**MENTAL STRESS DETECTION USING MACHINE LEARNING ALGORITHMS**

**MINOR PROJECT - I**

Submitted by:

**SHUBHANGI MISHRA (9917103077)**

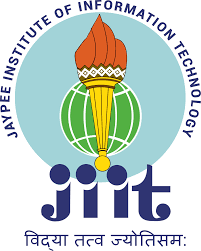
**VANSHIKA BHATNAGAR (9917103086)**

**SHIVANSH MITTAL (9917103087)**

**SANYA ATTRI (9917103101)**

Under the supervision of:

**DR. MUKTA GOYAL**

****

**Department of CSE/IT  
Jaypee Institute of Information Technology, Noida**

**OCTOBER 2019**

**Abstract**

**Background:** Stress is a necessary part of our lives and can have both beneficial and negative effects. The stress response is primarily determined by our perception of an event, transition, or problem. Finding a balance in our lives and managing our stress can be a challenge. An important first step is recognizing the degree to which we are affected by the stress in our lives and then move toward strategies to make it better. If stress could be detected without requiring special equipment, remedies to these situations could be undertaken.

**Aim:** The goal of this project is to detect stress via a questionnaire – the analysis of individual’s responses (on Likert scale) on questionnaire – and to map those responses and, analyse and classify using various machine learning algorithms.

**Data:** The questionnaire in a form was circulated to individuals of different age groups and belonging to different professions. Data was collected in csv format. Further, the data we received was hot encoded so as categorical variables were converted into a form that could be provided to ML algorithms to do a better job in prediction. The questions belonged to different sub-categories, stated as: (1) overcommitment, workload (2) enjoyment of work, self-realisation, empowerment   
(3) social distress, social support, social approval (4) recreational capacities, exhaustion (5) anxiety about the future, uncertainty.

**Method**: A questionnaire, consisting of 20 questions was used, in which each subject belonged to a sub-category of common stressors in an individual’s life. Their responses to the questionnaire were analysed, using a variety of machine learning techniques, to elucidate level of stress.

**Results:** Machine learning algorithms were successfully trained with an acceptable accuracy. Clusters were formed indicating the five stress levels: (1) very low (2) low (3) medium (4) high (5) very high. The selection of the final levels was dependent on several criteria: no (or very few) missing values and high variance in the answers meaning that category 0 to 4 had been chosen as far as possible.

