Escola d'Enginyeria de Telecomunicació i Aeroespacial de Castelldefels

UNIVERSITAT POLITÈCNICA DE CATALUNYA

MASTER THESIS

TITLE: Design and development of a students' performance predicting LMS utilizing machine learning based on mental stress level measured through a Bluetooth enabled smart watch

MASTER DEGREE: Master's degree in Applied Telecommunications and Engineering Management (MASTEAM)

AUTHOR: Syed Haider Ali Kazmi

ADVISOR: Cristina Barrado, Doctor Faiyaz, Angelica Reyes

DATE: June, 5th 2023

Abstract

Stress and academic anxiety problems can negatively impact numerous aspects of students' lives, resulting in degrading their academic achievement, quality of life, and social behaviour. Various research suggests that depression is associated with lower academic performance of students. The aim of this research is twofold. Firstly, in order to establish a correlation between students' mental stress level and their academic performance, a dataset has been compiled through gathering the data by conducting a survey in a university located in Punjab, Pakistan. The questionnaires were based on measuring the stress level of students using Perceived Stress Scale (PSS) , Cognitive performance assessment scale, in addition to some other demographic questions. Afterwards, this dataset has been analysed utilizing various machine learning algorithms. The second objective was to develop an innovative, affordable and smart performance predicting Learning Management System that takes into account students' mental stress while predicting the students' performance using machine learning models. The technique that was used for the mental stress measurements of the students was based on a phenomenon known as the Heart Rate Variability (HRV). A smart watch was utilized to measure the Heart Rate Variability of the students that was used to assess the stress level of students in academics. A Machine Learning (ML) model was trained using various parameters that were derived from the Heart Rate Variability. The original dataset that was used to train the model is known as Swell dataset. The SWELL dataset consists of HRV indices computed from the multimodal SWELL knowledge work dataset for research on stress and user modelling. The ML model effectively made prediction about the stress levels of the students with an accuracy of 98.1%.

CONTENTS

CHAPTER 1. INTRODUCTION

“Stress” is commonly defined as the external pressures on a person's bodily and mental well-being, whether they be physical or psychological. Stress is a process of interpreting and adjusting to external events, not merely a stimulus or a reaction. Humans are frequently affected by perceived and potential stress, which leaves them open to psychological issues and negative effects on their physical health. Increased level of stress may cause an individual to become disorganized and unsure about their objectives and aspirations. This could make it difficult for them to excel in their lives and, make it difficult to manage their time well in accordance with the situation. Several research suggests that mental stress can degrade the performance of individuals (Kiselica et al., 1994). In order to develop a deep understanding of such a critical issue, an in depth research has been conducted in this thesis, additionally, with the assistance of machine learning techniques, and utilizing artificial intelligence, a mental stress predictive model has been generated to understand how stress can hamper academic performance. The model thus created will provide a better insight regarding academic stress among students and would help in a better coordination between students, parents and teachers.

Chapter 1 gives an introduction to the thesis topic including the overall aim and research targets. A comprehensive scientific study was applied to the concept and the theme was broken down into an objective approach in order to evaluate the findings in a systematic way. In order to dive deeper into the content being presented in the thesis, Chapter 2 tries to understand the underlying theory and the brief description associated with the technology; the fundamental algorithms and the governing concepts related to machine learning (ML). Chapter 3 is related to the methodology that has been followed, and tries to dissect the various options that were taken under consideration for conducting the experiment and study. Chapter 4 gives detail insights of the dataset and provides an overview of the data analysis approach. In chapter 5, the advantages of using IBM Watson studio platform have been discussed. Additionally, we discuss the steps involved in implementing and configuring an AI machine learning experiment using IBM Watson studio. The experiment results are documented and presented in Chapter 6 as dedicated tables, plot diagrams and bar graphs. The thesis is concluded in Chapters 7 and 8, with an investigative overview of the research, drawing on conclusions and justification of thesis topic as well as recommendations to enhance further research.

