

Method/ Work done	Object(s)/Deal with	Data Type	Problem Domain	Sample Size	Feature Selection Applied	Algorithm	Accuracy
Our work	Mental stress	Survey based text	Detection	830 Records	✓	LR	86.7%
Ding et al. [13]	Depression	Text, Emoji	Recognition	1000 Users	✓	DISVM	86.15%
Mulay et al. [12]	Depression	Image	Detection	35,887 Images	✗	CNN	66.45%
Asad et al. [17]	Depression	Text, Question	Detection	150 Users	✓	SVM, NB	74%
Deshpande et al. [18]	Depression	Text, Question	Detection	10,000 Tweets	✓	SVM, NB	NB- 83%
Khan et al. [11]	Depression	Text	Detection	10,000 Tweets	✓	SVM, Multinomial NB	NB- 86%
Zhou et al. [21]	Depression	Video	Recognition	490 Videos (Dataset AVEC 2013 and 2014)	✓	[2]*NA*	[1]*NM*
Stankevich et al. [42]	Depression	Text	Detection	887 Reddit users	✓	SVM, RF	63%
Patel et al. [23]	Chatbot	User text analysis	Recognition	ISEAR dataset	✓	CNN, RNN and Hierarchical Attention Network (HAN)	75%
Shukla et al. [14]	[1]*NM*	Speech signals of persons	Detection	Ryerson Audio-Visual Database of Emotional Speech and Song	✓	Multi-Layer Perceptron (MLP)	81.567%
Orabi et al. [16]	Depression	Text	Detection	1,145 Twitter users	✓	CNN and RNN	87.957%
Ranade et al. [26]	Chatbot	Based on emotion	Recognition	[2]*NA*	✓	[2]*NA*	[2]*NA*

[1]*NM:* Not Mentioned

[2]*NA:* Not Applicable

4.4 Discussion

In this section, the representation of algorithms, specificity, accuracy, recall, sensitivity and precision, as well as the ROC curve and F_1-score are examined. Here, it is also described how evolutionary models work mathematically and how effective they are. We can observe that the maximum accuracy provided by the LR technique is 86.7%. Additionally, 92% sensitivity, 94.8% specificity, 94% accuracy, 92% recall and 90% F_1-score were attained by the LR technique. Finally, we find that the highest performance for our mental stress detection model may be obtained using LR.

CHAPTER 5

IMPACT ON SOCIETY, ENVIRONMENT AND SUSTAINABILITY

5.1 Impact on Society

Mental health is an important part of human life. Most people think that people with mental disorders are dangerous. However, many people have mental disorders and they contribute greatly to the society. People with mental disorders affect every part of society with their mental health as well as physical health. Many people have mental health issues including mental stress, anxiety and schizophrenia. These people are affected by excessive stress and are unable to function normally. Many of these issues are caused by societal problems such as bullying or financial issues. Societal issues caused by mental stress include crime and divisiveness. Many people who have mental health problems attempt to take care of themselves by buying drugs or gambling. This causes them to gain money and causes trouble for themselves and the country in general. Individuals can control their mental stress by exercising and eating well. Exercise helps people cope with anxiety; it gives them a sense of accomplishment and reduces their level of anxiety hormones such as cortisol. Eating well reduces stress by providing your body with the nutrients it needs to function well. Your physical condition depends on how well you eat and exercise; if you do not, your body will not get enough nutrients to function properly. Basically, you should live your life in such a way that you reduce your mental stress level as much as possible. It is tough to help people with mental disorders who are also part of the society; this makes things even worse for everyone involved. Humans are capable of greatness- when everyone is able to focus on themselves without being affected by excess stress, amazing things happen in the world.