**i**

**Table of contents**

**Page No.**

Abstract i

List of Figures ii

Abbreviation & Nomenclature ii

1. **Introduction 1**
2. **Background Study 2**
3. **Requirement Analysis 3**

3.1 Software Requirements

3.2 Hardware Requirements

3.3 Functional Requirements

3.4 Non-functional Requirements

1. **Detailed Design 4**
2. **Implementation 5-6**

5.1 Description of dataset

5.2 Screenshots of Graphs

1. **Testing Reports 7**
2. **Conclusion 8**
3. **Gantt Chart 8**

References iii

Appendices – 1, 2, 3 iii-iv

PowerPoint Presentation v-x

**List of Figures**

**Figure Title Page No.**

Fig 5.1 Visualisation of Responses 5

Fig 5.2 Responses in set of five clusters 5

Fig 5.3 y\_pred without PCA using SVM 6

Fig 5.4 y\_pred without PCA using Logistic Regression 6

Fig 5.5 y\_pred with PCA using Logistic Regression 6

Fig 5.6 y\_pred with PCA using SVM 6

Fig 6.1 Accuracy and confusion matrix of SVM with PCA 7

Fig 6.2 Accuracy and confusion matrix of SVM with 7   
Logistic Regression

Fig 6.3 Accuracy and confusion matrix of SVM without PCA 7

Fig 6.4 Accuracy and confusion matrix of SVM without 7  
Logistic Regression

**Abbreviations & Nomenclature**

PSS - Perceived Stress Scale

PCA - Principal Component Analysis

SVM - Support Vector Machine

**ii**

**1. Introduction:**

Stress is one of the main determinants of health status therefore an instrument to adequately measure stress is of prime interest not only in public health research but also for the examination of educational returns. School and workplace requirements are both essential sources of stress. We can define stress by saying that it involves the "set of emotional, physical, and cognitive reactions to a change." Thinking about stress as a reaction to change suggests that it is not necessarily bad, and sometimes, could even be a good thing. Stress is a common, familiar, and pervasive phenomenon. While typically viewed as a nuisance at worst and a good motivator at best, stress can be used as an indicator for a wide variety of phenomena.

Unfortunately, current methods for accurately detecting stress are intrusive and expensive. Such methods often cost a lot of money and typically involve attaching high-grade sensors and electrodes to an individual’s body to continuously monitor physiological measures for changes that are indicative of stress. At best, this is impractical. Not only is the cost highly prohibitive, but it would be nearly impossible to go through a normal day with such equipment attached. At worst, it is impossible; individuals will not be wearing such equipment. What is needed is a cheap and accurate technique to detect stress without the use of such sensors.

The purpose of this project is to determine the level of stress in individuals by using state-of-art machine learning algorithms. A questionnaire, consisting of 20 questions (Refer to Appendix 1) was used, in which each subject belonged to a sub-category of common stressors in an individual’s life. Their responses to the questionnaire were analysed, using a variety of machine learning techniques, to elucidate level of stress. The questionnaire in a form was circulated to individuals of different age groups and belonging to different professions. Data was collected in csv format.

The questions belonged to different sub-categories, stated as: (Refer to Appendix 2)   
(1) overcommitment, workload   
(2) enjoyment of work, self-realisation, empowerment   
(3) social distress, social support, social approval   
(4) recreational capacities, exhaustion   
(5) anxiety about the future, uncertainty.

Our questionnaire is inspired from the PSS Test by Sheldon Cohen. A set of 20 questions related to certain common conditions that generally occur and ask the subject to respond in their concern. It basically focuses to know their feeling and the extent of their stress. PSS can reduce the threat of extreme stress in an individual by diagnosing it at an initial level without investing much money at a preliminary stage.

**1**

**2. Background Study:**

In paper [4], the authors have developed and validated an instrument for measuring academic stress in accounting students. Statistical testing indicated that the instrument is a reliable and valid measurement of academic stress in accounting students. The validated instrument is subsequently administered to 98 accounting undergraduate students. The results of the present study indicate that differences in stress scores are based upon gender and educational level.

In paper [5], the authors have measured the existence of stress due to academics among university students. The study employed a quantitative research design where participants were screened using Academic Stress Scale. The five dimensions of sources such as personal inadequacy, fear of failure, interpersonal difficulties with teachers, teacher pupil relationship and inadequate study facilities were further analysed and gender differences were also obtained.

From papers [1] and [2], the authors have provided with the PSS Test and a stress questionnaire respectively. These are duly validated questions for testing stress levels in individuals – with their responses recorded on Likert scale.

In paper [3], the authors have calculated the mental stress of students one week before the exam and during the usage of the internet. Their objective is to analyse stress in the college students at different points in his life.

**2**

**3. Requirement Analysis:**

**3.1 Software Requirements**

**3.1.1 Library used:**

* Numpy
* Pandas
* Matplotlib
* Scikit

**3.1.2 Other Requirements:**

* Anaconda Platform (Jupyter Notebook)
* Python 3 or higher

**3.2 Hardware Requirements**

* Microsoft Windows 10
* Processor: Intel ® Core (TM) i5 -6200U CPU @2.30GHz 2.40GHz
* Ram : 4 GB and above
* Disk Space : 1 TB

**3.3 Functional Requirements**

* Appropriate data set to work on.
* Eligible software to implement our ideas.
* Suitable libraries for using algorithm in the source code.