1.1. Research motivation

Although short term stress might be considered as an academic incentive, long term chronic stress can impair performance, hamper growth and can cause health issues. Regarding the link between stress and students' academic performance, contentious conclusions have been published in the literature.

Consistent research has revealed that students with exceedingly higher levels of stress had poor grade point averages (Deng et al., 2022). Depending on the levels and factors, stress may or may not impede academic performance (Deng et al., 2022). Students may experience exceeding stress and feelings of academic burnout if they are unable to timely manage and finish their work load in the designated time. Additionally, students who are under a lot of stress have a tendency to put off tasks like finishing projects on time and meeting deadlines (Lin et al., 2020). Naturally, this will have an impact on their ability to study and the calibre of their work. This study is more inclined towards a general category of classification which classifies how students' academic performance is affected through stress.

The motivation behind this research work is to assist students and their teachers/parents to develop a better understanding of determining the impact of mental stress in an academic environment. Teachers/parents would be able to provide extra support to their students that would give these students a fair chance to achieve success.

1.2. Overall aim

Through the use of analysis of higher education data gathered through a survey, this project aims to assist students in achieving better academic performance while categorizing their mental stress. In terms of students' mental stress and academic performance, it would be self-explanatory if a ML-modelled thematic framework could be developed to assist university academic staff as well as students in determining if a particular student is tapping his/her maximum academic potential. The primary objective is to establish a correlation between the students´ perceived stress and their academic performance. The secondary aim of this quantitative study is to predict students mental stress while taking into account their Heart Rate Variability (HRV).

The method section provides a more deep explanation of the ML parameters being used. This study intends to assist university students and faculty members in gauging the developments of the students' stress and forecast future academic outcomes and scores.

1.3. Research agenda

The main agenda and the focus of the study is broken down further into the following objectives:

1. Determining the legitimacy of using ML algorithms for predicting student performance in academics.
2. Determining if there is a possible correlation between students' academic performance and their mental stress.
3. Taking advantage of IBM Watson ML platform for implementing the particular ML model that delivers the best accuracy, precision, recall and F1 score when fed with data collected from university students.
4. Implementing a ML model to predict the mental stress using Heart Rate Variability.
5. How accurately can we measure HRV from a wrist wearable device like a smart watch?

CHAPTER 2. THEORETICAL FRAMEWORK

Machine learning consists of a considerable amount of literature and research studies on numerous algorithms. The main objective of this domain is to tackle complex issues, hence, can be a daunting subject to approach.

2.1. Machine learning

Being a sub domain of artificial intelligence, machine learning targets problems in the same approach as humans do i.e. formulating a computational ML algorithms to assess the outcome of a challenge or problem. The basic driving principle of a ML model is to mimic the neural networks enmeshed in the human brain and making calculated conclusions through the information fed to the neural network model. Figure 1, elaborates the various domains of the artificial intelligence (William et al., 2021).

