Stress Factors Identification and Comparative Study of Stress Levels among Undergraduate Students

by

Samiha Khan
16301165
Mahinur Akter Mousumi
16301206
Noshin Saiyara
16301116
Mehanaz Fatema
16301109

A thesis submitted to the Department of Computer Science and Engineering
in partial fulfillment of the requirements for the degree of
B.Sc. in Computer Science

Department of Computer Science and Engineering
Brac University
April 2020

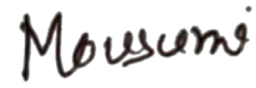
Declaration

It is hereby declared that

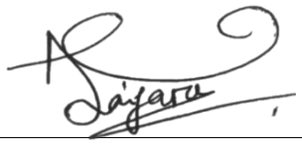
1. The thesis submitted is my/our own original work while completing degree at Brac University.
2. The thesis does not contain material previously published or written by a third party, except where this is appropriately cited through full and accurate referencing.
3. The thesis does not contain material which has been accepted, or submitted, for any other degree or diploma at a university or other institution.
4. We have acknowledged all main sources of help.

Student's Full Name & Signature:

Samiha Khan
16301165



Mahinur Akter Mousumi
16301206



Noshin Saiyara
16301116

Mehanaz Fatema
16301109

Approval

The thesis titled "Stress Factors Identification and Comparative Study of Stress Levels among Undergraduate Students" submitted by

1. Samiha Khan (16301165)
2. Mahinur Akter Mousumi (16301206)
3. Noshin Saiyara (16301116)
4. Mehanaz Fatema (16301109)

Of Spring, 2020 has been accepted as satisfactory in partial fulfillment of the requirement for the degree of B.Sc. in Computer Science on April 2, 2020.

Examining Committee:

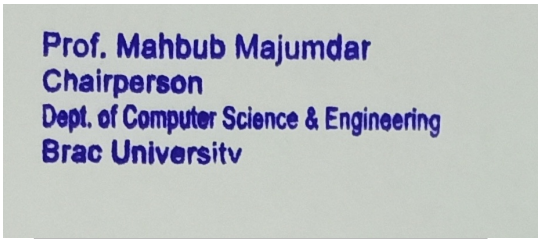
Supervisor:
(Member)

Dr. Md. Golam Rabiul Alam
Associate Professor
Department of Computer Science and Engineering
Brac University

Co-Supervisor:
(Member)

Nadia Rubaiyat
Lecturer
Department of Computer Science and Engineering
Brac University

Head of Department:
(Chair)



Dr. Mahbubul Alam Majumdar
Professor
Department of Computer Science and Engineering
Brac University

Ethics Statement

The proposition is carried out in total compliance with research morals standards, and the codes and hones set by BRAC University. In our proposal we gather the information from primary sources. We collect information from distinctive members and we utilize our possess dataset for our proposition. We are guaranteeing we utilize references and in content citations appropriately. We the four co authors take full duty for the thesis code infringement. For fathoming issues we perused distinctive websites, research paper, youtube tutorials, and different apparatuses. We moreover took offer assistance from our university faculty individuals. At last, we pronounce that we donate credit to each individual from whom we took offer assistance. We did not make any extortion able implies for completing the thesis. Our work is in compliance with the morals standard set by BRAC university.

Abstract

Stress has become a part of the academic life of the students, thanks to the different internal and external demands put on their shoulders constantly. Excessive and protracted stress has consequences that transcend pure dedication to health by causing various diseases, and it's well-known that such stress-related somatic events can affect the standard of life of the students. Many criteria are wont to categorize stress, and this results in many terms for different varieties of stress. We basically focus on the academic stress during this study, and that we are especially inquisitive about measuring the level of the stress and understanding the tutorial background. Using supervised machine learning algorithms and feature extraction methods to find out the classification accuracy of stress factors that students are going through in their university life. We will also present a comparative analysis of the stress level between students in first year and students in fourth year. So as to live the strain we have got used two kinds of scale, one is PSS another one is modified PAS in order to calculate stress level accurately.

Keywords: Stress level, Stress Measurement, PAS ,PSS

Dedication

We would like to commit this inquiry to our guardians. Without their back we may not be able to total. We too need to devote the research to our companions who made a difference to us to do way better and most imperatively the members who made a difference us collecting data from surveys. Uncommonly we need to commit it to those resources who have permitted us to conduct our survey in their classes. In conclusion we need to devote our study to our supervisor who has helped us all through the entire time amid this work.

Acknowledgement

Firstly, we would like to say much appreciated to our cherished family individuals whom we will continuously be obligated to.
Besides, we thank our supervisor Md. Golam Rabiul Alam for all his offer assistance and consistent guideline.
Thirdly, we would like to thank all the faculty individuals and stuffs for giving us with such a superb study environment where we make ourselves better as well as this research appropriately.

Table of Contents

Declaration i

Approval ii

Ethics Statement iii

Abstract iv

Dedication v

Acknowledgment vi

Table of Contents vii

List of Figures ix

List of Tables 1

1 Introduction 2
- 1.1 Background . . . 2
- 1.2 Motivation . . . 2
- 1.3 Symptoms of stress . . . 3
- 1.4 Problem Statement . . . 4
- 1.5 Objective . . . 5
- 1.6 Work plan . . . 6

2 Related Work 8
- 2.1 Literature Review . . . 8

3 Proposed Method 13
- 3.1 System Model . . . 13
- 3.2 Dataset description . . . 13
  - 3.2.1 Questionnaire Description . . . 13
  - 3.2.2 Feature Selection . . . 18
  - 3.2.3 Methodology . . . 18
  - 3.2.4 Dataset Visualization . . . 19
  - 3.2.5 Analysis of Dataset . . . 21

4 Algorithms **29**
4.1 SVM(Support Vector Machine) . . . 29
4.2 Logistic Regression . . . 30
4.3 Decision Tree . . . 30
4.4 Naive Bayes . . . 32
4.5 Random Forest . . . 33
4.6 Chi-Square . . . 33
4.7 Cronbach's Alpha . . . 34
4.8 Shapiro-Wilk . . . 35

5 Result Analysis and Discussion **36**
5.1 Results of Chi-Square Analysis . . . 36
5.2 Results of Support Vector Machine . . . 37
5.3 Results of Logistic Regression . . . 42
5.4 Results of Decision Tree . . . 42
5.5 Results of Random Forest . . . 42
5.6 Results of Naive Bayes . . . 42

6 Conclusion **44**
6.1 Conclusion . . . 44
6.2 Future Work . . . 44

Bibliography **46**

List of Figures

1.1 The most popular reasons of students' stress . . . 3
1.2 Workflow . . . 6

3.1 System Model of Finding Factors . . . 14
3.2 Flowchart of the Research . . . 14
3.3 The ratio of Gender in the dataset . . . 17
3.4 The ratio of University Year in the dataset . . . 18
3.5 Working Methodology . . . 20
3.6 Heatmap correlation matrix for PSS scale . . . 21
3.7 Heatmap correlation matrix for modified PAS scale . . . 22
3.8 PSS scale's QQ plot . . . 23
3.9 Modified PAS scale's QQ Plot . . . 24
3.10 modified PAS stress level for LOW PSS stress level . . . 25
3.11 modified PAS stress level for MODERATE PSS stress level . . . 25
3.12 modified PAS stress level for HIGH PSS stress level . . . 26
3.13 PSS Stress level of 1st year's students . . . 26
3.14 PSS Stress level of 4th year's students . . . 27
3.15 modified PAS stress level of 1st year's students . . . 27
3.16 modified PAS stress level of 4th year's students . . . 28

4.1 PSS scale's decision tree . . . 31

5.1 Feature Importance of PSS Scale . . . 37
5.2 Feature Importance of modified PAS Scale . . . 38
5.3 confusion matrix for PSS Scale(All features) for the svm classifier 0 is LOW Stress level, 1 is MODERATE Stress level and 2 is HIGH Stress level . . . 39
5.4 confusion matrix for PSS Scale(without Q4 feature) for the svm classifier 0 is LOW Stress level, 1 is MODERATE Stress level and 2 is HIGH Stress level . . . 39
5.5 confusion matrix for PSS Scale(without Q4 and Q7 features) for the svm classifier 0 is LOW Stress level, 1 is MODERATE Stress level and 2 is HIGH Stress level . . . 40
5.6 confusion matrix for modified PAS Scale(All features) for the svm classifier 0 is LOW Stress level, 1 is MODERATE Stress level and 2 is HIGH Stress level . . . 40
5.7 confusion matrix for modified PAS Scale(without Q12 feature) for the svm classifier 0 is LOW Stress level, 1 is MODERATE Stress level and 2 is HIGH Stress level . . . 41

5.8 confusion matrix for modified PAS Scale(without Q12 and Q24 features) for the svm classifier 0 is LOW Stress level, 1 is MODERATE Stress level and 2 is HIGH Stress level 41

List of Tables

3.1 Calculation of PSS Scale 15
3.2 Calculation of Modified PAS Scale 17
3.3 Result from Shapiro-Wilk Test 22
3.4 Overall Result from Cronbach's Alpha Test 24
3.5 Feature based Result from Cronbach's Alpha Test 24

4.1 Consistancy of Cronbach's Alpha 34

5.1 Chi-Square of PSS Scale's features 36
5.2 Chi-Square of Modified PAS Scale's features 37
5.3 Least scored(Chi-square) features for both Scale 38
5.4 Accuracy of SVM classifier for PSS Scale 38
5.5 Accuracy of SVM classifier for modified PAS Scale 39
5.6 Accuracy of Logistic Regression classifier for PSS Scale 39
5.7 Accuracy of Logistic Regression classifier for modified PAS Scale . . . 40
5.8 Accuracy of Decision Tree classifier for PSS Scale 42
5.9 Accuracy of Decision Tree classifier for modified PAS Scale 42
5.10 Accuracy of Random Forest classifier for PSS Scale 42
5.11 Accuracy of Random Forest classifier for modified PAS Scale 43
5.12 Accuracy of Naive Bayes classifier for PSS Scale 43
5.13 Accuracy of Naive Bayes classifier for modified PAS Scale 43

Chapter 1

Introduction

1.1 Background

We all endure stress, the anxiety we experience when faced with a replacement situation or a threatening one. Students at universities are not any different from everyone else, because they too are feeling the pressures and strains of living within the world today. They are frequently faced with new circumstances where the end result is often unknown. They can get angry or nervous, sad or disappointed without really knowing why. When young adults get overwhelmed by stress, they often withdraw or hit at others. Stress can cause many mental and physical health issues, and might be caused by a spread of things. Suicide is one among the foremost egregious outcomes of student stress at an academic institute. In 2008, Edison Media Research performed a study of college student pressure at several schools around the U.S.[1] The survey showed 4 by all of 10 university students claim that they feel stressed most often [1]. 4 by the whole of 5 people say they feel stressed the majority of the time[1]. 1 out of 4 students report day-to-day stress and 9 per cent report suicidal thoughts [1]. Looking at the patterns over the last 25 years, the view of the students regarding their own mental health has been gradually deteriorating. [2] In 2010, the view of males and females about their own mental well-being reached the lowest levels in 25 years, dropping from 2009 to 2010, by around 13 per cent. [2] Stress can have a high adverse impact on classroom results. The National College Health Assessment's Spring 2013 edition, where the average age of those studied was 21 years, revealed that nearly half (46.3 per cent) of all undergraduate students feel stressed by their academic obligations.[3] According to the American Anxiety and Depression Association, about 30 percent of university students say that their studies have had an adverse effect for stress.[4]

1.2 Motivation

Stress is a feeling that is generated when we respond to real events and prepare to face a difficult situation with concentration, energy, stamina and increased alertness. Events which trigger stress are called stressors. There are several factors we found that would affect the stress level among university students consisting of curriculum and self-expectation, financial standing, relationship and family support, social assistance and personality. [5] These days the universities need to keep up their great review about financing, quality inspecting and positioning, tall quality of education

and learning. Hence, the university undergraduate students will experience high weight with the change of instruction framework. In this way, the exam-oriented system will lead to push among students since the students claim that getting an "A" within the exam makes a difference to be a great leader and get a great job within the future. Other than being overstepped by expectations within the coursework, students take part in different sorts of activities organized by faculty or university to pick up delicate ability which is appallingly required in any job environment these days. As a result, efficient time administration is crucial for a survivor of students in university life. In 2019 a blog by Sandra Zheng, it has been claimed that a high percentage of students experience stress while at university. She received 34 percent of survey respondents who have felt stress in university for various reasons. [6]

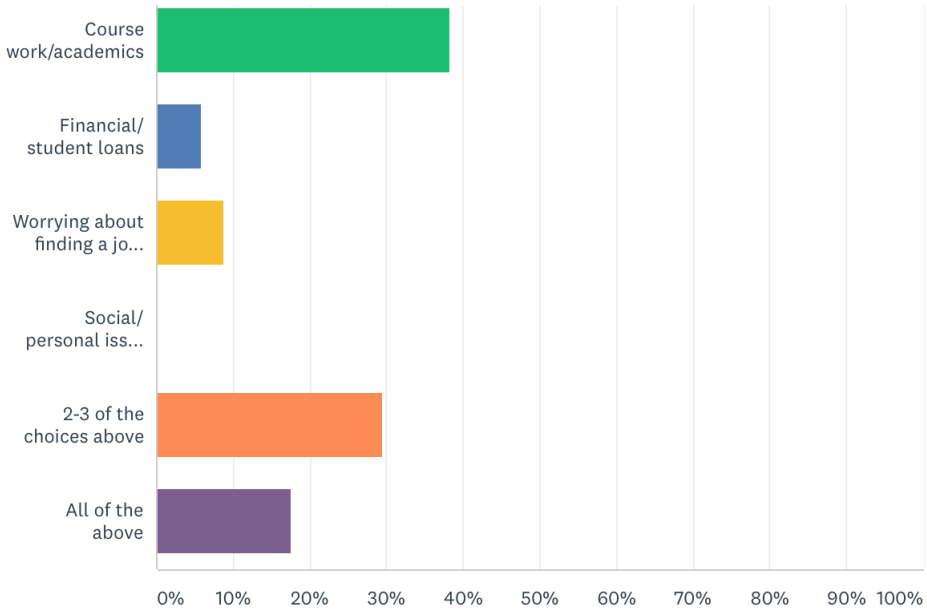


Figure 1.1: The most popular reasons of students' stress

1.3 Symptoms of stress

The upsetting university environment can frequently compound or worsen pre-existing mental well being issues. It is evident that extreme stress is experienced by the most university students while they are completing their graduations. But that can work with a few basic stress management strategies and small lifestyle changes. A Uni Well Well-being think about found that 80% of those who are considering higher instruction detailed side effects of push or uneasiness, whereas one NUS study found that nine in 10 understudies experienced push [7]. All viewpoints of life can get influenced by stress. Stress can also influence counting emotions, behaviors, considering capacity, and corporal health. No portion of the body is safe. But, since everyone has an unique way to handle stress, bad impacts of stress can be changed. The National Collusion on Mental Illness's overview uncovered that nearly two-thirds (64 percent) of the understudies who dropout of college do so for mental health reasons.[3] To address the challenges related with stress, people must be able to recognize the indications.

- The National Organized of Mental Health claimed that nearly one third (30 percent) of college students experienced significant misery that has an impact on their day by day exercises. Besides, students can befuddle push with the signs of discouragement. [8]

- Physical Indications of stress incorporate migraines, muscle hurts, acid reflux, indigestion, Chest pain and rapid heartbeat etc.

- Emotional Indicators are getting to be effortlessly unsettled, baffled, fickle, feeling repulsed, having trouble resting, low conceit, worthless, dodging others etc.