**3.4 Non-functional Requirements**

* Validation: Applying benchmark functions for checking the accuracy and performance of algorithms.

**3**

**4. Detailed Design:**

**Data Collection**  
The questionnaire in a form was circulated to individuals of different age groups and belonging to different professions. Data was collected in csv format. A total of 307 were recorded and stored

**Data Pre-processing**  
Further, the data we received was hot encoded so as categorical variables were converted into a form that could be provided to ML algorithms to do a better job in prediction. Missing and redundant values were taken care of. Data was rescaled - since data comprised of attributes with varying scales, machine learning algorithms can benefit from rescaling the attributes to all have the same scale.

**PCA**The main idea of PCA is to reduce the dimensionality of a data set consisting of many variables correlated with each other, while retaining the variation present in the dataset, up to the maximum extent.

**Support Vector Machine(SVM)**  
This classifier that generally works upon the hyper plane. This algorithm works upon the ideal hyper plane which is more useful in sorting new illustrations. In a 2-Dimensional plane, it is a line isolating a plane in two sections where each class lies in either of the sides.

**Logistic Regression**  
Like all regression analyses, the logistic regression is a predictive analysis. Logistic regression is used to describe data and to explain the relationship between one dependent binary variable and one or more nominal, ordinal, interval or ratio-level independent variables.

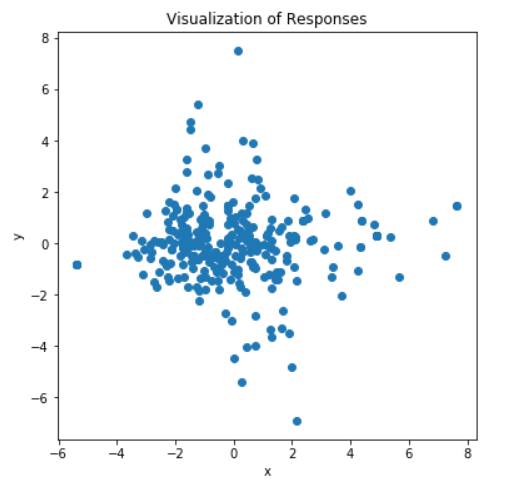
**4**

**5. Implementation:**

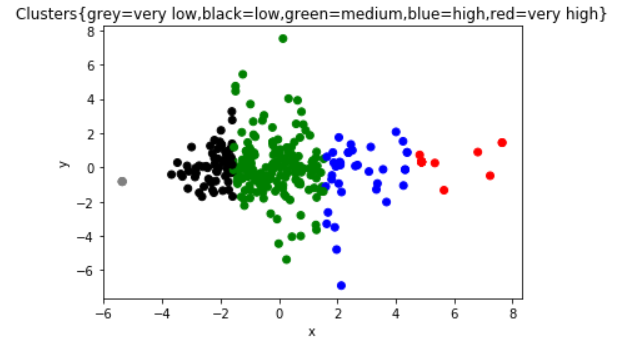
**5.1 Description about dataset**

Dataset consists of 307 recorded responses to a 20 set of questions provided by a questionnaire form. The responses were recorded on a Likert Scale ranging from 0 to 4, namely – very low, low, medium, high, very high stress level. Data was collected in csv format. Further, the data we received was hot encoded so as categorical variables were converted into a form that could be provided to ML algorithms to do a better job in prediction.