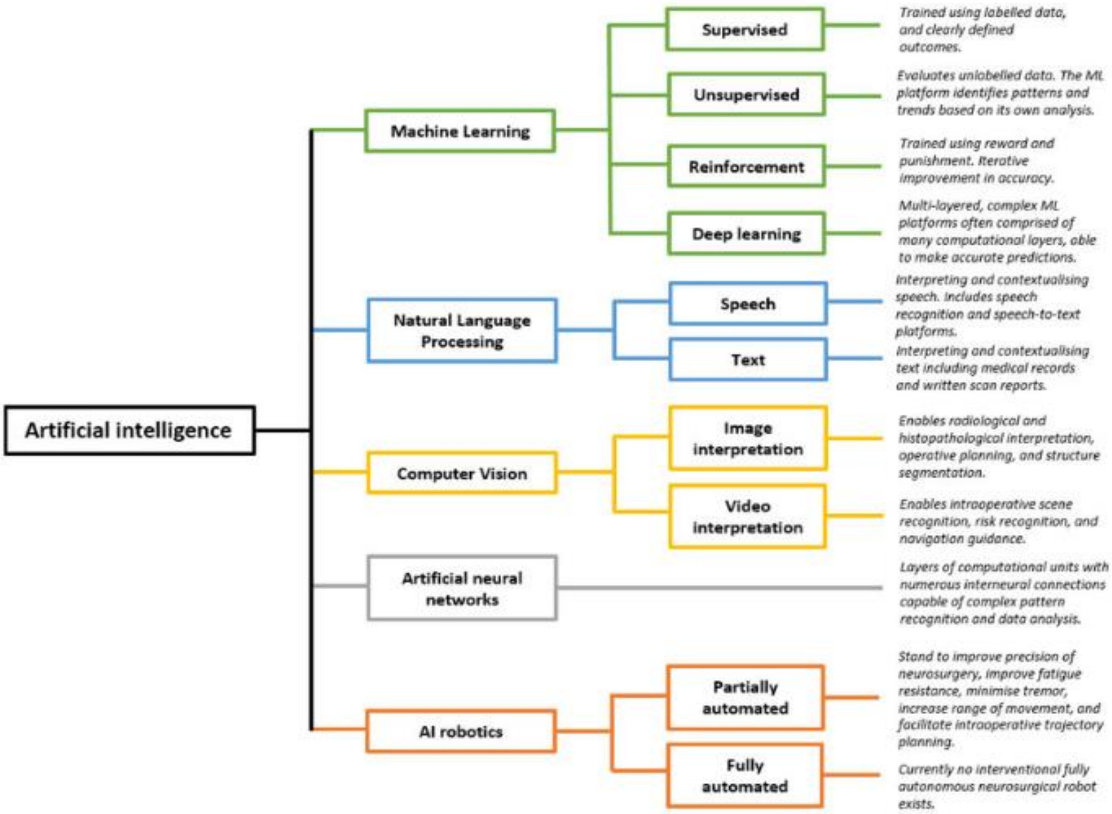

Fig. 1 Sub-domains of Artificial Intelligence

2.1.1. Supervised & Unsupervised machine learning

Supervised learning is a branch of machine learning and artificial intelligence that utilizes labeled datasets to train algorithms for accurate data classification or outcome prediction. In the training process, the algorithm adjusts its weights based on input data until it achieves an appropriate model fit, typically through cross-validation (IBM, n.d.). This methodology provides organizations with a potent tool to tackle various real-world issues, including spam filtering, image recognition, speech analysis, and medical diagnosis, provided there is an ample supply of labeled data (IBM, n.d.).

Supervised machine learning can be categorized into two primary types: regression and classification. Classification problems involve mapping input data to discrete class labels, such as distinguishing between "cat" or "dog," "malignant" or "benign," or "spam" or "non-spam." Several popular classification algorithms are available, including Naive Bayes, K-Nearest Neighbors, Decision Trees, and Support Vector Machines. Regression problems, on the other hand, deal with mapping inputs to continuous numerical outputs, like predicting the price of a house. Regression analysis aims to establish relationships between independent and dependent variables. Common regression algorithms include Linear Regression, Ridge Regression, Lasso Regression, and Neural Network Regression (Pykes, 2021).

Unsupervised learning, in contrast, is employed to uncover functional patterns, associations, and relationships within data. It can also be used for dimensionality reduction and data grouping. Clustering, a data mining technique, involves identifying similarities and dissimilarities among uncategorized data points and grouping them accordingly. The granularity and size of these groups are typically determined using the K-number, with K-means being a popular clustering algorithm.

2.1.2. Decision Tree

Decision trees are non-parametric supervised learning algorithms that can be used for both classification and regression tasks. They have a hierarchical, tree-like structure consisting of a root node, branches, internal nodes, and leaf nodes (IBM, 2023).