1.4 Problem Statement

Stress is defined as a reaction caused by a number of external circumstances, which can be viewed as a positive or detrimental experience. Fontana defines stress as an order tailored physically and mentally to adaptive abilities. When such skills can handle the request and take joy in the opportunity in question, then tension becomes a constructive driving factor and vice versa. In distinctive terms, feedback to stresses are likely to be distinctive by individual and the comparative environment can result in different results. Whereas a few individuals may succeed, others may shrivel beneath the strain. Cox Brockley commented that stress is an understanding event which comes about from a contrast between the request set and capability of somebody to achieve the mission productively. Unsteady circumstances in this implies will coordinate into stress encounters and in the long run into stress responses.
Within the final decade, there has been a developing consideration in exploring stress chance components and well-being results among university student's populace. Stress and mental health of university understudies could be a significant open health subject as sound students will be the more beneficial laborers of the longer term. Going to university has the potential to end up a positive and fulfilling involvement for students' life. In any case, it is observed that being a student may get an upsetting involvement.
As a result our study focused on the actual factors of university level students. We try to bring out the dependency of the stress level on those factors or reasons. In order to identify the factors we set two scales. One of them is Psycho geriatric Assessment Scales or PAS, but we modified it for a better result. It gives a whole view about a person with the help of psycho geriatric assessments. We have tried to figure out the relationship between the academic and personal stress factors. For more accuracy in our result, we have also used PSS or Perceived Stress Scale. It is the foremost in a big way utilized psychic utensil for measuring the judgement of stress. It can be an amount of the range to which scenes in one's life are called as stressful. Things were contemplated to feel how stray, wild, and over-burden litigants found out their lives. A number of coordinates are also added by this measuring scale and it can measure almost present levels of experienced stress as well .
Additionally, we have also performed a comparative study of stress levels between the 1st year and the last year students. The prompt point of this study is to explore

the sort and level of stress which a first-year student goes through and we will compare this to the stress which a final year student goes through . We point out how the factors of stress differ for a student from one starting stage to the final stage of their university lives.

1.5 Objective

In a number of ponders, creators found that the foremost habitually detailed variables contributing to stress and uneasiness around the examination periods were broad course loads, need of physical work out, and long length of exams, detailed by the students .One of the key reasons of human stress is the duty to act like pedantic in every steps for any level of students. Housework is exceptionally requested and the struggle for obtaining the best marks is exceptionally furious. Those pupils who are arranging to apply for higher studies from a great positioned university can be beneath an extraordinary bargain of weight as they battle to exceed expectations presently. The same is genuine for those who are looking for grant financing or who must keep their grades up in arranging to keep existing scholarship grants. Numerous students have involvement with monetary stress. This also adds the battle to manage sufficient cash for the payment of educational costs, as well as attaining the stores required to advocate their living costs whereas going to university. Students frequently lock in different activities of the university. In the expansion to taking a few classes at the same time, students may too legerdemain in employment, extracurricular activities, tender jobs, home duties, and many more. Whereas figuring handling these different synchronous subjections are great duties for adulthood and a number of students feel stressed to pull off all the work or duties at the same time. Whereas a bright perspective of the endurances is obtained by a few students and they need to consider themselves as adults, most feel vanquished while visualizing or attempting to find out what is the main purpose of their existence. The burden of choosing an explanatory career can have a huge impact on the rest of their days. The process of how a student chooses his major is quite unpleasant, as he or she can make choices approximately which connections to attempt right after one is done, where to stay and many more. The university is long characterized by a bit of alter. Managing with alter may be a major stressor for most people. Coming to the university is the start of a method for so many people to get to the ultimate power of independance and set themselves free as a bird . Taking off domestic and begins taking on extra duty can be exceptionally stretch actuating. A student of university level faces difficulties while making big choices and plans for the first time is something that is actually stressful for them. At the university level, peer pressure is something that is very much insane. Testing drugs, sexual action and other possible bad manners are frequently forced upon the sophomores and create pressure more often. The people who ignore or stand against these activities, are being pressured immensely sometimes it causes suicide as well. Already, there were a few works on student's stress level,like studies on the variety of stress stages between sexual orientations are ordinarily performed straightforwardly or in an indirect way. Thus, the deliberation of this examination is to approve the result of going before research that expressed female students suffer from extra stress than male students.The basic objectives of our research are-

- Our fundamental purpose is to anticipate the reliance of the variables of stress and analyzing the variables utilizing machine learning algorithms.
- Here we are centering on the PSS and modified PAS questions to degree the student stress level. We have working on our own datasat by evaluating the reaction of undergraduate students.
- We will be basically working on the dataset which is done on the premise of 1st year students' reactions and final year students' reactions. As a result the comparative study between their stress levels will be an easy going task for us.

1.6 Work plan

At first we set the arrangement of the work. Without a legitimate arrangement, it would not be done. We had to read a few research papers, we had to get what PSS and PAS ,how these scales work, set the questions for the students and at last the noteworthiness of students' stress. In our survey we have two groups, one is the 1st year university students another one is the last year university students.

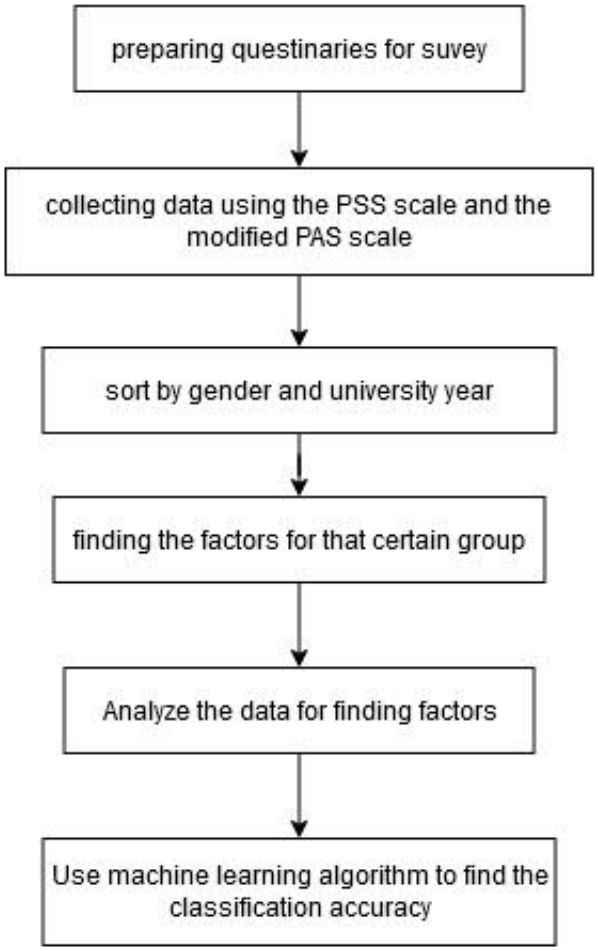


Figure 1.2: Workflow

Figure 1.2 shows the workflow graph of the total framework. The framework will take after the over specified steps to fulfill our reasons for searching the reliance of stress on it's components. To begin with, we'll make questions that nearly cover each segment when a student feels stressed . At that point we are going to orchestrate the questions agreeing to PSS and PAS scale. We are going to have the manual information from this and to perform feature extraction. The answers ought to be fair a tick mark on any of the numbers given close to the question. According to the previous description these each number includes a critical value. We'll provide appropriate instruction within the question paper to answer questions. The question paper will contain the significance of the numbers which students ought to select one as their answers concurring to their sentiments or involvement. The question paper will also contain the name of the participants, age, gender and semester number as well. We will analyze the collected data and sort them by gender and university year. Then we will extract features in order to find the factors for certain groups. Finding these factors will be done by machine learning algorithms. We will find dominating factors of a certain group. These algorithms will also show us the classification accuracy. From the extracted features we will also compare the 1st year students' stress vs the final year students' stress.

Chapter 2

Related Work

2.1 Literature Review

Stress is a natural feeling which people experience when they cannot cope up with any incident or pressure. From the viewpoint of a student, stress is a physiological and psychological condition in response to constant challenges inside a university environment[16]. If stress is extreme and constant, it can affect academic performance and disrupt a student's capabilities in many ways[9]. The Perceived Stress Scale (PSS) is a psychological tool most commonly used for evaluating stress perception. It is a measure of the scale for which circumstances are rated as stressful in one's life [10]. Many studies have been conducted to evaluate academic stress level of students and its effect on their life based on perceived stress scale (PSS).

In [9] how academic stress can affect the progress of students and the influence of factors such as level of education, gender and age was examined. Students are vulnerable to experiencing stress in an advanced educational institution like University where any workload put on a student always has a time limit time and pressure to do better than others. Researchers have previously identified signs of stress such as loss of energy, high bp, depression, elevated cravings, concentrating trouble, impulsiveness, anxiety etc. They studied 150 students selected from different universities of Islamabad. Their age was from 18 t0 20 years. For measuring stress among students, they used perceived stress scale (PSS). There are 10 items in that scale and reliability is 0.74. Scoring and analyzing were performed after data had been collected. Five hypotheses have been established and tested to examine how academic stress can affect participant's performance. First hypothesis was academic stress affects student success. They observed that the greater the stress, the poorer the student's academic achievement. Students who can efficiently deal with their stress do so much better than others who can't manage or deal with their stress. The second presumption was academic stress affect female students more than male students. The findings showed there were no substantial difference in stress level of male and female students. In fact, female students knew all about how they should manage time and handle stress such as male students. They were also very committed to their studies, attentive and reliable. Third presumption was younger student's academic stress would be higher than older students. Younger students had to face more transition challenges and older students showed improved organi-

zational skills than younger students as with time they had learned how they should manage their time. Number four theory was in less experienced students (juniors) the academic stress would be greater than in highly experienced students (seniors). Juniors had low degree of maturity and were more vulnerable to university academic stress compared to seniors who really were mature, responsible, and very well suited to academic stress. Finally, the last hypothesis was students would feel more stressed around final than at the start of the semester. Study showed that there was no significant difference in this scenario. If it was the beginning or the end of the semester, students needed to stay vigilant the whole semester.

The objectives of this study[11] were to identify academic stress in the students 'different categories and Explore the connection between student's characteristics and their perceived stress. Survey took place in five universities of Sweden. There were 187 participants (131 female and 56 male) from six different classes of the social and nursing sciences of those universities. Approximately two-thirds of the students were under 24. They used modified version of the Perceived Stress Scale (PSS-14). In this analysis, the modified version of PSS had ten items to evaluate the degree to which circumstances in the life of students were perceived as unexpected, unmanageable and overburdening. The data were graded numerically and quantified. Using SPSS version 22.0 these quantitative scores were entered in a program for review. They performed descriptive and inferential statistical analysis. Fifty percent of participants recorded high stress levels in the range of 8 to 15 and only 12 percent (n=23) of students recorded high stress rates in ¿= 20. Most participants in this study documented physical stress in the form of headache symptoms, and psychological stress in the form of depression and anxiety symptoms. Students with high stress rates found seen notably more amongst women than men (16.0% vs 3.6%). High stress rates amongst students with overweight or obesity were dramatically higher than for students with healthy weight (26.3% vs. 7.9%). High stress rates among nursing science students were notably more present compared to social science students (19.4% vs. 8.3%). Students who exercise frequently or very regularly high stress was less among them than students with lower physical engagement (4.7% vs. 31%). Students with high stress rates were substantially more in the student community who reported worse state of health than students who self-identified themselves as healthy.

This is a research [12] carried out on medical students. The medical school is known as a stressful place that can adversely affect the academic success, health and wellbeing of students. Many researches have shown that medical students experience more stress. Students use different coping mechanisms to relieve tension during medical school. There can be a number of causes of stress, in medical college some factors have been established as academic, personal, etc. Understanding the stress rates of medical students of depth will help to avoid potential mental ill-health such as depression. The study was done in Mysore Medical College and Research Institute. It is the only government run medical college in Mysore. The research population was composed of medical undergraduates. The research was performed from July to September 2017. The perceived stress scale (PSS-10) was used for measuring

stress levels. The questions in this scale query was about attendee's emotion and experiences over the previous month. This was a survey of 10 items with answers in the form of how frequently they have undergone those events, from 0-never to 4-very frequently. They used Medical student stressor questionnaire (MSSQ) evaluate associated stressors. The MSSQ was composed of 40 objects that display the six stressor regions. The queries represented any circumstance or behavior and the participants were asked to answer the stress level they would feel if they are involved in that specific activity in the scale of 0 to 4 (0- no stress to 4-very highly stressful). A total of 303 students (156 females and 147 males) took part in the study. The participants age groups ranged from 18 to 24. Stress increased higher among students over 20 years old (84.6 %). There was less stress between students whose parents were doctors (65.4%) compared to other students (81.2%). Stress among 1st and 4th year students were greater (84% and 87.3% respectively). Again, the majority of students (40.9%) reported that academic related stressors were the cause of high stress accompanied by interpersonal and intrapersonal based stressors 33.3%. in the study most participants (70%) encountered moderate stress in and only 6% suffered extreme stress.

In this [13] study they design a scale for measuring perceived sources of academic stress amongst students at universities.it took place at Tanta University Egypt. For the study 100 students (75 male 25 female) were selected from the third year, Educational Psychology class. Their age was from 19 to 26 years old. They were going to take their third-year final course exams. It was a written (three hours) examination (75 marks) and an oral examination (25 marks) immediately after that. Students have earned no points for any assignments or research they have completed during the year. Both male and female experts (12 experts) were invited to contribute in this report and they were faculties of Psychology and Educational Psychology at Tanta University. The experts offered their opinion on the instrument's overall quality. Every expert checked and commented on the relevance of the scale to be formed prior before testing it with participants. The created PAS had 18 items which were 5 points liker type questions. The participants needed around five minutes to complete the PAS. On the basis of Kaiser rule the percentage of variation was responsible for and the cohesiveness of the variables a four-factor approach seemed ideal. The factor were 'pressure to perform' (5 items) , 'perception of workload and examinations'(5 items) , 'self-perception' (five items) , 'time restraints'(6 items).the study reflects 18% of the variance, the academic stress experience, the huge expectations of faculties and parents and the social pressure to be successful. Participants who performed the PAS reported low stress levels for most items delivered, given the fact that the scale was conducted at the time of the final which is known to be a significant source of stress. Medium to extreme sources of stress were correlated with instructor's criticism on academic results of students, and were correlated with intense peer rivalry. Outcomes from this study support the fact that there were positively recorded scores indicating that students were substantially optimistic of their educational outcomes and potential career prospects and were confident of academic strategic thinking.

For [14] study experiment was conducted in the INZUA algorithm programming

contest summer camp. It is held annually in Istanbul, Turkey. This algorithm programming camp is planned to develop programming skills for high school and undergraduate students. 84 participants with different experience level took part in this contest. 21 students (18male 3 female) were selected to collect data were (physiological signal and questionnaire perceived). Participants were on average 20 years old. There were three types of session- training or lecture session, the contest and the day off. It took place for 9 days from 10.00 to 17.00. At first attendees had two hours classes with faculties from top ranked universities of Turkey. Then students participated a problem-solving contest where the questions were from that day's class. During all sessions, the users received the raw NASA-TLX questionnaires. The questions were used to evaluate the perceived stress level of the participants. They used wearable devices- two Samsung Gear S1, ten Samsung Gear S2, four Samsung Gear S3 smartwatches and four Empatica E4 wristbands. It had been shown that device's data quality improves the accuracy in the classification of stress levels. For their collected data two best performing classifier were the random forest and the multiplayer perceptron algorithms, these classifiers outperformed other algorithms. They also observed the pre received stress and physiological stress of participants could not be the same. Contests were intended to cause most stress, the lectures were expected to offer some intellectual load and lesser stress and free time sessions were believed to be comfortable and relaxing. They were asked '' How irritated, stressed and annoyed versus content, relaxed and complacent did you feel during the task". The response was no a measure of 0 to 100 with increments of five points. They found that the perceived classification of stress levels results in less accuracy than the classification accuracy of the physiological stress level for all ML algorithms. The relation between the known context and perceived stress labels was calculated 0.356. It is a decent correlation. There was no analysis onto the relationship between perceived stress and physiological stress in their paper.