**5.2 Screenshots of Graphs:**

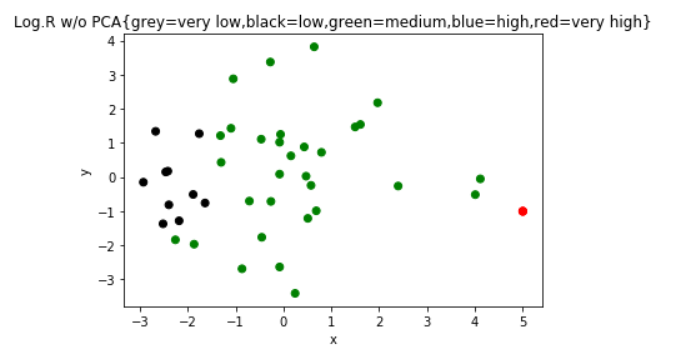
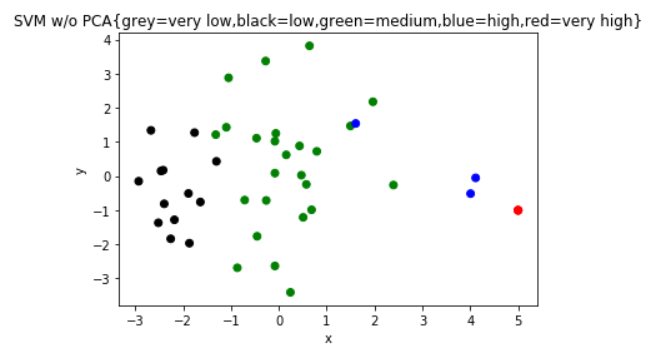


**(Fig 5.1 – Visualisation of Responses)**



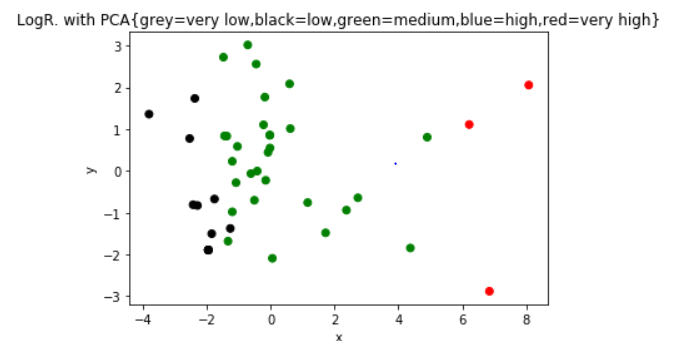
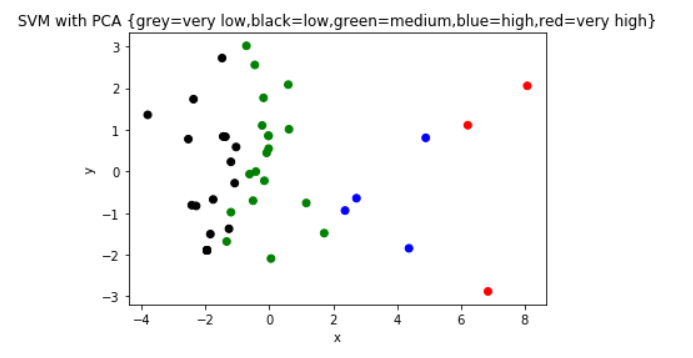
**(Fig 5.2 - Responses in set of 5 clusters)**

**5**



**(Fig 5.3 – y\_pred without PCA, using SVM) (Fig 5.4 –y\_pred without PCA , using**

**Logistic Regression)**

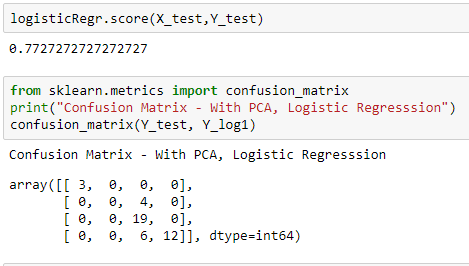
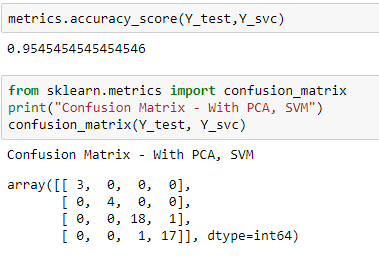
****