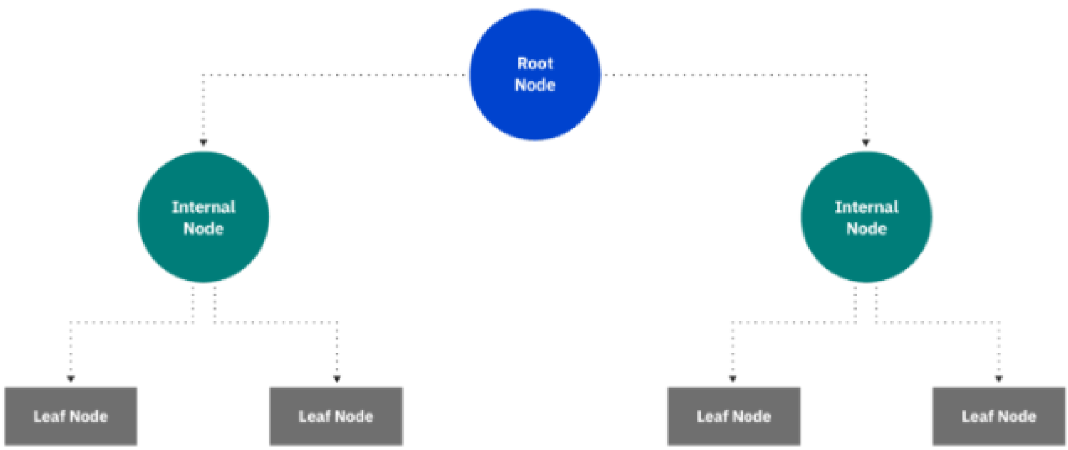

Fig. 2 A typical Decision Tree (IBM, 2023)

Decision trees are a type of machine learning algorithm that uses a tree-like structure to make decisions as depicted in figure 2. The algorithm starts with a root node, and branches off into internal nodes or decision nodes, which evaluate available features to form homogenous subsets represented by leaf nodes. For example, when deciding whether to go surfing, one might use a decision tree with rules like "Is the temperature warm enough?" and "Is the wind too strong?". The following decision rules can be followed to make the possible choice (IBM, 2023):

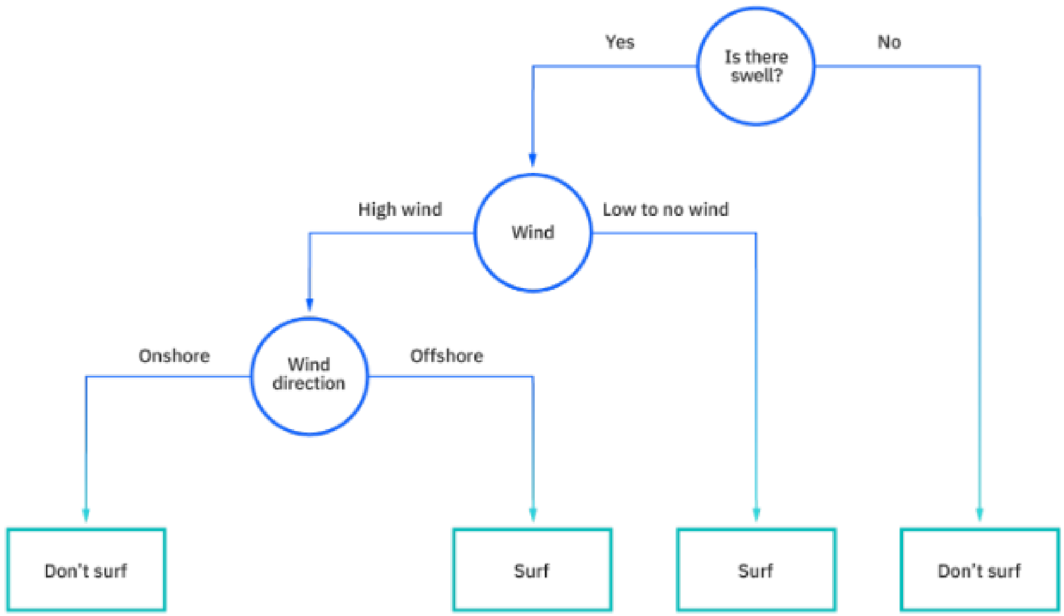

Fig. 3 Decision Tree initialization (IBM, 2023)