The goal of the research [15] was to examine the amount of perceived academic stress amongst university students in Nigeria. The research sample consisted of 427 (228 males ,199 females) randomly picked undergraduate students from three professors in a University of Nigeria. 128 students were aged 16 to 19 years and 229 were aged 20 years or above. 142 participants were from the professor of Education, 159 were from professor of Physical Science and 126 were from the professor of Social Science. They used a 20 items instrument. Provided questions were in two parts. Part A sought demographic data of participants about age, sex, faculty and year of study and Part B was to evaluate academic stress scale by 20-items. In their lecture theaters, the instrument was provided to the participants just before the start of the second semester exams and the completed reports were collected. They developed four null hypotheses to lead the study. First one was about finding no noticeable difference between man and woman's stress level. But there wasn't any major difference. To compare the mean scores of male and female they used independent sample t-test. The findings showed that the mean difference was 1.6577 which was important at a significance point of 0.05. So, the hypothesis was not accepted and male students reported considerably greater academic stress than their female peers. The second hypothesis was about finding no noticeable difference in stress level between younger and older university students. It was also tested using dependent sample t-test. The mean difference was 0.1885(for younger 49.7188 and for older

49.5302) which could be ignored at 0.05 level of significance. So, this theory was accepted. Third hypothesis was about not finding any major different in stress level of three different professors. The mean score of students of Physical science, Education and social science were respectively 51.6962, 49.0211, 47.5974. This theory was tested using one- way analysis of variance statistics. Noticeable differences were found among their mean values at 0.05 level of significance. So, third theory got rejected. The fourth and last hypothesis was about finding no noticeable difference in perceived academic stress level among participants of different education levels. The mean score of students of 400,300,200and100 were respectively 50.8333, 49.761, 49.761 and 48.3571. Again, this hypothesis was tested using one- way analysis of variance statistics at 0.05 level of significance. Last year undergraduates reported to experience noticeably higher rate of academic stress than first year undergraduates. No major difference was found between the degree of academic stress faced by the 200 and 300 level undergraduates.

The research [16] explored the prevalence of stress and coping strategies of undergraduates. The research also investigates if there are any major variations in the coping mechanisms of male and female students. 334 students (222 females,112 male) from university of technology in southern Gauteng, South Africa participated in this research. They were 18 to 20 years old. The provided questions were segmented into three parts. The questionnaire of part A was about biographical information, part B was the Perceived Stress Scale-10 (PSS-10) and part C collected knowledge on ways of dealing stress using the Brief COPE scale. They used descriptive analysis to evaluate collected data and used Exploratory factor analysis to determine the dimensions of stress and stress dealing. Research had shown that students with stress issue faced problem to overcome difficulties(M=3.21), not being able to manage important things of their lives(M=3.26), got disturbed by events outside their grasp (M=3.29), quickly became irritated by sudden incidents(M=3.41), could not deal with irritations in general (M=3.53) and quickly got nervous (M=3.74). Study also shown that students practice various types of coping techniques such as introspection and veneration (M=3.83), emotional support (M=3.30), humor (M=2.95), negativism and refusal (M=2.60) and substance abuse (M=1.80). It had shown no major variations between males and females in negativism and denial factor (males median=2.41, females median=2.57), emotional support factor (males median=3.25, females (median=3.25), and humor factor (males median=3.00, females median=3.00). There were major differences between males and females in the introspection and veneration factor (males median=3.80, females median=4.00), substance abuse factor (males median=2.00, females median=1.00).

Chapter 3

Proposed Method

3.1 System Model

Figure 3.1 shows the methodology of our research. Firstly, we collected data from students of BRAC University by conducting the survey. we will use the important factors from the survey reports and drop the unnecessary factors. The clean process will be done by algorithms like feature extraction algorithms and finding out the relation between stress level and the factors of being stressed. So, we can easily identify the vital factors of stress level in university students. Now, For processing the data Machine Learning algorithms will be used (SVM, Random forest etc).Finally, using these algorithms we will get the result of very important features of stress level.

Figure 3.2 shows how the machine learning model work for this research. based on all results, we will choose the best from different datasets based on different stress survey.

3.2 Dataset description

3.2.1 Questionnaire Description

To get our dataset we have to do a survey. For that survey we need some questionnaires. As we are working on university students' stress detection, we took two stress scale.One is PSS(Perceived Stress Scale) and another one is PAS(Perception of academic stress). As we focused on university students' stress levels during examinations time, we modified our PAS Scale question and used some of our own questionnaires related to our work.So, in our survey we provide textual questionnaires to the students and they gave us answer based on their perspective.these answers help us to get our dataset and complete our work.

PSS(Perceived Stress Scale): PSS is one of the known and broadly utilized Scale for measuring Stress.This scale is intended to test the stress level of one's life[10]. There are 10 questions on this scale which are about feelings and thoughts during the last month. The questions are designed in such a way so that one can understand and answer these questions easily[10]. Also. There are 5 scales to answer these questions which are 0,1,2,3,4. Here,0 indicates Never, 1 indicates Almost

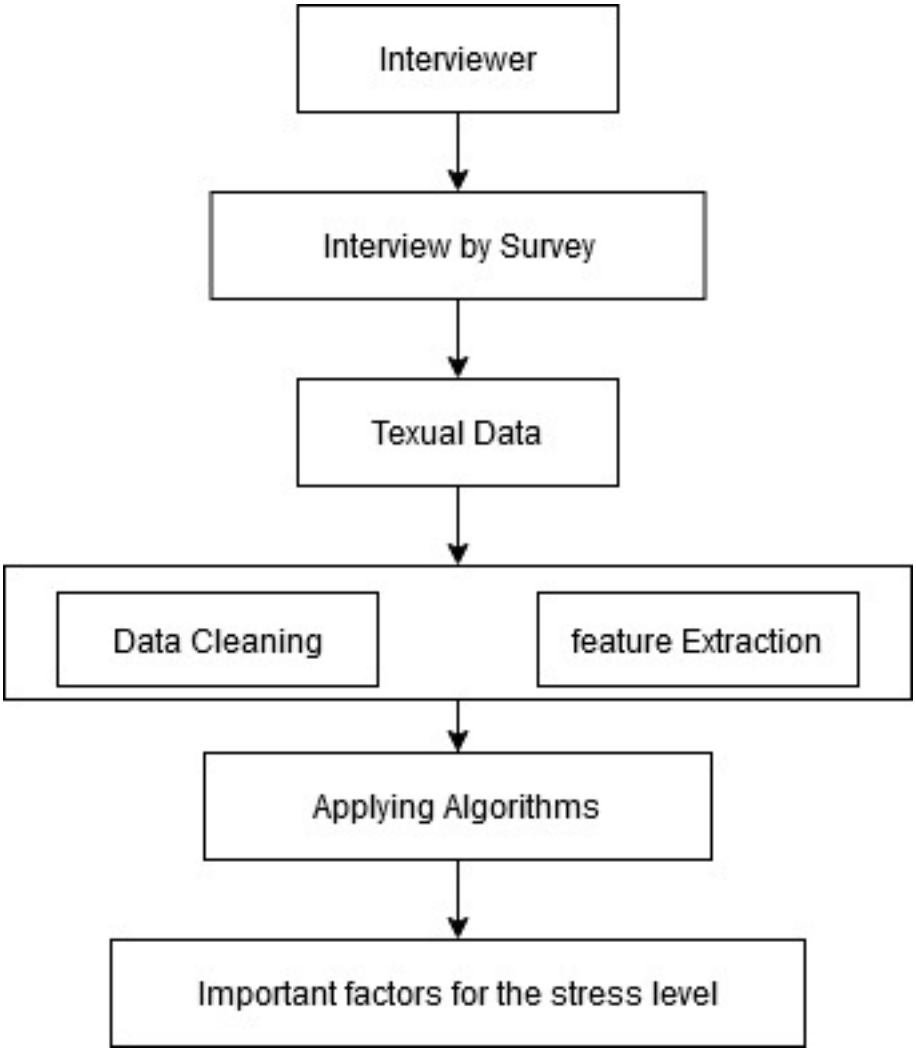


Figure 3.1: System Model of Finding Factors

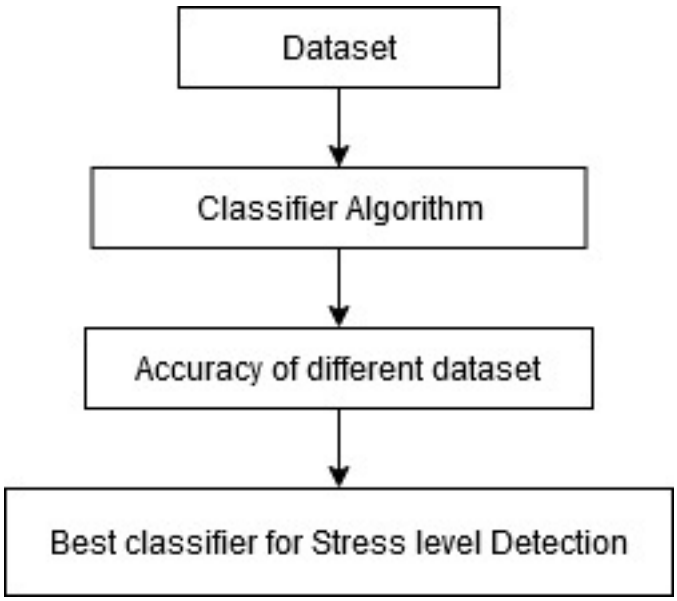


Figure 3.2: Flowchart of the Research

Never, 2 indicates Sometimes, 3 indicates Fairly Often, 4 indicates Very Often. In these questions, there are some positive stated questions also which are number 4,5,7 8. The PSS scores obtained reversely for this question. For example 0=4, 1=3, 2=2, 3=1, 4=0 [10].In this survey, all the questions are based on their last month expectations, feelings, nervousness, problems etc. If someone agrees with that question she or he might answer based on the five given scale.

Result: Based on these above question during Perceived Stress Scale or PSS scale the stress is Low when the Total Value of PSS scale is between 0-13.It is moderate when the Total Value of PSS scale is between 14-26 and it is High when the Total Value of PSS scale is between 27-40.

Stress	PSS(Perceived Stress Scale)
Low	0 - 13
Moderate	14 - 26
High	27 - 40

Table 3.1: Calculation of PSS Scale

Modified PAS(Perception of academic stress): As we are working on, stress detection of University students,so we took help from PAS scale.Here all the questions are based on their last month expectations,feelings, nervousness,problems etc. If someone agrees with that question she or he might answer based on the five given scale[11]. That is why we choose the PAS scale for our work but we did not take all the questions from the PAS scale. As we mainly focused on University students stress detection during examination time so we prepare our question based on that.There are 15 questions in that Scale where the question have 5 levels to answer these questions which are 0,1,2,3,4. Here,0 indicates Never, 1 indicates Almost Never, 2 indicates Sometimes, 3 indicates Fairly Often, 4 indicates Very Often. In this scale, there are some positive stated questions also which are 12 24 numbers. The PAS scores obtained reversely for this question.For example 0 =4,1=3,2=2,3=1,4=0.
0=Never;1=Almost Never;2=Sometimes;3=Fairly Often;4=Very Often

Script Checking :

1. I felt nervous while checking my scripts.
2. During script checking, the mark met my expectations.
3. The competition with my peers for marks is quite intense.

Here these questions are based on Script checking.In 1 Number questions it is asked whether the student felt nervous or not during his or her script checking time. Then, In the 2 number question it asked when the student got the mark whether that mark can meet the students expectation or not. After that, In the 3 number question it was asked whether the competition with their classmates or peers for marks is dreadful or not.

Examination:

1. The examination time is short to complete the answers.

2. The examination questions were difficult.

3. During the Examination, I became very stressed.

4. I am worried whether I can meet my expected result.

From question 1 to 4, we have talked about the quiz and examination. We make these questions based on examination.In the 1st question we have asked whether students can complete their answer during the examination time or not. Then in 2nd question, It has asked whether the question was difficult or not.By doing that, we wanted do know their opinion about examination time.After that, In the 3rd question, the students has asked during examination time whether they felt stressed or not.Then again,In 4th question, they have asked what they think about their examination, whether they will get their expected result or not.

Understanding Lectures or Videos

1. The way of explaining the lecture is not easy for me to understand.

2. The actual content or topic of the lecture is difficult.

3. I am not prepared enough for the topic.

4. My background is not related to the topic or inappropriate.

In this section, we have shown a flipped video to the students to find out their stress level. After showing them the video we have asked them these 4 questions.In the 1st question we have asked whether they understand the explanation of that video or not.Then, in the 2nd question, the question said whether the lecture was difficult or easy.Then again,In 3rd question the question has asked whether the student have prepare enough for the topic or not .Finally in 4th question it has asked whether the students background is related to the topic or not.

Future Career Or Job:

1. My experience is not enough to get a job.

2. My result is not up to mark to get a job.

3. My family pressurizes me for my future career.

4. future career/job the job market is too competitive for me.

These questions are asked to the students based on their future career expectation to get a job. In the 1st question, it has asked whether their experience is enough for getting a job or not.Then, In the 2nd question,it has asked what they think about their result. Whether their result is enough to get a job or not. After that, In the 3rd question,the question asked whether their family pressurized them for their future career or not.FInally,In the 4th question, the question was asked,what do they think about the job market?whether the job market is competitive for them or not.

Result: Based on these above questions during Perception of academic stress or PAS scale the stress is Low when the Total Value of PAS scale is between 0-20. It is moderate when the Total Value of PAS scale is between 21-39 and it is High when the Total Value of PAS scale is between 40-60.

Stress	Modified PAS(Perceived Academic Stress Scale)
Low	0 - 20
Moderate	21 - 39
High	40 - 60

Table 3.2: Calculation of Modified PAS Scale

Survey Dataset

For the dataset we have divided our questions into two part.One is PSS scale where to know the primary stress level of a student and another one is PAS scale which is modified based on examination, flipped videos, script checking and future career to know about their stress level of academic life.There is total 25 questions to collect our data. We have asked students about these 25 questions by giving them the environment related with the questions and collected our data from the answer of these questions.

Basic Information Of the Students: The basic information we need for our research work is gender and university year. We take a survey from 132 University Students. They are from different genders,different ages, different years,they have different mediums to answer those questions also.Everyone has different answers or opinions for those questions.So, everything has to be analyzed to categorize them. These basic information helps us to separate them according to groups

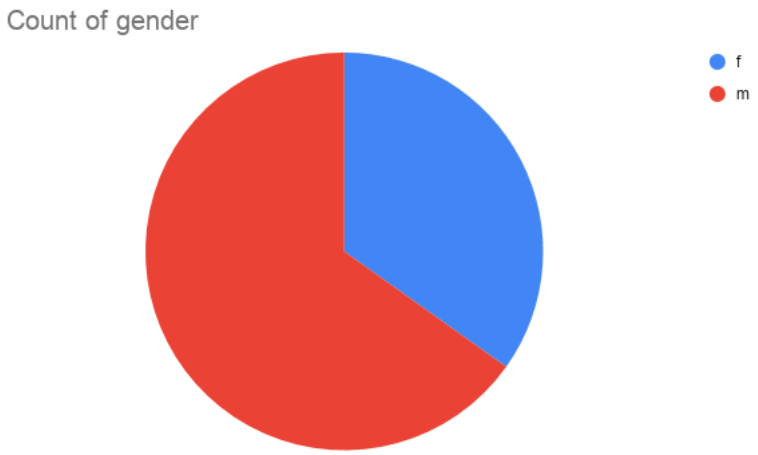


Figure 3.3: The ratio of Gender in the dataset

Figure 4.1, 34.8% female students participated in our survey where male participants were 65.2%.