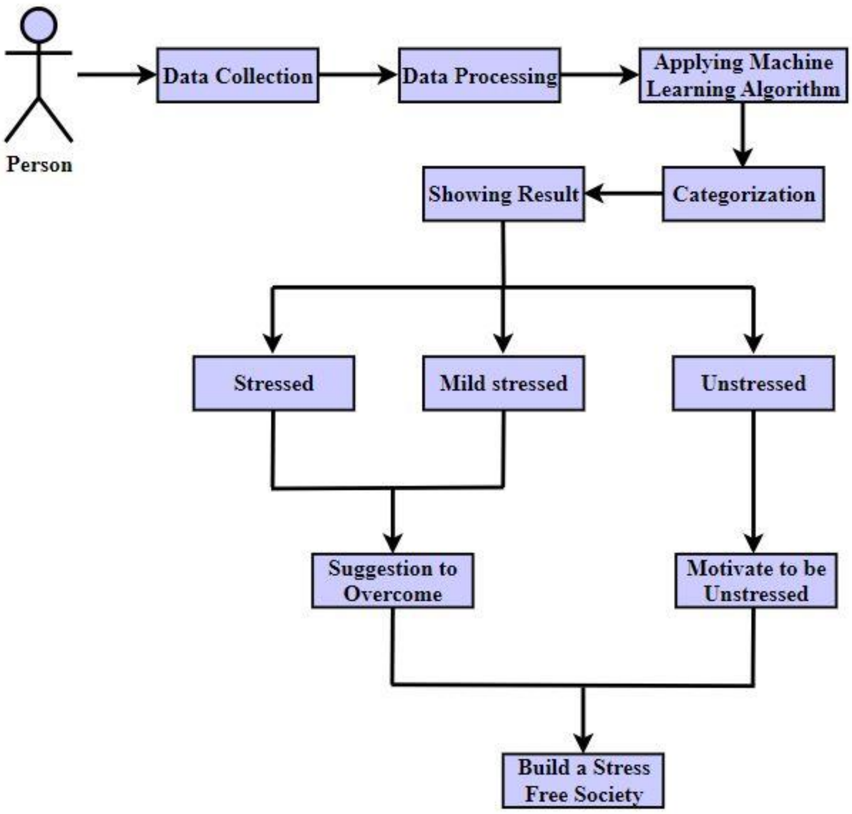


Figure 5.1: Impact on Society

5.2 Ethical Aspects

We have a long way to go before we can complete our whole mental stress model. However, we can tell you that it does not use any personal information, name or identity that could be damaging to someone. As a result, there will be no breach of privacy. We can also guarantee that this model does not interfere with a person's right to enjoy but rather serves an important part in raising awareness. As a result, the model of mental stress detection may be easily handled utilizing machine learning technology.

5.3 Sustainability Plan

The Sustainability Plan provides us with practical suggestions for any project and our long-term goals. Our model's goal is to determine the likelihood of stress. It is important to be aware of this concept so that individuals can quickly adapt. Additionally, it is crucial to make sure that people understand their positions before applying this paradigm. This paradigm can help psychologists, psychiatrists and mental health organizations to complete their tasks more quickly.

CHAPTER 6

SUMMARY, CONCLUSION AND IMPLICATION FOR FUTURE RESEARCH

6.1 Summary and Conclusion of the Work

Our effort is divided into several stages, including data collection, data preprocessing, model implementation and evaluation. We acquired the necessary information from schools, colleges and universities, public places and online sources. We used to google colab after data collection to work on data cleaning and algorithm development. After preprocessing the data, we used LR, RF, SVM, *k*-NN, GB, XGBoost, DT and NB as our eight machine learning methods. Furthermore, the accuracy, precision, sensitivity, recall and F_1-score of each approach are calculated and considered for determining the best model. Notably, the LR method produces the best results. But in order to detect stress in our model we have used the *k*-NN algorithm. Because *k*-NN is the best for multiclass classification approach.

6.2 Limitations of the Work

We created a machine learning algorithm to detect stress. Our research and models have various flaws and limitations. Because the level we chose was fairly small, it might be preferable to utilize a dataset separated into numerous levels. Due to several constraints, we were unable to collect data on a wide range of people. The dataset is also not that large. For data processing and preprocessing, a variety of highly developed, cutting-edge methods and models can be used, and the model or system can then be wonderfully displayed using a variety of cutting-edge and well-liked algorithms. It is unquestionably possible to detect the stress condition using our provided model. When this model is completed, we are confident that people will be able to use it more easily and understand its importance and value in assuring emotional state. We are hopeful and sure that our model will appropriately reflect stress, prompting people to recognize their disease and move towards a cure.

6.3 Implication for Further Study

Modern technology, data science and artificial intelligence have recently accelerated, simplified and enhanced all aspects of human life. In the future, we hope to turn our concept into an Android version. In the future, we will try to increase the accuracy of our models. We will create a larger dataset with a lot of people's data and add more categorization layers to our dataset. Furthermore, by offering an immersive user-friendly GUI, the mobile application produced utilizing the concept may be made accessible to everyone, even doctors. Implementing new methods, adding additional parameters and other characteristics to produce a robust dataset can significantly improve the model's effectiveness and practicality. In the future, a robust dataset can be developed by collecting data from more different classifications of people based on location, age and activities. In addition, the Department of Mental Health can be helped with model expansion.