**(Fig 5.5 –y\_pred with PCA, using (Fig 5.6 – y\_pred with PCA, using SVM)**

**Logistic Regression)**

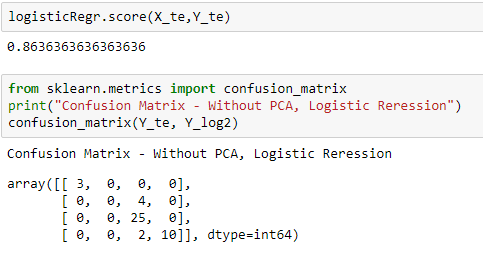
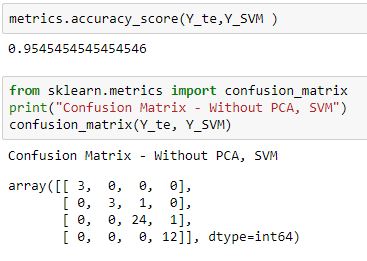
**6**

**6. Testing reports:**



**(Fig 6.1 – Accuracy and Confusion Matrix (Fig 6.2 – Accuracy and Confusion Matrix of**

**of SVM with PCA) Logistic Regression with PCA)**



**(Fig 6.3 – Accuracy and Confusion Matrix (Fig 6.4 – Accuracy and Confusion Matrix**

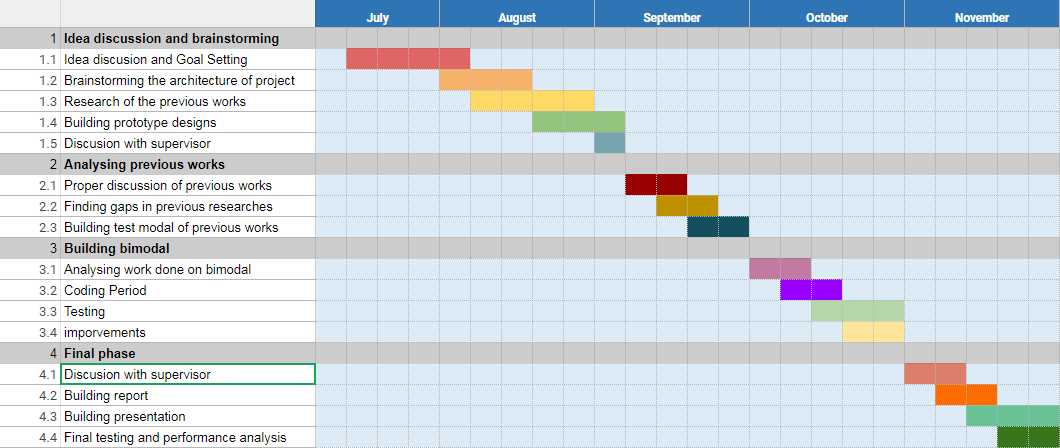
**of SVM without PCA) of Logistic Regression without PCA)**

**7**

**7. Conclusion:**

From the dataset (Refer to Appendix 3), we infer that the level of stress individuals face due to the mentioned factors are as follows:   
(1)anxiety about the future, uncertainty – High Level Stress (33.1%),   
(2)recreation, exhaustion – High Level Stress (32%),   
(3)self-realisation – Low to Medium Level Stress (24.9%),   
(4)workload, overcommitment – High Level Stress (29.6%) and   
(5)social distress – High Level Stress (30%).  
The accuracy for Logistic Regression without PCA is 86.36% and with PCA decreases to 77.27%. The accuracy for SVM without PCA is 95.45% and remains same even when PCA is applied.

1. **Gantt Chart:**



**8**

**References**

1. <http://www.mindgarden.com/documents/PerceivedStressScale.pdf>

2. <https://www.isma.org.uk/wp-content/uploads/2013/08/Stress-Questionnaire.pdf>

3. Ahuja, R. and Banga, A. (2019). *Mental Stress Detection in University Students using Machine Learning Algorithms*.

4. Zhenghong, J. Hou and Shinde, Jaysinha S. and Willems, John (2013) Academic stress in accounting students: an empirical study. Asian Journal of Accounting and Governance, 4 . pp. 1-10.