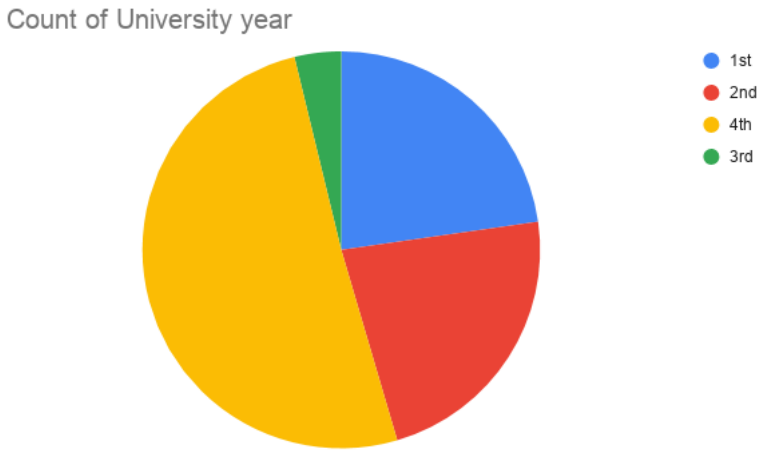


Figure 3.4: The ratio of University Year in the dataset

Figure 4.2, 22.7% students are from 1st year.Again similar percentage means 22.7% from 2nd year also.Then,50.8% students from 3rd year from 4th year there are 3.8% Students.
In the above we are seeing that the participants of different classes who participate in the survey. Different ages of students from different years also from different genders participate in the survey. We also notice that the number of males are a bit more than the number of female participants.

3.2.2 Feature Selection

As we are considering the questionnaires as the features so we have 10 features from the PSS scale and another 15 features from the PAS scale along with the university year and gender.

We will find out the important features using the Chi square algorithm. This method is used to select the important factors of each scale. We will find out the most important features for finding the best classification algorithms to get the most accuracy out of it. As Not all the features are equally relevant for the research.

3.2.3 Methodology

The data used in our research is collected by a survey. The survey data is collected by printed form. Then we made an excel file for analyzing the data. All the data was sorted in an excel file. It helps to represent the statistical data. We also used the Pie chart or Bar diagram for statistical assessment. For the survey, we created questionnaires. These questions are related to the exam, understanding class lecture and future or career-related stress level.
After collecting the data, we select some specific features for predicting the highest rate of accuracy. So, among that many features, we work on some specific features of our dataset. It is the part of our pre-processing of our dataset before applying classifier algorithms. Before the pre-processing of data, we cleaned the data. Some data are unnecessary as they are not relevant to our research.

Firstly, we conduct a survey for data collection among the students of different courses. Then we split the data. Here the pre-processing stage started. We cleaned the data and removed the unnecessary data. Then converts all the string type data to numerical data.
After splitting the data, we used Cronbach's alpha and Shapiro-Wilk test for questionnaire analysis. It also determines the reliability of the survey dataset. In the time of analyzing the dataset, we found every feature is not that much important to determine our factors. So, we use Chi-square methods for feature extraction. So that we can easily identify which features are appropriate for our research. These are the steps of our data processing.
Moreover, We used the Chi-Square method to analyze the important features of the dataset. As we are predicting the stress level of university students, the Chi-square method uses the features vs the training set. When we train the features of the PSS scale, the same goes for the features of the modified PAS scale also.
The next phase was training and labeling data. This is useful for making the dataset more accurate. It will help to gain a good accuracy score. Better accuracy indicates better performance of the dataset. Finally, to get accuracy and result we use five classification algorithms. These are Logistic Regression, Decision Tree, SVM, Naive Bayes and Random Forest. Each of the classifiers shows different accuracy scores and confusion matrix.

Figure 3.5 is the working methodology of our research. Here we described the algorithms we used for questionnaire analysis and feature extraction for the research. We also discussed the classification algorithms in the diagram.

3.2.4 Dataset Visualization

Heatmap: The heat map is a visualization of two-dimensional data on which color values are displayed. The graphic heat map produces an immediate visual summary of the data. More accurate heat maps allow viewers to analyze complex data sets. Perhaps there is a number of techniques to visualize heat maps, and they all have one common trait. They use shade to convey data-value relationships that could be more complicated to understand if they were seen numerically in a graph[27].

The figure 3.6 shows that Q2 is strongly correlated with the PSS stress level and Q4 is the least one. In the following figure 3.7, Q16 feature is highly correlated with the modified PAS stress level and Q3 is the least one to be correlated.

QQ PLOT: This is a quantile-quantile plot, a statistical method that lets us determine if a data set is obtained by a valid mathematical distribution, such as regular or exponential. For instance, we used a common Q-Q plot to check the hypothesis if we perform a statistical analysis predicting that our dependent variable is normally distributed. It's just a visual assessment, not an exact evidence, so it's sort of arbitrary. So it helps us see how true the hypothesis is, and if not, how faulty the argument is, and what data points lead to the breach. A Q-Q plot is a scatter plot formed by mapping two sets of quantiles against each other. Since all sets of quantiles derive from the same array, the dots can be seen forming an almost straight line. It is the traditional Q-Q plot instance where all quantile sets come

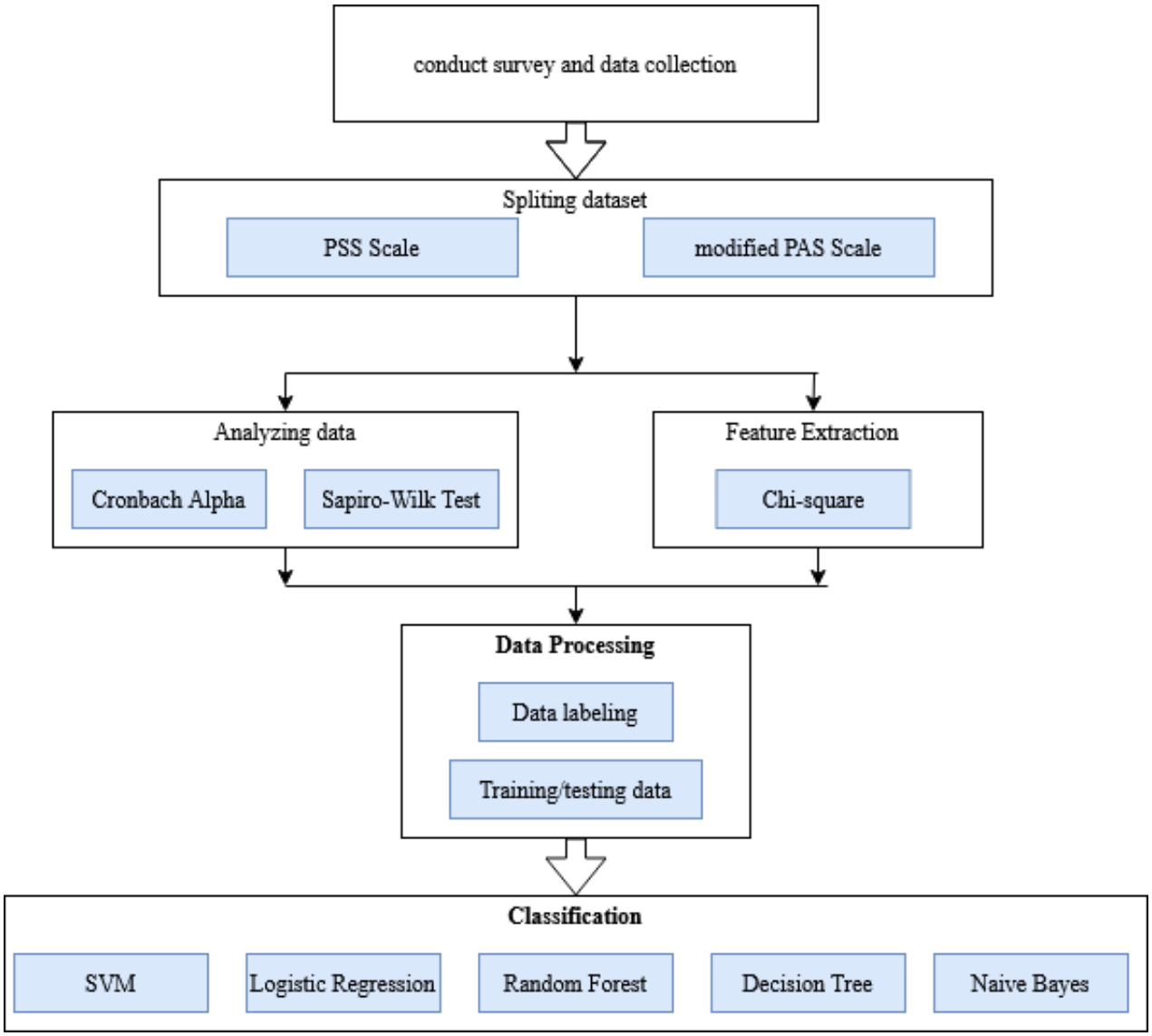


Figure 3.5: Working Methodology

from normal distributions.
Therefore, quantiles are essentially all ascending ordered data, with different data points being marked as the point below which a certain percentage of the information drops. Nevertheless, there are many methods to measure quantiles. Q-Q plots are taken from the sample data, ordered in ascending order, and plotted towards the hypothetical distribution of determined quantiles. The quantity of quantiles is selected to suit the value of the sample data. While generic Q-Q plots are the most widely used in practice due to too many statistical methods that assume normality, Q-Q plots can be generated for any distribution[28].

Prediction from the preceding graph in Figure 3.8 shows that the PSS scale is usually distributed. In addition, this following graph (Figure 3.9) indicates that the adjusted PAS scale is also usually distributed.

Bar Graph: A bar graph is a rectangular bar graph with lengths and heights equal to the values they represent. They reflect the data vertically graphically. This in-

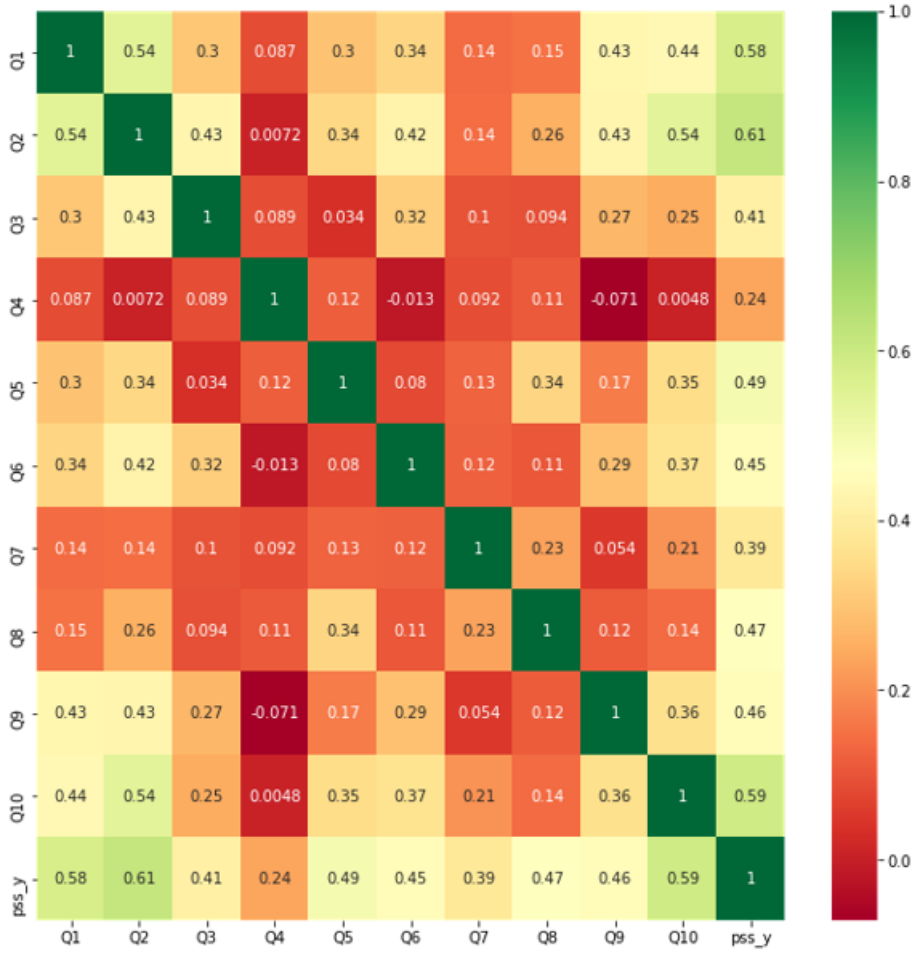


Figure 3.6: Heatmap correlation matrix for PSS scale

dicates the types of data being measured on one side of the graph. The other axis reflects the values of each type of results. The length of each bar is equivalent to the size of the type of results, and all bars extend from bottom top [29]

Figure 3.11, 3.12, 3.13 shows that for all levels of PSS Stress level the moderate MODERATE PAS stress level is seen most of the time.

The following bar graph from figure 3.14, 3.15, 3.15, 3.16, the 1st and last year students have the mostly the MODERATE stress level for both PSS and modified PAS scale

3.2.5 Analysis of Dataset

We have divided our dataset in two parts. First we have taken the values as input for a 0 - 4 scale as discussed before and tested that for different types of classification algorithms based on the stress level (low, moderate, high). All the dataset was set in that way to find the classification better and to see whether we could classify stress levels based on these two datasets or not. After forming out dataset we test the dataset for normality using the following machine learning algorithms and feature selection techniques

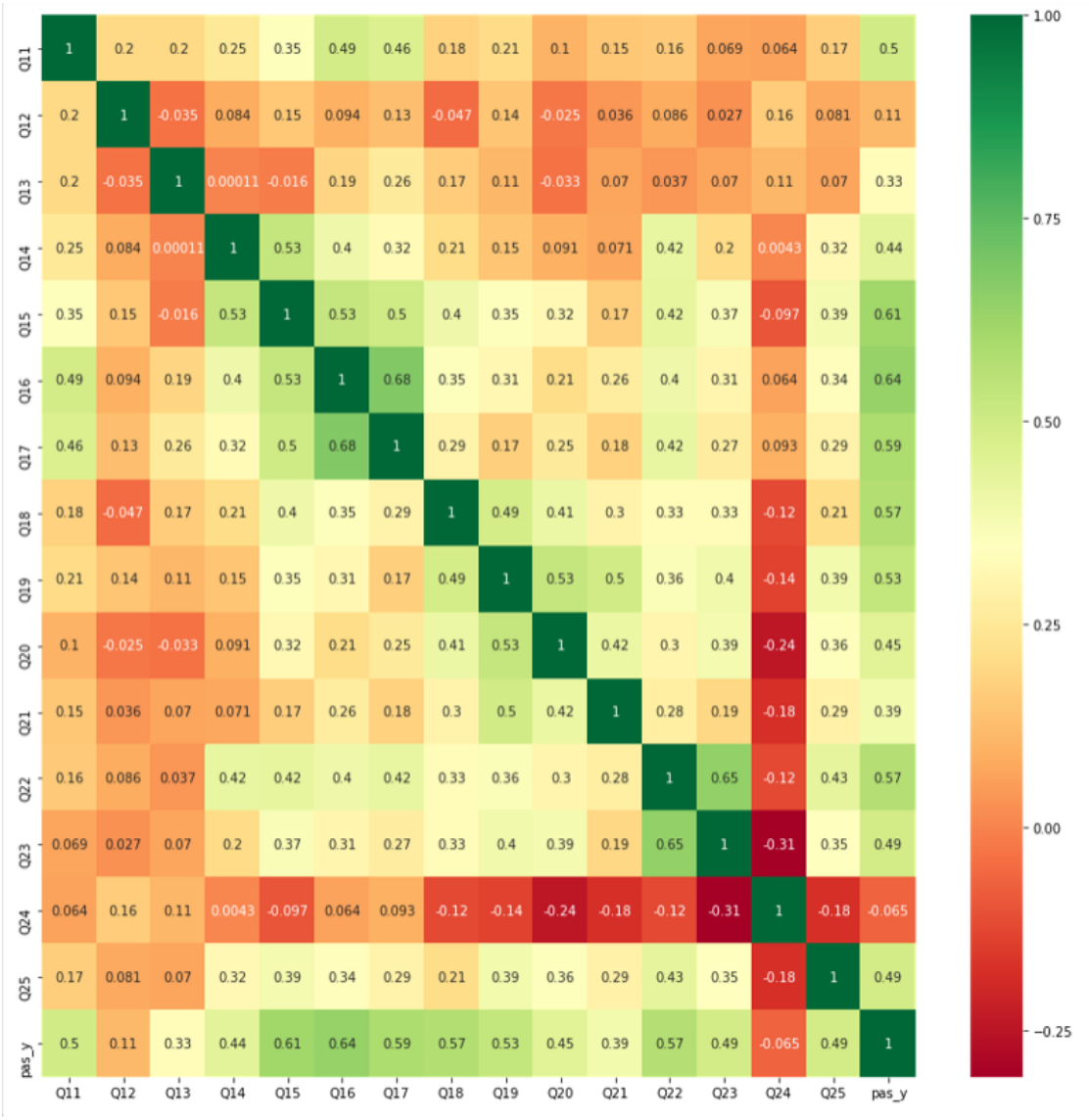


Figure 3.7: Heatmap correlation matrix for modified PAS scale

Shapiro-Wilk Test

This test is designed to evaluate whether a collection of data is from a normal distribution. We have tested our dataset to see the statistics and P value and also whether they belong to a normal distribution. The Shapiro Wilk test was carried out on all the different models in our datasets but mostly ended up giving us different responses. The statistics and p-value of these variations show the normality of our dataset. The following table 3.3 shows the values of this test for various datasets.