The flowchart, shown in figure 3, structure also creates an easy to digest representation of decision-making, allowing different groups across an organization to better understand why a decision was made (IBM, 2023).

To construct a decision tree, the algorithm follows a divide-and-conquer approach to find the optimal split points in the data for classifying records into specific class labels. This iterative process continues until either all records or the majority of records are correctly classified. It is preferable to have smaller trees as they can result in pure leaf nodes, while larger trees are more prone to overfitting and data fragmentation. To address these concerns and reduce complexity, pruning techniques can be employed. The accuracy of the decision tree model can be assessed through cross-validation or by creating an ensemble using the random forest algorithm, which combines multiple decision trees to improve predictive performance (IBM, 2023).

2.1.3. Boosting algorithms

Boosting is a technique used to enhance the accuracy of weak learner algorithms by transforming them into stronger learning algorithms. In boosting, each tree is trained on a modified version of the original dataset. For instance, the AdaBoost Algorithm assigns equal weights to each tree and trains them. After evaluating the outcomes of the initial tree, the algorithm increases the weights of observations that are challenging to classify and decreases the weights of those that are easier to classify. This iterative process allows subsequent trees to classify observations that were previously misclassified. The final ensemble model's predictions are determined by the weighted sum of the predictions made by the preceding tree models (IBM, 2023).

2.1.4. Light Gradient Boosting Machine

Figure 4, shows the scheme that is adopted by typical decision tree algorithms where the growth of the trees is level-wise.

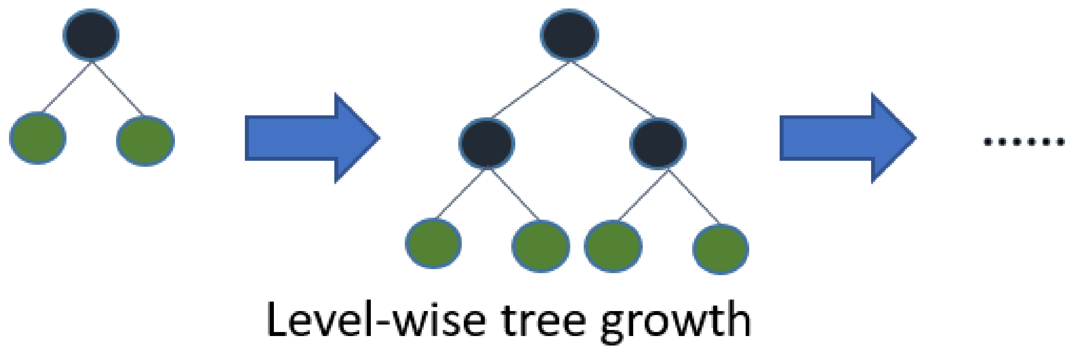

Fig. 4 A symmetric level-wise tree growth strategy in which each node in a particular level has child nodes, thus, developing an additional layer of depth (Microsoft Corporation, 2023)

On the other hand, for Light Gradient Boosting Machine (LGBM) ,developed by Microsoft, the growth of the trees is leaf-wise, as shown in figure 5. For growth, LGBM would choose leaf with maximum delta loss.