5. Reddy K. J, Menon K. R, Thattil A. Academic Stress and its Sources Among University Students. Biomed Pharmacol J 2018;11(1)

**Appendix**

**Appendix 1**

**Stress Questionnaire:**

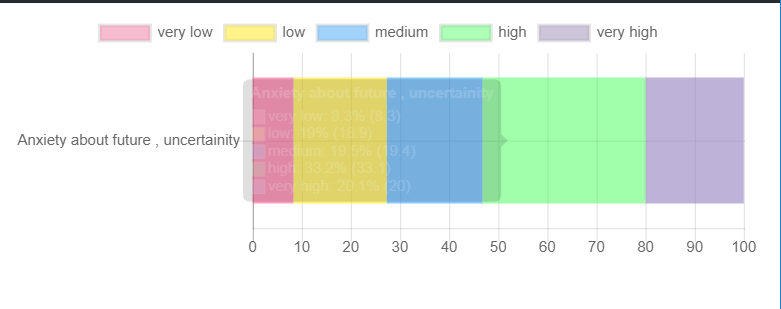
1. How often have you felt that you were unable to control the important things in your life?
2. How often have you felt that things were going your way?
3. How often have you been upset because of something that happened unexpectedly?
4. I find that I don't have time for my interests/hobbies outside of regular work.
5. How often have you been able to control irritations in your life?
6. How often have you found that you could not cope with all the things that you had to do?
7. How often do you get a full, restful night of sleep?
8. To what extent do you believe your life has purpose?
9. Most of my activities in life are shaped by society?
10. How often have you felt that you were on top of things?
11. How often have you been angered because of things that were outside of your control?
12. How often do you feel fear of failure in any competitive scenario?
13. How often have you felt difficulties were piling up so high that you could not overcome them?
14. How often have you felt confident about your ability to handle your personal problems?
15. I experience mood swings, difficulty in making decisions, concentration and memory is impaired.
16. I frequently have guilt if I procrastinate tasks.
17. In comparison with other people, how likely are you to see others as threatening, uncooperative or exploitative?
18. How often are you confused about the intentions of others towards you?
19. Do you feel stressed about not having enough money to participate in the same activities as peers?
20. Do you feel stressed about not managing your time well to be productive enough?