Dataset	Statistic	p-value	Normality
PSS Scale	0.741	0.000	non-parametric
Modified PAS Scale	0.748	0.000	non-parametric

Table 3.3: Result from Shapiro-Wilk Test

In the above table we see the statistics value which is one of the indicators of the

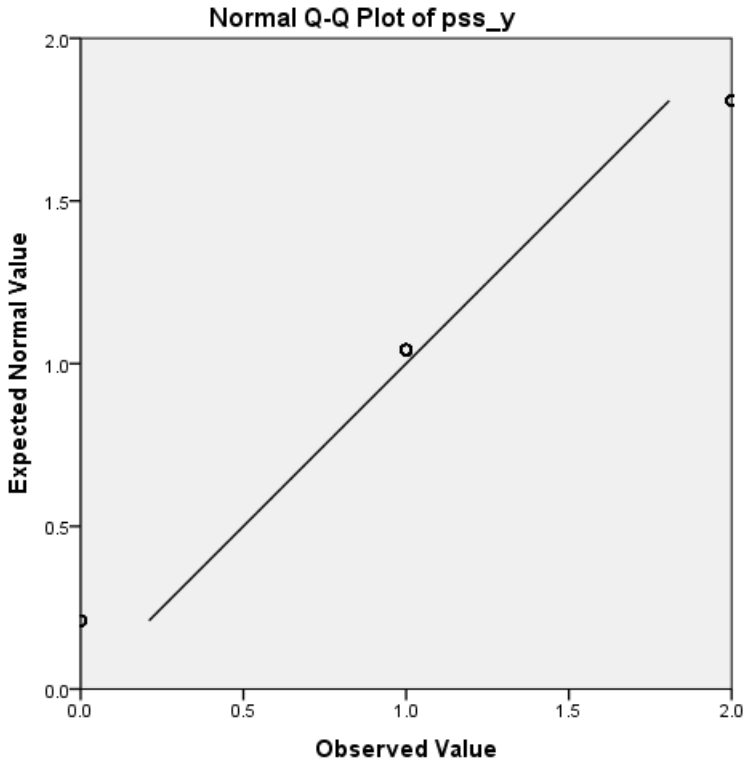


Figure 3.8: PSS scale's QQ plot

Gaussian distribution shows the values at a range of 0.741 and 0.748. When all the values are generalized the values drops due to less numbers of scale. P value is actually the easier measure of dataset normality. Even though the datasets show different statistics value the p value is the same for all of them. Our reference p-value was 0.05 as seen above the p value our datasets show that our value is 0.000 which is lower than the set threshold. Thus rejecting the null hypothesis showing that the dataset does not belong to a gaussian distribution.

Cronbach's Alpha Test

For calculating the internal consistency of a set of data the cronbach's alpha method is used, to see how closely the features are related as a set. It shows the reliability of different dimensions. We measure the reliability of the features of our dataset by the cronbach's value. The higher the value, the more reliable that feature is. We have run Cronbach's Alpha algorithm based on the two different classifiers and it shows us different features which are important for that dataset. In the different datasets different features come up with reliability, which is explained in the table 3.4 and 3.5 . Comparing the alpha values we found that both survey datasets conclusion draw that the features of the both datasets are reliable and quite good as all features' alpha value is greater than 0.5. Moreover the overall result shows that both datasets' alpha values are greater than 0.7 and very close to 0.8 which indicates that they are very reliable.

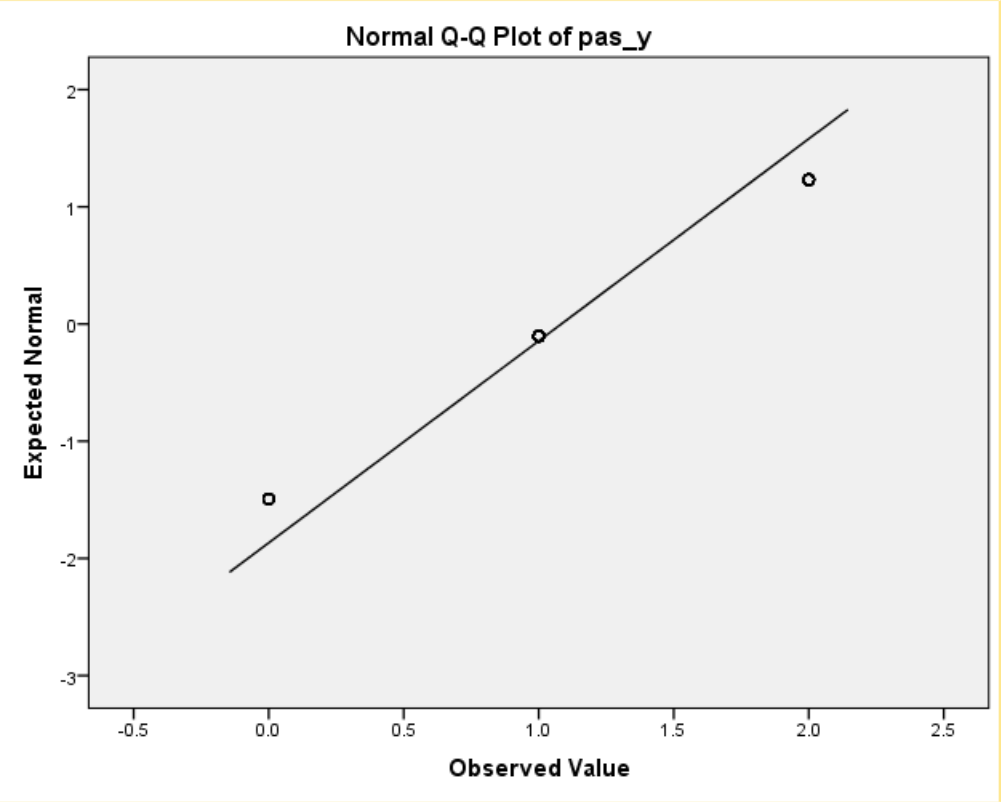


Figure 3.9: Modified PAS scale's QQ Plot

Dataset	Cronbach's Alpha	Consistency
PSS Scale	0.746	Acceptable
Modified PAS Scale	0.798	Acceptable

Table 3.4: Overall Result from Cronbach's Alpha Test

Dataset	$\alpha \geq 0.8$(Good)	$\alpha \geq 0.7$(acceptable)	$\alpha \geq 0.6$(questionable)
PSS Scale(10)	0	8	2
Modified PAS Scale(15)	2	13	0