APPENDIX A

Appendix A1: Google form

https://forms.gle/JWxL8XgnTYwWSzYUA

Appendix A2: Raw dataset

https://docs.google.com/spreadsheets/d/13X5NsbLiZkrlz4uKTluLZBWOJjevDQYVru7CN18BbaQ/edit?usp=sharing

Appendix A3: Processed dataset

https://docs.google.com/spreadsheets/d/1mXGfKhO1okiv_J5LaP4B9vCJiB5HCqj-FEOcP7I5Hnc/edit?usp=sharing

Appendix A4: Google colab

https://colab.research.google.com/drive/13tEqadGIGXos3_935QdpjduXdki3POvz?usp=sharing

REFERENCES

[1] "What is depression? | UNICEF Parenting." https://www.unicef.org/parenting/mental-health/what-is-depression?gclid=CjwKCAiAqaWdBhAvEiwAGAQltk1rLlWniSDBGwav7V7FYdUxpfrPH12GF_4SuAiK1mdQ_Jqaqjj4ExoCfLwQAvD_BwE (accessed Dec. 26, 2022).

[2] N. Bayram and N. Bilgel, "The prevalence and socio-demographic correlations of depression, anxiety and stress among a group of university students," *Soc. Psychiatry Psychiatr. Epidemiol.*, vol. 43, no. 8, pp. 667–672, 2008, doi: 10.1007/s00127-008-0345-x.

[3] M. A. I. Arif, S. I. Sany, F. Sharmin, M. S. Rahman, and M. T. Habib, "Prediction of addiction to drugs and alcohol using machine learning: A case study on Bangladeshi population," *Int. J. Electr. Comput. Eng.*, vol. 11, no. 5, pp. 4471–4480, 2021, doi: 10.11591/ijece.v11i5.pp4471-4480.

[4] M. D. Hossain, H. U. Ahmed, W. A. Chowdhury, L. W. Niessen, and D. S. Alam, "Mental disorders in Bangladesh: A systematic review," *BMC Psychiatry*, vol. 14, no. 1, pp. 1–8, 2014, doi: 10.1186/s12888-014-0216-9.

[5] M. Akter and F. Ahmed, "A Machine Learning Approach To Predict Social Media Addiction During COVID-19 Pandemic," no. Icaaic, pp. 401–405, 2022.

[6] M. Z. Islam, Z. Jannat, M. T. Habib, M. S. Rahman, and G. Z. Islam, *Detection of Facebook Addiction Using Machine Learning*, vol. 514 LNNS. Springer International Publishing, 2022.

[7] A. R. Subhani, W. Mumtaz, M. N. B. M. Saad, N. Kamel, and A. S. Malik, "Machine learning framework for the detection of mental stress at multiple levels," *IEEE Access*, vol. 5, no. c, pp. 13545–13556, 2017, doi: 10.1109/ACCESS.2017.2723622.

[8] B. Hu *et al.*, "Signal Quality Assessment Model for Wearable EEG Sensor on Prediction of Mental Stress," *IEEE Trans. Nanobioscience*, vol. 14, no. 5, pp. 553–561, 2015, doi: 10.1109/TNB.2015.2420576.

[9] J. Choi, B. Ahmed, and R. Gutierrez-Osuna, "Development and evaluation of an ambulatory stress monitor based on wearable sensors," *IEEE Trans. Inf. Technol. Biomed.*, vol. 16, no. 2, pp. 279–286, 2012, doi: 10.1109/TITB.2011.2169804.

[10] J. Choi and R. Gutierrez-Osuna, "Removal of respiratory influences from heart rate variability in stress monitoring," *IEEE Sens. J.*, vol. 11, no. 11, pp. 2649–2656, 2011, doi: 10.1109/JSEN.2011.2150746.

[11] R. H. Khan, "Sentiment Analysis from Bengali Depression Dataset using Machine Learning," 2020.

[12] A. Mulay, A. Dhekne, R. Wani, S. Kadam, P. Deshpande, and P. Deshpande, "Automatic

depression level detection through visual input," *Proc. World Conf. Smart Trends Syst. Secur. Sustain. WS4 2020*, pp. 19–22, 2020, doi: 10.1109/WorldS450073.2020.9210301.

[13] Y. DIng, X. Chen, Q. Fu, and S. Zhong, "A Depression Recognition Method for College Students Using Deep Integrated Support Vector Algorithm," *IEEE Access*, vol. 8, pp. 75616–75629, 2020, doi: 10.1109/ACCESS.2020.2987523.