**iii**

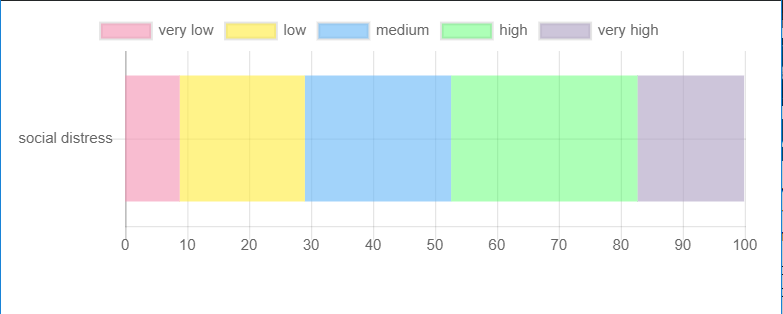
**Appendix 2**

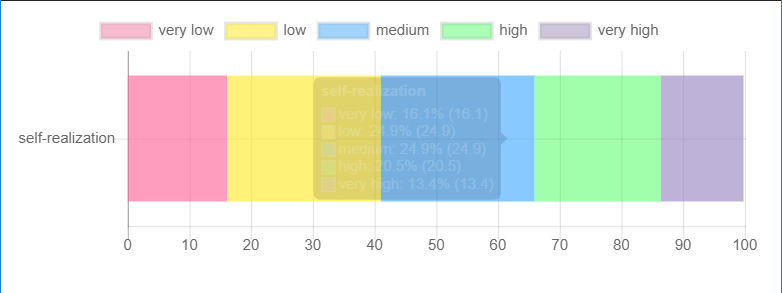
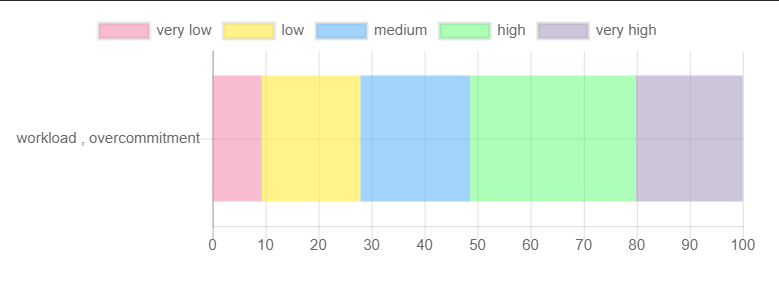
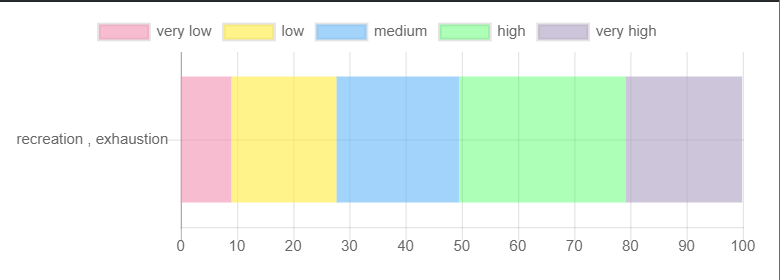