Table 3.5: Feature based Result from Cronbach's Alpha Test

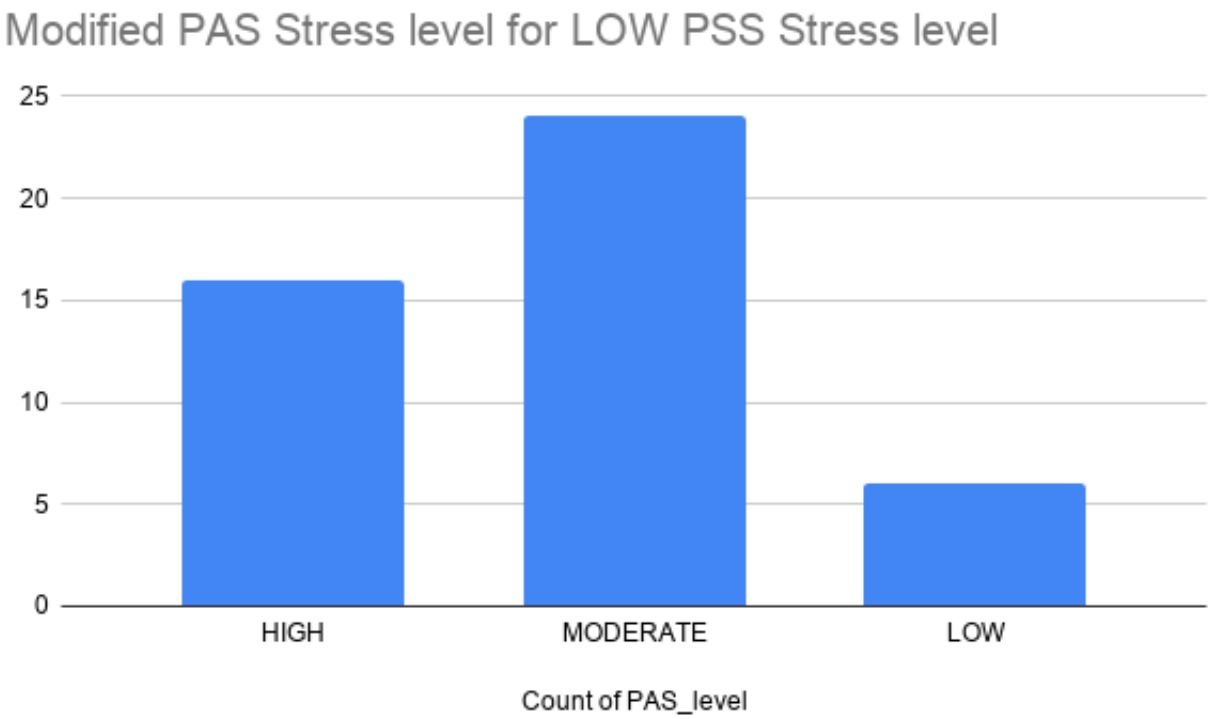


Figure 3.10: modified PAS stress level for LOW PSS stress level

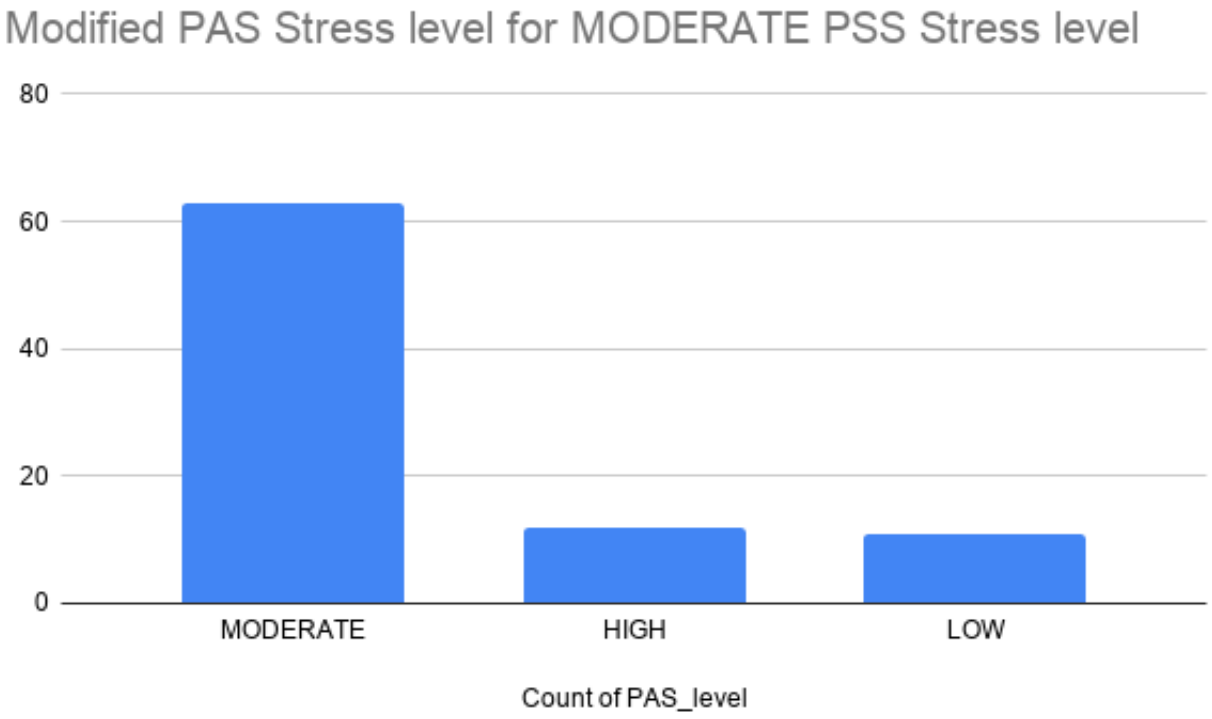


Figure 3.11: modified PAS stress level for MODERATE PSS stress level

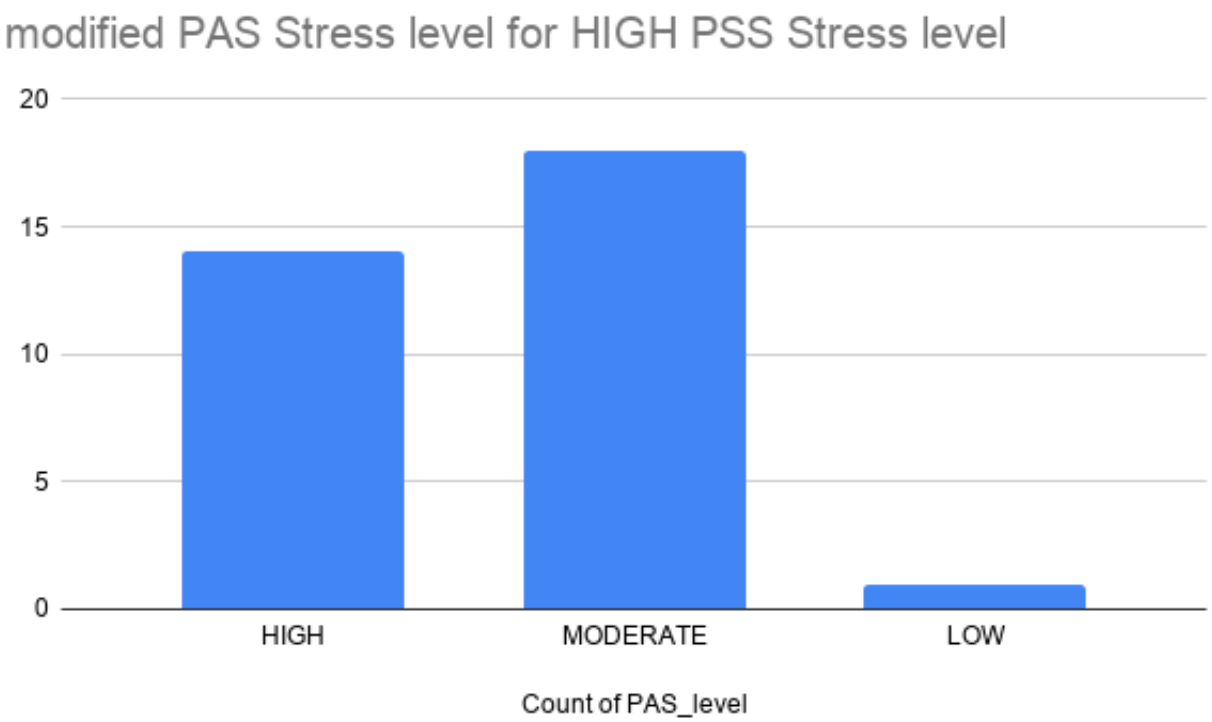


Figure 3.12: modified PAS stress level for HIGH PSS stress level

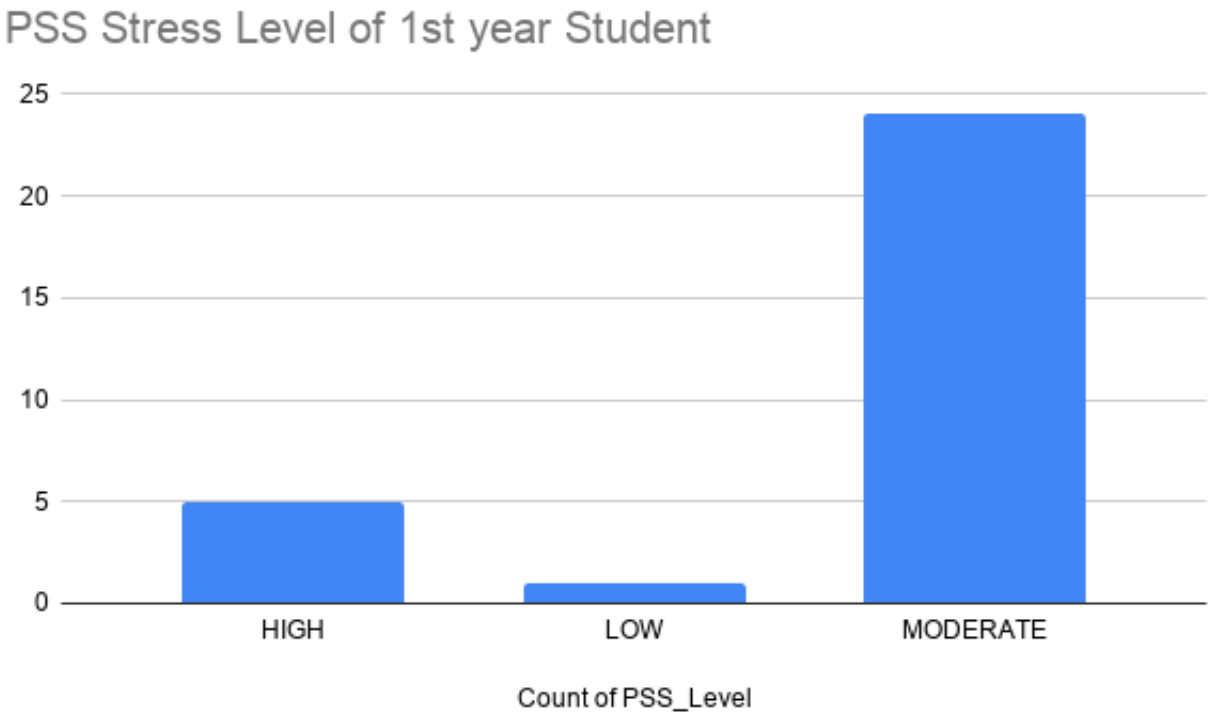


Figure 3.13: PSS Stress level of 1st year's students

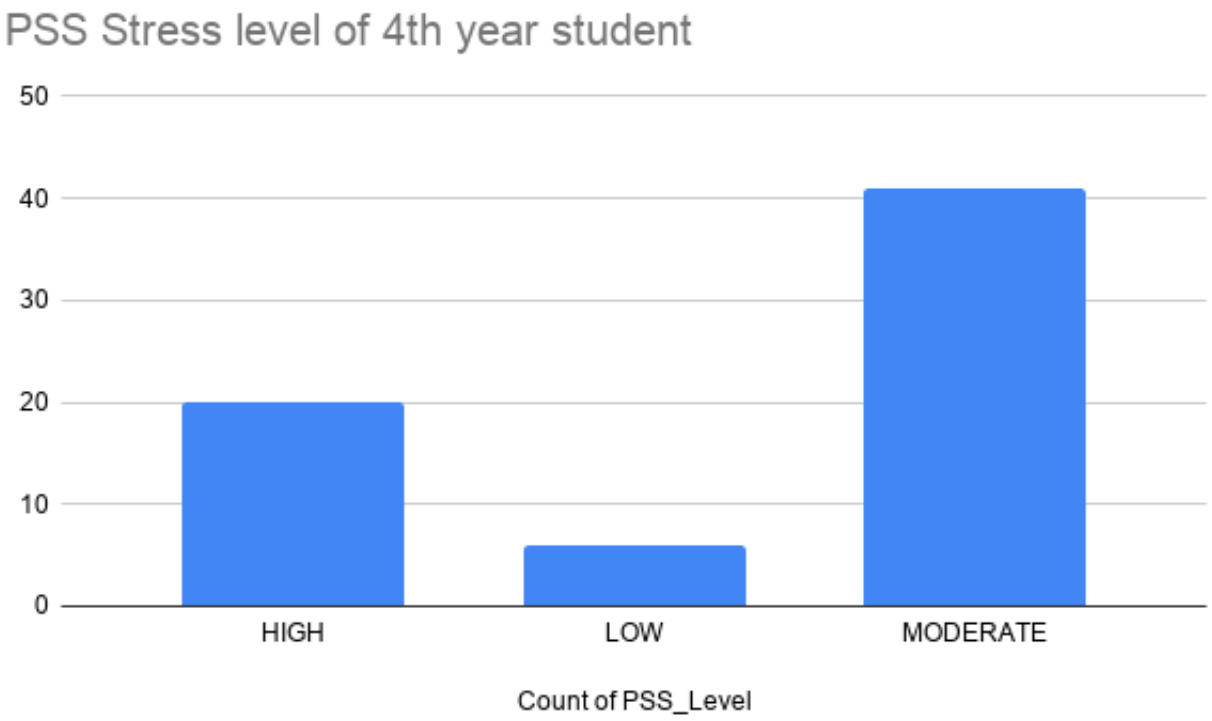


Figure 3.14: PSS Stress level of 4th year's students

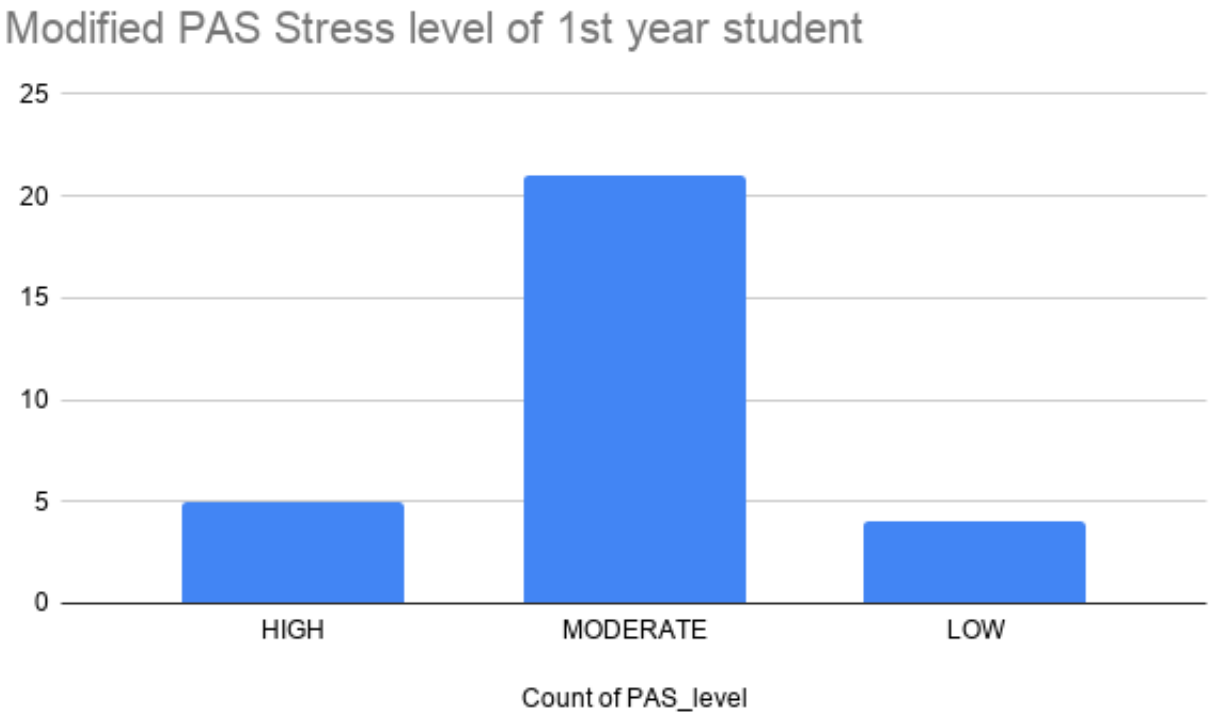


Figure 3.15: modified PAS stress level of 1st year's students

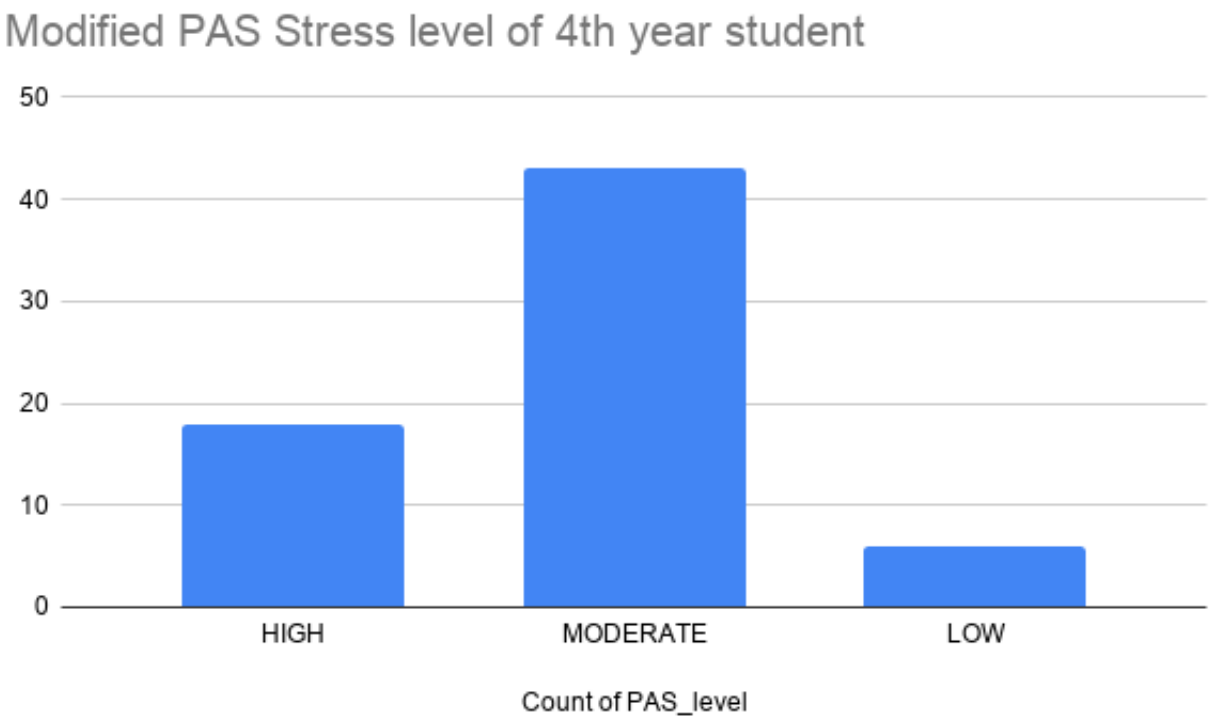


Figure 3.16: modified PAS stress level of 4th year's students

Chapter 4

Algorithms

We used some machine learning algorithms for our research. Those are explained in this section.

4.1 SVM(Support Vector Machine)

SVMs are intrinsically two-class classifier. This paper will briefly discuss how we can use SVMs when we have more than two classes [23]. There are two main approaches will be discussed: (1) one-against-all classifiers and (2) multiclass SVMs. One-against-all classification is the simplest way to adapt SVMs to multiclass classification. Under this Algorithm, rather than solving a common learning problem using the Ly categories, Ly solves boolean cases, one of which involves classifying the existing class j among the other ones. So it simply constructs Ly datasets, for each of which the label is :

$$y_i j = (y_i = j), \tag{4.1}$$

and we get Ly weight vectors w1, . . . , wLy . To classify a new point x, the idea is very simple: The further point is from the assessment boundary throughout the "positive" direction, the more likely it can be a part of that class. But also precisely pick the point where the point is quite far from the border line in a positive direction and set the class according to:

$$y_? = argmaxjh(x_?)\Delta w_j + w0, j \tag{4.2}$$

One-against-all classification is reasonable and can work quite well, though it requires training multiple SVMs. A single SVM can be trained to perform multiclass classification directly. The intuition behind the multiclass SVM is that, if the classification rule is

$$y_i = argmaxjh(x_i)\Delta w_j + w_0, j \tag{4.3}$$

then it should simply make sure that if y i = j, then h(x i)· wj +w0,j is greater than h(x i) · wj 0 + w0,j0 for all j 0 6= j by the largest margin, in the same way that we make sure that h(x i) · wj + w0,j 1 in the binary SVM. So we can directly optimize over all of our decision boundaries with constraints that enforce it, and the same objective as before (but now summed over all decision boundaries):

min w1,...,wL1 ,w0,1,...,w0,Ly ,s1,...,sN 1 2 X Ly j=1 ——wj ——2 + X N i=1 si such that wyi · h(x i) + w0,yi + si wj 0 · h(x i) + w0,j0 + 1 i 1, . . . , N, j 0 6= y i si 0 i 1, . . . , N

4.2 Logistic Regression

logistic regression is one of the foremost crucial and broadly utilized Machine Learning Calculations. Logistic regression is ordinarily among the primary few themes which people pick whereas learning prescient modeling. Logistic relapse isn't a regression calculation but a probabilistic classification model.
Multinomial logistic regression may be a type of logistic regression used to predict an aim factor of more than two categories. It could be a modification of logistic regression using softmax function instead of just the sigmoid function of cross-entropy failure work. Softmax research demolishes all values for run [0,1] and all components are 1[24].

$$softmax(x)i = exij = 1nexi \tag{4.4}$$

Cross entropy is a measurement about how close the two probability distributions are with each other. If p and q both are discrete:

$$Hp, q = xp(x)(x) \tag{4.5}$$

This study has an extension of [0, inf] and grows to when p = q and limitlessness when p is exceedingly minimal relative to q or poor habit. In case x, the course scores are taken from vector z = Wx+b, in which W could be a CM lattice and b may be the length of C vector of inclinations. Throughout this series, the name y is defined as a one-hot vector that rises to 1 for the proper lesson c and everything else. The tragic loss for the study of case x with expected dispersion of lesson y and restoration of course c.
There are now two popular ways for doing a multi-class classification using the double classification equation of the measured relapse: one-vs-all and one-vs-one. In one-vs-all, the C independent parallel classifier is trained for each course and runs all of the above classifiers onto each unused x-case and takes the course with the largest score. In one-vs-one, C selects 2 classifiers = C(C-1)/2 for each possible mix of course, and selects the course with the highest votes, while expecting a modern case.

4.3 Decision Tree

This is a decision support system that enables a tree-like decision pattern and its possible effects, including the consequences of chance incidents, resource costs and utility. That is one way to interpret an algorithm that includes only conditional control statements. This method can be used for both binary-class and multi-class classifiers. This is a supervised machine learning mechanism for the initialization of a decision tree from training data. The decision tree is often referred to as a category

tree or a reducing tree. This is a statistical model, consisting of a representation of observations on the topic of results on its final value. In this tree system, leaves contain classifications (also called labels), non-leaf nodes are features, and branches have combinations of features that refer to classifications.