[14] Di. M. Shukla, K. Sharma, and S. Gupta, "Identifying Depression in a Person Using Speech Signals by Extracting Energy and Statistical Features," *2020 IEEE Int. Students' Conf. Electr. Electron. Comput. Sci. SCEECS 2020*, pp. 5–8, 2020, doi: 10.1109/SCEECS48394.2020.60.

[15] X. Yang, D. Fan, A. Ren, N. Zhao, and M. Alam, "5G-Based user-centric sensing at C-band," *IEEE Trans. Ind. Informatics*, vol. 15, no. 5, pp. 3040–3047, 2019, doi: 10.1109/TII.2019.2891738.

[16] A. H. Orabi, P. Buddhitha, M. H. Orabi, and D. Inkpen, "Deep Learning for Depression Detection of Twitter Users," pp. 88–97, 2018.

[17] A. M. Pranto, "Depression Detection by Analyzing Social Media Posts of User," *2019 IEEE Int. Conf. Signal Process. Information, Commun. Syst.*, pp. 13–17, 2019.

[18] M. Deshpande and V. Rao, "Depression detection using emotion artificial intelligence," *Proc. Int. Conf. Intell. Sustain. Syst. ICISS 2017*, no. Iciss, pp. 858–862, 2018, doi: 10.1109/ISS1.2017.8389299.

[19] N. Z. Gurel, H. Jung, S. Hersek, and O. T. Inan, "Fusing Near-Infrared Spectroscopy with Wearable Hemodynamic Measurements Improves Classification of Mental Stress," *IEEE Sens. J.*, vol. 19, no. 19, pp. 8522–8531, 2019, doi: 10.1109/JSEN.2018.2872651.

[20] Æ. N. Schmidt, Æ. J. Babin, and M. Pharr, "Coping with Stress : The Effectiveness of Different Types of Music," pp. 163–168, 2007, doi: 10.1007/s10484-007-9043-9.

[21] X. Zhou, K. Jin, Y. Shang, and G. Guo, "Visually Interpretable Representation Learning for Depression Recognition from Facial Images," *IEEE Trans. Affect. Comput.*, vol. 11, no. 3, pp. 542–552, 2020, doi: 10.1109/TAFFC.2018.2828819.

[22] R. Castaldo, P. Melillo, U. Bracale, M. Caserta, M. Triassi, and L. Pecchia, "Biomedical Signal Processing and Control Acute mental stress assessment via short term HRV analysis in healthy adults : A systematic review with meta-analysis," *Biomed. Signal Process. Control*, vol. 18, pp. 370–377, 2015, doi: 10.1016/j.bspc.2015.02.012.

[23] F. Patel, R. Thakore, I. Nandwani, and S. K. Bharti, "Combating depression in students using an intelligent ChatBot: A cognitive behavioral therapy," *2019 IEEE 16th India Counc. Int. Conf. INDICON 2019 - Symp. Proc.*, Dec. 2019, doi: 10.1109/INDICON47234.2019.9030346.

[24] U. Lundberg, "Stress, subjective and objective health," *Int. J. Soc. Welf.*, vol. 15, no. SUPPL. 1, pp. 41–48, 2006, doi: 10.1111/j.1468-2397.2006.00443.x.

[25] M. C. Pfaltz, P. Grossman, T. Michael, J. Margraf, and F. H. Wilhelm, “Physical activity and respiratory behavior in daily life of patients with panic disorder and healthy controls,” *Int. J. Psychophysiol.*, vol. 78, no. 1, pp. 42–49, 2010, doi: 10.1016/j.ijpsycho.2010.05.001.

[26] A. G. Ranade, M. Patel, and A. Magare, “Emotion model for artificial intelligence and their applications,” *PDGC 2018 - 2018 5th Int. Conf. Parallel, Distrib. Grid Comput.*, pp. 335–339, 2018, doi: 10.1109/PDGC.2018.8745840.

[27] D. Kim, Y. Seo, J. Cho, and C. H. Cho, “Detection of subjects with higher self-reporting stress scores using heart rate variability patterns during the day,” *Proc. 30th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. EMBS’08 - "Personalized Healthc. through Technol.*, pp. 682–685, 2008, doi: 10.1109/iembs.2008.4649244.

[28] M. Kumar, M. Weippert, R. Vilbrandt, S. Kreuzfeld, and R. Stoll, “Fuzzy evaluation of heart rate signals for mental stress assessment,” *IEEE Trans. Fuzzy Syst.*, vol. 15, no. 5, pp. 791–808, 2007, doi: 10.1109/TFUZZ.2006.889825.