**Appendix 3**



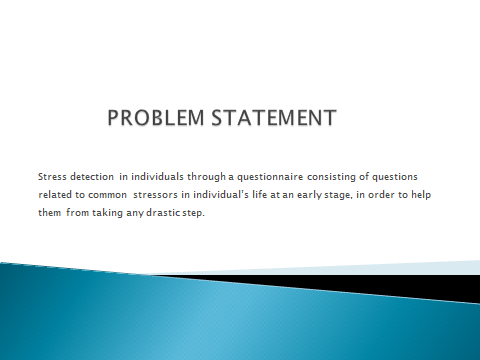
**iv**

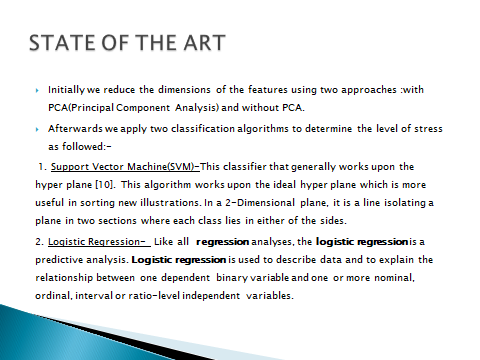




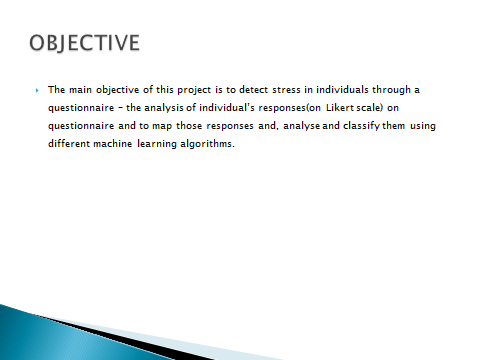
**v**

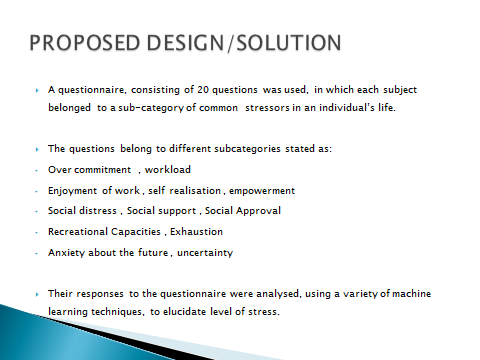
**PowerPoint Presentation**

****

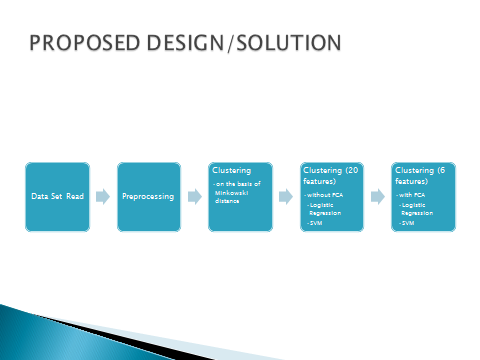
****

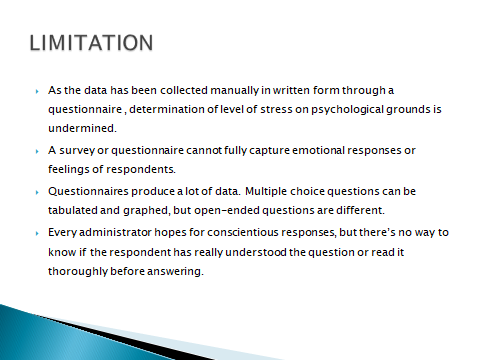
**vi**

****

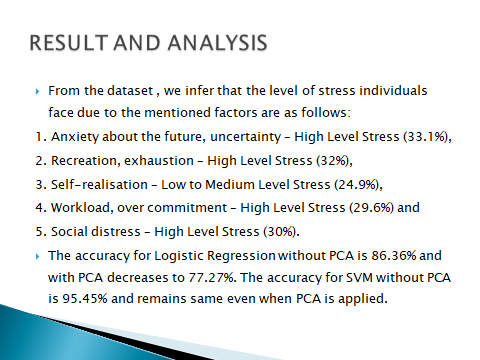
****

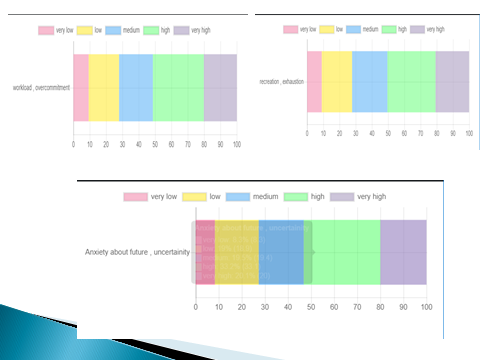
**vii**

****

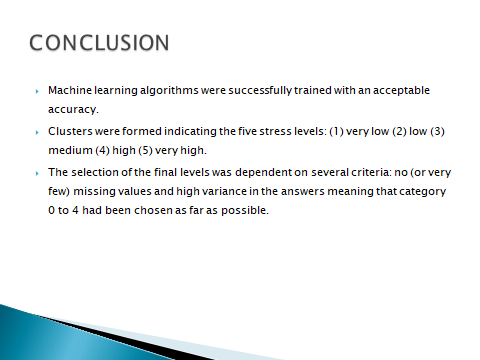
****

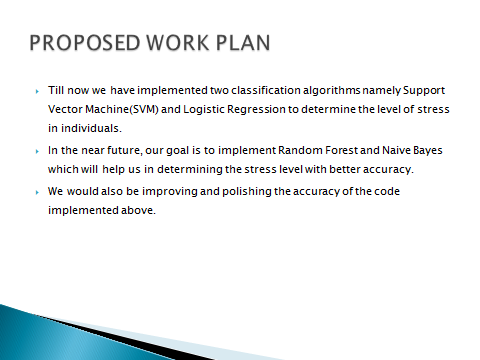
**viii**

****

****

**ix**

****

****

**x**