The decision tree includes a number of nodes for testing features and edges for splitting by values of the specified feature and leaves the labeling groups anywhere each leaf contains a distinctive subclass. The decision tree consists of two primary procedures: one for the creation of a tree and the other for classification[25].

Building the tree: Building a decision tree is a way to create a tree from head to toe. The whole range of training starts at the base. The goal is to find the best test attribute in each tree decision node, permitting the mixture of categories in each test-generated subset to be minimized as far as possible. This process shall continue for each sub-decision tree until the leaves and their respective groups are achieved.

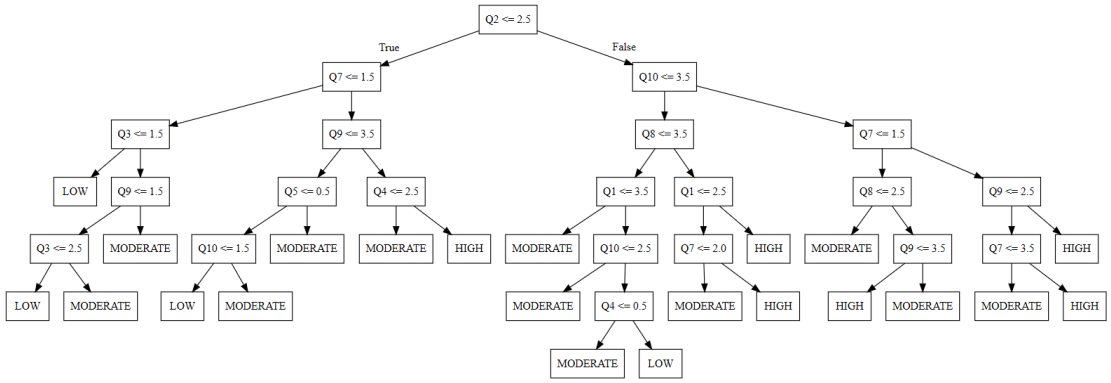


Figure 4.1: PSS scale's decision tree

In figure 4.1, the tree is built by doing entropy calculation and the root is Q2 which means it has the highest impurity for classifying the classes and the leafs are the output classes(LOW, MODERATE, HIGH Stress Level).

Classification: The category of new items relies on the built tree. It starts from the root to define the object, determines the specific test attribute, and takes the branch that fits the test outcome. This process is repeated before a leaf is found. The new entity is then marked as a class that labels the leaf. When there is a dataset of "m" training examples, each of which contains information in the form of different features and a label, then there is a finite set of classes in multiclass classification and each training example has "n" features as well. Each direction from the root to the leaf in the decision tree correlates to a combinations of test attributes, and the tree is defined as a disjunction of certain attributes.

4.4 Naive Bayes

The Naive Bayes classification is a form of predictive machine learning and is used for grouping tasks. Naive Bayes Classifiers are a group of Bayes 'Theorem-based classification algorithms. The entire argument of the classifier is based on the theorem of the Bayes. It's not a specific algorithm, but a collection of algorithms that all share a similar concept[26].

The Bayes theorem is defined below:

$$Py|X = PyP(y)P(X) \tag{4.6}$$

Using this Method, one will consider the expectation that Y will happen, provided that X has appeared. Here, X is the proof, and y is the assumption. The expectation here is that the predictors / features are unique. It is the addition of one element that does not impact the other and is considered naive.

X is given as,

$$X = (x1, x2, x3, ..., xn) \tag{4.7}$$

Here x1, x2, x3,....., xn represent the features. By substituting X and extending according to the chain rule we get,

$$Py|x1, \ldots, xn = PyPyPyP(y)Px1Px2P(xn) \tag{4.8}$$

By looking at the dataset, it will obtain the values for each one and replace them in the equation. The denominator does not change for all entries in the dataset, it remains unchanged. When the denominator remains unchanged for a particular data, the term may be removed:

$$Py|x1, \ldots, xnPyni = 1Py \tag{4.9}$$

Consider the chance of a particular set of inputs with all possible class variable y values and take the output with the greatest chance. This can be expressed as follows:

$$y = argmaxyPyni = 1Py \tag{4.10}$$

But for multiclass classification if test examples are considered as m and each class is considered as k then it is expressed as:

$$y = argmaxkPdatam \tag{4.11}$$

iAs a predictive method, the algorithm can be effectively applied and the predictions made very efficiently. So this is easily scalable, and is the algorithm of preference for actual-world implementations that is supposed to immediately respond to client needs.

4.5 Random Forest

This paper uses an improved-RFC (Arbitrary Forest Classifier) approach to multi-class classification. This consists of a combo of Atypical Woodland Machine Learning Analysis, a Feature Evaluator Strategy as well as an Occasion Channel Strategy. It is difficult to make improvements in the implementation of the Irregular Timberland estimation. The operation comes to the conclusion that perhaps the suggested improved-RFC solution performs better than the Arbitrary Timberland estimate with an improvement in classification precision of up to 97.80% for multi-class dataset.
The enhanced RFC approach begins with the collection of a multi-class data set for classification. The CFS, SU and Pick up Proportion Attribute Evaluator Strategy is picked and added to the dataset preparation to acquire essential classification attributes (calculation 1 step 1). After implementing the property evaluator, the occasion filter-Resample is effectively related to adjust the flow of the multi-class data set (calculation 1 step 2). Use of such a case filter-Resample is deliberate in the better RFC strategy. On the off situation that the data spreads of the course can be skipped at this point on step 2.

In the above way Irregular Timberland classification estimation (calculation 1 step 3) is related to the result obtained from calculation 1 step 2. The subsequent classification accuracy is obtained from calculation 1 step 4. Efficiency measurements-Classification Accuracy, F-Measurement, ROC, Affectability and Specificity are famous for the application of the improved-RFC method with respect to each Feature Evaluator Strategy-CFS, SU and Benefit Ratio.

4.6 Chi-Square

This is an algorithm that is widely used to evaluate relations between categorical variables. Null hypothesis of the Chi-square test is that there is no connection between the categorical factors in the population; they are not linked.
The test is most widely used to assess Independence Tests through the use of a cross tabulation (otherwise named a bi-variate table). Cross Tabulation indicates the presentation of two absolute factors one at a time, with convergences of the groups of the factors resulting in the cells of the table. The Independence Test determines whether there is an interaction between the two factors by comparing the observed sample of reactions in the cells with the scenario that would be usual if the factors were completely separate from each other. Computing and comparing the Chi-Square calculation to the vital value of the Chi-Square distribution allows the observer to determine if the observed cell counts are not exactly the same as the normal cell counts[20].

The analysis of the Chi-Square calculation is very straight forward and intuitive.

$$\chi^2 = \sum \frac{(f_0 - f_e)^2}{f_e} \tag{4.12}$$

where f_0 = observed frequency (the observed counts in the cells)

f_e = expected frequency if NO relation existed between the variables

As illustrated in the equation, the Chi-Square calculation relies on the difference between what is really observed in the details and what would be expected if there was really no relation between the variables.
Different important criteria arise when using the Chi-square calculation to test the cross tabulation. According to how the value of the Chi-Square is measured, it is very difficult to measure the size of the sample when the size of the sample is excessively large (500), even every slight difference would seem measurably remarkable. It is also important for the conveyance within the cells, and SPSS sends a warning message if the cells have fewer than 5 occurrences. All of which can be achieved by the continuous use of unmitigated variables with a specific number of classes[20].

4.7 Cronbach's Alpha

A straightforward indicator used to determine the reliability of the quality or internal reliability of the composite score is Cronbach's alpha. The alpha of Cronbach gives us a basic method of determining whether or not a score is good.For example,, a company might grant a work fulfillment study to their re-presentaive. High reliability means the job satisfaction of this butt and low reliability means something else is calculated. The Test has to see multiple questionnaires. The test is not that easy for actual life.The general rule of thumb is that the alpha of a Cronbach.70 and from 0.8 t0 0.9 is good and above that is excellent[21].

The Cronbach's Alpha is calculated by the following formula:

$$\alpha = \frac{N.\bar{c}}{\bar{v} + (N-1).\bar{c}} \tag{4.13}$$

Where, N= The number of items.
$\bar{c}$= average co-variance between each factors.
$\bar{v}$= average variance.

The Table for Alpha Values:

Cronbach's Alpha	Internal Consistency
$\alpha \geq 0.9$	Excellent
$0.9 > \alpha \geq 0.8$	Good
$0.8 > \alpha \geq 0.7$	Acceptable
$0.7 > \alpha \geq 0.6$	Questionable
$0.6 > \alpha \geq 0.5$	Poor
$0.5 > \alpha$	Unacceptable

Table 4.1: Consistancy of Cronbach's Alpha

4.8 Shapiro-Wilk

The way of explaining if a random sample comes from a normal distribution is called Shapiro-Wilk test. The method gives you a value which is W; if the value of W is LOW then it indicate that the data is not normally distributed. This method has several drawbacks. It has a disadvantage for larger datasets. If the sample is larger in size, it is difficult to achieve a statistically valid result[22].
The formula for Shapiro-Wilk is:

$$W = \frac{\sum_{i=1}^{n} a_i {x_i}^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2} \tag{4.14}$$

Where, x_i = The ordered sample dataset values
a_i = Constants generated by the variances, means and co-variances of the sample (size n) from a normally distributed dataset.

Chapter 5

Result Analysis and Discussion

As we already analyzed the dataset using the Shapiro-Wilk test and Cronbach's Alpha test to check the reliability of the dataset and we will be using Chi-square for selecting features and those selected features will be used for the further analysis to find the accuracy of different machine learning algorithm on multiple datasets. The accuracy of SVM, logistic regression, decision tree and naive bayes classifiers is given:

5.1 Results of Chi-Square Analysis

Features	Score
Q2	38.189
Q10	34.884
Q1	27.219
Q5	27.214
Q8	18.881
Q3	17.937
Q9	17.923
Q6	16.084
Q7	13.341
Q4	8.023

Table 5.1: Chi-Square of PSS Scale's features

In table 5.1, the chi square value for the features of PSS scale says that "Q4" and "Q7" has the least value among the all features of PSS scale. That determines that we can actually calculate the accuracy for the algorithms without having these features.

Both "Q9" and "Q4" are the least important , as they have the least feature importance for the PSS scale according to figure 5.1

From table 5.2, it is clear that "Q12" and "Q24" have the least Chi square value for the modified PAS Scale. So we calculated the accuracy without these two features.

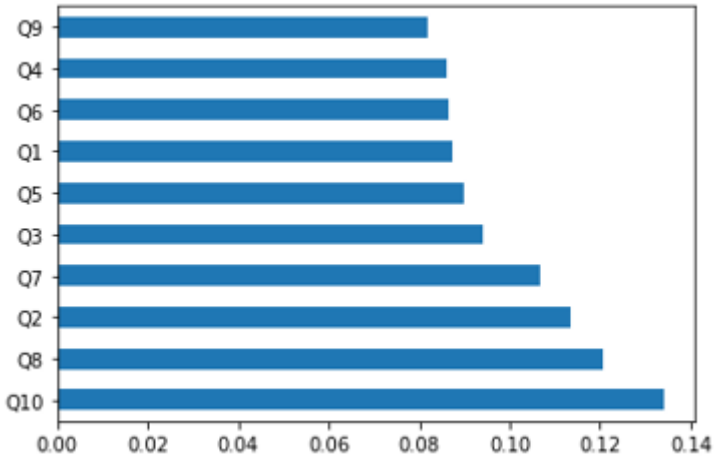


Figure 5.1: Feature Importance of PSS Scale

Features	Score
Q16	43.952
Q17	34.686
Q22	33.007
Q11	32.452
Q18	30.552
Q23	28.376
Q19	27.272
Q15	26.510
Q21	21.930
Q20	19.768
Q25	19.245
Q14	17.353
Q13	14.351
Q24	5.345
Q12	1.305

Table 5.2: Chi-Square of Modified PAS Scale's features

Furthermore, from the graph of figure 5.2 the feature "Q20" and "Q24" are the least important features among them.

Table 5.3 is very useful when the dataset is too big, that time it is very hard to consider all of the features of the dataset and need to prioritize the features for a very optimized solution.

5.2 Results of Support Vector Machine

From table 5.4, we can see that svm classifier has better accuracy without the Q4 feature on the PSS dataset. So it is better not to consider the Q4 feature if we use SVM classifiers for classification.

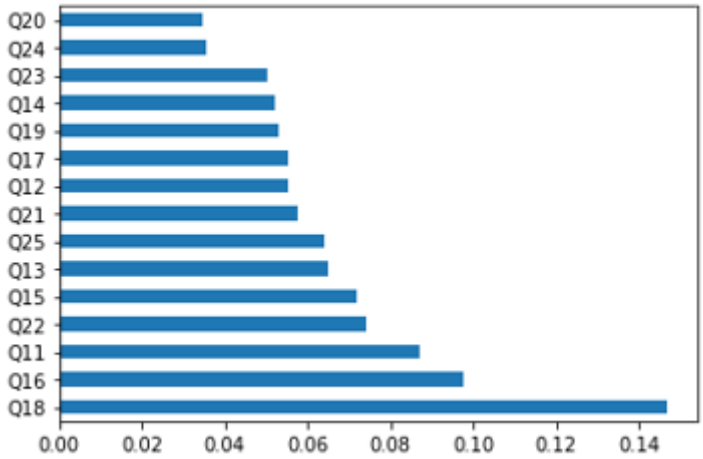


Figure 5.2: Feature Importance of modified PAS Scale

Dataset	Least one	Least two
PSS Scale	Q4	Q4,Q7
Modified PAS Scale	Q12	Q12,Q24

Table 5.3: Least scored(Chi-square) features for both Scale

Dataset	Classification Accuracy
PSS Scale (all 10 features)	0.889
PSS Scale (without Q4 feature)	0.926
PSS Scale (without Q4 and Q7 feature)	0.778

Table 5.4: Accuracy of SVM classifier for PSS Scale

24 outputs are predicted correctly when we considered all the features of PSS scale for the SVM Classifier(confusion matrix from figure 5.3.
Figure 5.4 shows that 27 outputs are predicted correctly when we consider all features except Q4 feature which proves that the calculation without this feature is the best.
Furthermore, Figure 5.5 shows that 21 outputs are predicted correctly when we consider all features except the Q4 and Q7 feature for modified PAS scale.

Table 6.8 shows that the SVM classifier does more accurate classification of modified PAS Scale without the Q12 and Q24 feature.

Figure 5.6, 5.7, 5.8 indicate that correctly predicted datas are 20, 22 and 23 respectively. As a result if we use SVM classifiers for the modified PAS scale then we should consider every feature except Q12 and Q24.