[29] R. Costin, C. Rotariu, and A. Pasarica, “Mental stress detection using heart rate variability and morphologic variability of EeG signals,” *EPE 2012 - Proc. 2012 Int. Conf. Expo. Electr. Power Eng.*, no. Epe, pp. 591–596, 2012, doi: 10.1109/ICEPE.2012.6463870.

[30] S. Boonnithi and S. Phongsuphap, “Comparison of heart rate variability measures for mental stress detection,” *Comput. Cardiol. (2010).*, vol. 38, no. January 2011, pp. 85–88, 2011.

[31] K. W. Murdock, A. S. LeRoy, and C. P. Fagundes, “Trait hostility and cortisol sensitivity following a stressor: The moderating role of stress-induced heart rate variability,” *Psychoneuroendocrinology*, vol. 75, pp. 222–227, 2017, doi: 10.1016/j.psyneuen.2016.10.014.

[32] S. Sharma, G. Singh, and M. Sharma, “A comprehensive review and analysis of supervised-learning and soft computing techniques for stress diagnosis in humans,” *Comput. Biol. Med.*, vol. 134, no. May, p. 104450, 2021, doi: 10.1016/j.compbiomed.2021.104450.

[33] R. Castaldo *et al.*, “Detection of Mental Stress due to Oral Academic Examination via Ultra-short-term HRV Analysis,” pp. 3805–3808, 2016.

[34] R. Castaldo, L. Montesinos, P. Melillo, C. James, and L. Pecchia, “Ultra-short term HRV features as surrogates of short term HRV : a case study on mental stress detection in real life,” pp. 1–13, 2019.

[35] “Causes of Depression: Genetics, Illness, Abuse, and More.” https://www.webmd.com/depression/guide/causes-depression (accessed Jan. 03, 2023).

[36] “Causes of Depression: Genetics, Illness, Abuse, and More.” https://www.webmd.com/depression/guide/causes-depression (accessed Dec. 26, 2022).

[37] “Depression (major depressive disorder) - Symptoms and causes - Mayo Clinic.” https://www.mayoclinic.org/diseases-conditions/depression/symptoms-causes/syc-20356007

(accessed Dec. 26, 2022).

[38] I. M. Cameron, J. R. Crawford, K. Lawton, and I. C. Reid, “Psychometric comparison of PHQ-9 and HADS for measuring depression severity in primary care,” *Br. J. Gen. Pract.*, vol. 58, no. 546, pp. 32–36, Jan. 2008, doi: 10.3399/BJGP08X263794.

[39] “Causes of depression - Mind.” https://www.mind.org.uk/information-support/types-of-mental-health-problems/depression/causes/ (accessed Dec. 26, 2022).

[40] “Causes - Clinical depression - NHS.” https://www.nhs.uk/mental-health/conditions/clinical-depression/causes/ (accessed Dec. 26, 2022).

[41] “Depression after COVID-19: Causes, statistics, and more.” https://www.medicalnewstoday.com/articles/depression-after-covid#factors (accessed Dec. 26, 2022).

[42] M. Stankevich, V. Isakov, D. Devyatkin, and I. Smirnov, “Feature engineering for depression detection in social media,” *ICPRAM 2018 - Proc. 7th Int. Conf. Pattern Recognit. Appl. Methods*, vol. 2018-Janua, no. Icpram, pp. 426–431, 2018, doi: 10.5220/0006598604260431.

Plagiarism Report

ORIGINALITY REPORT

28%	24%	18%	%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	dspace.daffodilvarsity.edu.bd:8080 Internet Source	10%
2	Khalid Masood, Mohammed A. Alghamdi. "Modeling Mental Stress Using a Deep Learning Framework", IEEE Access, 2019 Publication	3%
3	Putri Azmira R Azmi, Aida Wati Zainal Abidin, Sofianita Mutalib, Iskandar Shah Mohd Zawawi, Shamimi A Halim. "Sentiment Analysis on MySejahtera Application during COVID-19 Pandemic", 2022 3rd International Conference on Artificial Intelligence and Data Sciences (AiDAS), 2022 Publication	<1%
4	link.springer.com Internet Source	<1%
5	Madiha Mansoori, Hrishil Maliwal, Sharvil Kotian, Hersh Kenkre, Ishani Saha, Payal Mishra. "A Systematic Survey on Computational agents for Mental Health Aid",	<1%