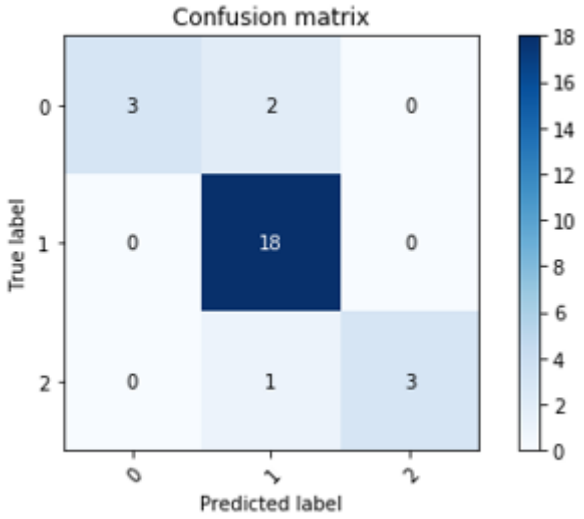


Figure 5.3: confusion matrix for PSS Scale(All features) for the svm classifier
0 is LOW Stress level, 1 is MODERATE Stress level and 2 is HIGH Stress level

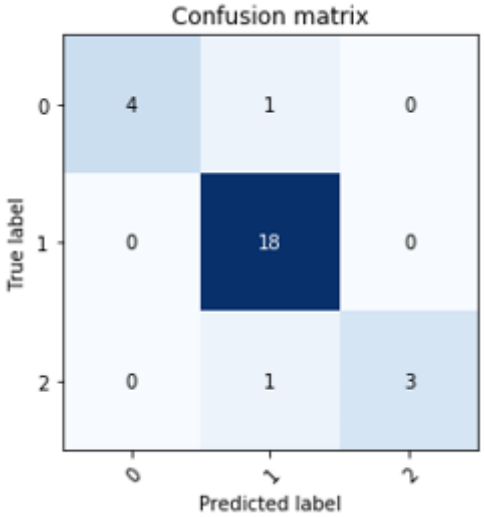


Figure 5.4: confusion matrix for PSS Scale(without Q4 feature) for the svm classifier
0 is LOW Stress level, 1 is MODERATE Stress level and 2 is HIGH Stress level

Dataset	Classification Accuracy
Modified PAS Scale (all 10 features)	0.741
Modified PAS Scale (without Q12 feature)	0.815
Modified PAS Scale (without Q12 and Q24 feature)	0.852

Table 5.5: Accuracy of SVM classifier for modified PAS Scale

Dataset	Classification Accuracy
PSS Scale (all 10 features)	0.667
PSS Scale (without Q4 feature)	0.741
PSS Scale (without Q4 and Q7 feature)	0.741

Table 5.6: Accuracy of Logistic Regression classifier for PSS Scale

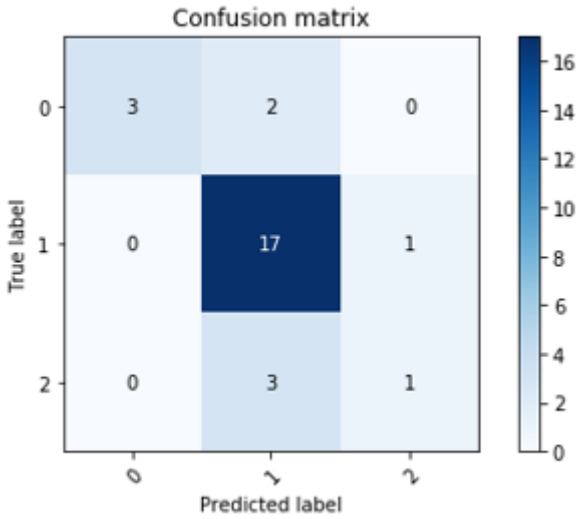


Figure 5.5: confusion matrix for PSS Scale(without Q4 and Q7 features) for the svm classifier
0 is LOW Stress level, 1 is MODERATE Stress level and 2 is HIGH Stress level

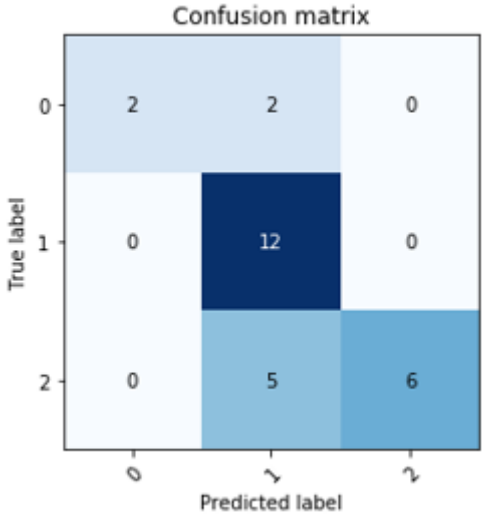


Figure 5.6: confusion matrix for modified PAS Scale(All features) for the svm classifier
0 is LOW Stress level, 1 is MODERATE Stress level and 2 is HIGH Stress level

Dataset	Classification Accuracy
Modified PAS Scale (all 10 features)	0.519
Modified PAS Scale (without Q12 feature)	0.593
Modified PAS Scale (without Q12 and Q24 feature)	0.519

Table 5.7: Accuracy of Logistic Regression classifier for modified PAS Scale

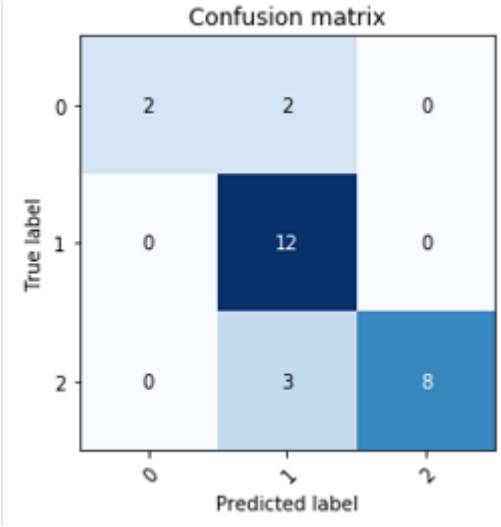


Figure 5.7: confusion matrix for modified PAS Scale(without Q12 feature) for the svm classifier
0 is LOW Stress level, 1 is MODERATE Stress level and 2 is HIGH Stress level

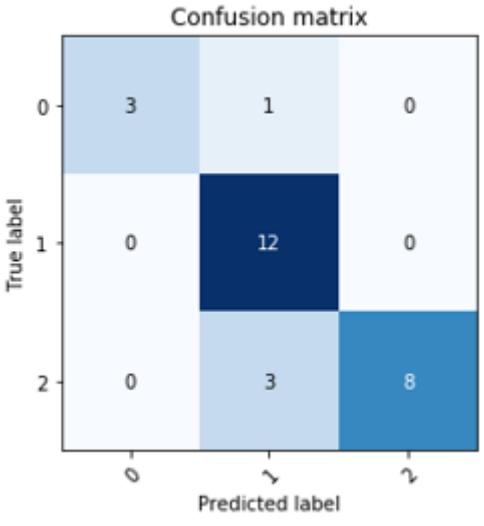


Figure 5.8: confusion matrix for modified PAS Scale(without Q12 and Q24 features) for the svm classifier
0 is LOW Stress level, 1 is MODERATE Stress level and 2 is HIGH Stress level

5.3 Results of Logistic Regression

Table 5.6 and 5.7 shows that logistic regression classifiers have better classification accuracy for pss scale(without Q4, without Q4 and Q7) and for modifies PAS scale without the Q12 feature.

5.4 Results of Decision Tree

Dataset	Classification Accuracy
PSS Scale (all 10 features)	0.778
PSS Scale (without Q4 feature)	0.778
PSS Scale (without Q4 and Q7 feature)	0.741

Table 5.8: Accuracy of Decision Tree classifier for PSS Scale

Dataset	Classification Accuracy
Modified PAS Scale (all 10 features)	0.593
Modified PAS Scale (without Q12 feature)	0.705
Modified PAS Scale (without Q12 and Q24 feature)	0.556

Table 5.9: Accuracy of Decision Tree classifier for modified PAS Scale

The accuracy of Decision Tree classifier is better when all the features are used or without the Q4 feature for PSS Scale and for modified PAS scale without Q12 feature (table 5.8 and 5.9)

5.5 Results of Random Forest

Dataset	Classification Accuracy
PSS Scale (all 10 features)	0.889
PSS Scale (without Q4 feature)	0.851
PSS Scale (without Q4 and Q7 feature)	0.851

Table 5.10: Accuracy of Random Forest classifier for PSS Scale

The accuracy of Random forest classifier is better when all the features are used for PSS Scale and for modified PAS scale without the two least valued chi features (Q12 and Q24) from table 5.10 and 5.11 respectively.

5.6 Results of Naive Bayes

For Naive Bayes classifier algorithm, PSS Scale without Q4 has better accuracy (table 5.11) and modified PAS scale has better accuracy without the two least valued

Dataset	Classification Accuracy
Modified PAS Scale (all 10 features)	0.704
Modified PAS Scale (without Q12 feature)	0.852
Modified PAS Scale (without Q12 and Q24 feature)	0.889

Table 5.11: Accuracy of Random Forest classifier for modified PAS Scale

Dataset	Classification Accuracy
PSS Scale (all 10 features)	0.926
PSS Scale (without Q4 feature)	0.926
PSS Scale (without Q4 and Q7 feature)	0.889

Table 5.12: Accuracy of Naive Bayes classifier for PSS Scale

chi features (Q12 and Q24) (table 5.12).

Among all of the classifiers, the accuracy of naive Bayes classifiers is the highest for both the PSS Scale and modified PAS Scale.

Dataset	Classification Accuracy
Modified PAS Scale (all 10 features)	0.926
Modified PAS Scale (without Q12 feature)	0.963
Modified PAS Scale (without Q12 and Q24 feature)	0.963

Table 5.13: Accuracy of Naive Bayes classifier for modified PAS Scale

Chapter 6

Conclusion

6.1 Conclusion

The main aim of our research is to predict the stress level of university students. We are using machine learning algorithms for detecting the factors of being stressed as a student of university. We hope that, by this research, it will be helpful for the university's authority to reduce the stress level of students which will make the students more productive. The main aim of our research is to help the university students to be more productive without being stressed which will create a positive impact on society.

In our research, there is huge scope for future amplification for further improvement. We thought of a few things to apply to them in the future. One such improvement is we will be able to measure the exact stress level to view help in academics and help psychologists to assume the stress level perfectly. But for immediate updates what we can do is to apply a deep learning algorithm and get much better accuracy. Applying deep learning will allow us to get a more appropriate stress level that will make the standard scale that we are willing to provide much more efficient. But now we are working mostly on increasing machine learning algorithm efficiency.

6.2 Future Work

We are interested in this field after our thesis also. In this thesis we only focused on the survey data (PSS and modified PAS Scale). In future we want to work on the EEG signals for the same stress scale to emphasize our work more for a strong impact on our society to help reduce the stress level by finding the most relevant factors of the university students. As EEG signals are more reliable than the manual hand written survey.

In this research we focused on the machine learning algorithms more for finding the classification Accuracy. In future we will implement deep learning algorithms for getting better accuracy for the classification of stress level. We know that deep learning algorithms help to learn about more accuracy. We will try to use the RNN algorithm for analyzing the stress level. As we already discussed that stress level is observed high for the university students for the academic purpose also.

In our thesis we worked on the stress level of university students. In future we will create a more relevant environment for getting the datas more accurate and will use eeg signal as it is reliable brain signal to get a more impactful result. Here, we will use CNN or Faster RCNN algorithm for getting better accuracy. These will help us to make more effective results to determine the stress level in future.

Bibliography

[1] Surveys.associatedpress.com, 2020. [Online]. Available:http://surveys.associated press.com/data/Edison/mtvU-Associated %20Press%20College%20Survey-Topline%20Frequencies.pdf. [Accessed: 31- Mar- 2020].

[2] Heri.ucla.edu, 2020. [Online]. Available: https://www.heri.ucla.edu/PDFs/ pubs/briefs/HERI_ResearchBrief_Norms2010.pdf. [Accessed: 31- Mar- 2020].

[3] G. Applebury, "Statistics on College Student Stress," LoveToKnow. [Online]. Available: https://stress.lovetoknow.com/Statistics_on_College_Student_Stress. [Accessed: 31-Mar-2020].

[4] "Facts," Anxiety and Depression Association of America, ADAA. [Online]. Available: https://adaa.org/finding-help/helping-others/college-students/facts. [Accessed: 31-Mar-2020].

[5] "Level of Stress among University Students." UKEssays.com. Accessed April 1, 2020. https://www.ukessays.com/essays/management/level-of-stress-among-university-students.php.

[6] Zheng, Sandra. "University Students Stress Levels." Medium. Medium, April 30, 2019. https://medium.com/@sandrazheng/university-students-stress-levels-6ce14da5690e.

[7] Lawton, Georgina. "Why Are Students at University so Stressed?" The Guardian. Guardian News and Media, May 31, 2019.https://www.theguardian.com/education/2019/may/31/why-are-students-at-university-so-stressed.

[8] "Publications about Depression," National Institute of Mental Health. [Online]. Available: https://www.nimh.nih.gov/health/publications/depression-listing.shtml. [Accessed: 01-Apr-2020].

[9] Khan , M. J., Altaf, S., Kausar, H. (n.d.). Effect of Perceived Academic Stress on Students' Performance. FWU Journal of Social Sciences, 146–151

[10] Cohen, S., Kamarck, T., Mermelstein, R. (1983). A global measure of perceived stress. Journal of Health and Social Behavior, 24, 385–396.

[11] Ebrahim, M. (2016). Perceived Academic Stress and Its Association with Student Characteristics. Journal of Applied Medical Sciences, 4, 1–15.

[12] M., B. N., Ahmed, M., B., P. (2018). Perceived stress and source of stress among undergraduate medical students of Government Medical College, Mysore. International Journal Of Community Medicine And Public Health, 5(8), 3513. doi: 10.18203/2394-6040.ijcmph20183090

[13] Bedewy, D., Gabriel, A. (2015). Examining perceptions of academic stress and its sources among university students: The Perception of Academic Stress Scale. Health Psychology Open, 2(2), 205510291559671. doi: 10.1177/2055102915596714

[14] Can, Y. S., Chalabianloo, N., Ekiz, D., Ersoy, C. (2019). Continuous Stress Detection Using Wearable Sensors in Real Life: Algorithmic Programming Contest Case Study. Sensors, 19(8), 1849. doi: 10.3390/s19081849

[15] Aihie, O. N., Ohanaka, B. I. (2019). Perceived Academic Stress among Undergraduate Students in a Nigerian University. Journal of Educational and Social Research, 9(2), 56–66. doi: 10.2478/jesr-2019-0013

[16] Zyl, Y. van, Dhurup, M. (2016). WHEN THINGS GET TOUGH, THE TOUGH GET GOING: UNIVERSITY STUDENTS' PERCEIVED STRESS AND COPING MECHANISMS. INTERNATIONAL JOURNAL OF SOCIAL SCIENCES AND HUMANITY STUDIES, Vol 8, No 1.

[17] Cohen, S. (n.d.). PERCEIVED STRESS SCALE.

[18] Wilks SE. Resilience amid academic stress: the moderating impact of social support among social work students. Adv Soc Work. 2008;9(2):106–125. [Google Scholar]

[19] Johnson S. Children's fear in the classroom setting. Sch Psychol Dig. 1979;8:382–396. [Google Scholar]

[20] Using Chi-Square Statistic in Research. (n.d.). Retrieved from https://www.statisticssolutions.com/using-chi-square-statistic-in-research/

[21] Stephanie. (2019, June 2). Cronbach's Alpha: Simple Definition, Use and Interpretation. Retrieved from https://www.statisticshowto.com/cronbachs-alpha-spss/

[22] Stephanie. (2017, October 12). Shapiro-Wilk Test: What it is and How to Run it. Retrieved from https://www.statisticshowto.com/shapiro-wilk-test/

[23] Levine, S. (n.d.). Week 7: Multiclass Support Vector Machines. Retrieved from https://courses.cs.washington.edu/courses/cse446/16sp/svm_3.pdf

[24] Jhunjhunwala, A. (2019, February 26). Is Logistic Regression a good multiclass classifier ? Retrieved from https://medium.com/@jjw92abhi/is-logistic-regression-a-good-multi-class-classifier-ad20fecf1309

[25] Bouchon-Meunier, B., Coletti, G., Yager, R. R. (2006). Modern information processing: from theory to applications. Amsterdam: Elsevier.

[26] Naive Bayes Classifiers. (2019, January 14). Retrieved from https://www.geeksforgeeks.org/naive-bayes-classifiers/

[27] What is heat map (heatmap)? - definition from whatis.com, https://searchbusinessanalytics. techtarget.com/definition/heat-map.

[28] Understanding q-q plots — university of virginia library research data services + sciences, https://data.library.virginia.edu/understanding-q-q-plots/.

[29] What is bar graph? - definition, facts example, https://www. splashlearn.com/math-vocabulary/geometry/bar-graph.