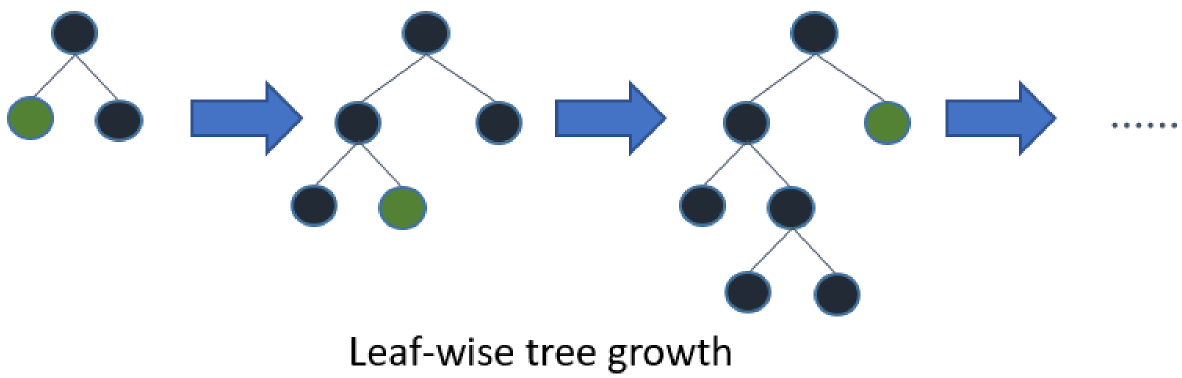

Fig. 5 Leaf wise tree growth in which only the node having the highest delta loss is split. Successive splitting occurs only on one side of the tree, thus, resulting in an asymmetrical tree (Microsoft Corporation, 2023).

In terms of memory consumption and computational speed, LGBM can substantially outperforms XGB (Ke et al., 2017).

2.2. Background and related work

Numerous machine learning (ML) projects have been developed for forecasting student performance. Some of the research work, related to the theme of this thesis, are elaborated in the following sections

2.2.1. Student performance evaluation in educational data mining

Two machine learning models were created and a comparative analysis was performed by researchers (Ahmed et al., 2021). The models being ANN model and the Random Forest. TensorFlow was integrated at the backend for both the models. They tried to develop the models in order to predict the academic success using students previous academic evaluation and geographical data. They achieved the required results by investigating that ANN can out preform the Random Forest model by using a sizeable amount of data. The increased interest among researchers in applying data mining techniques to evaluate student data served as a prerequisite for this study. Future implications could include using recurrent neural networks (RNN) for identifying students who are about to drop-out from college.

2.2.2. Predicting the academic performance of middle- and high-school students using machine learning algorithms

In their study (Rajendran et al., 2022), the authors utilized machine learning techniques to develop models for predicting the academic performance of high school students. The models took into account various socio-demographic factors such as age, gender, obesity, average household income, family size, and marital status of parents, as well as school-related variables like type of gender education and academic level, and student-related variables such as stress and lifestyle. The output variable considered in the models was the students' GPA. The results showed that the gradient boosting method outperformed other techniques, followed by random forest, in terms of generating better predictions. The analysis of the models led to the conclusion that maintaining a health-conscious lifestyle has a positive correlation with academic performance, while the presence of stress has a negative impact. Nevertheless, the impact of gender was not identified as a significant predictor of a student's academic performance.

2.2.3. Tracking and predicting student performance in degree programs

This paper (Xu et al., 2017) addresses several new challenges in the field. The authors propose an innovative approach for predicting the future performance of students in degree programs based on their current and past performance. To construct base predictors, they develop a course clustering method using a latent factor model. They also introduce an ensemble-based progressive prediction architecture to incorporate the evolving performance of students into the prediction process. These data-driven methods can complement other pedagogical approaches and provide valuable information for academic advisors.

The information can be used to recommend subsequent courses to students and implement pedagogical interventions if necessary. Furthermore, the findings of this study have implications for curriculum design in degree programs and the formulation of education policies. However, it should be noted that the study does not account for students' performance and stress specific to a particular college program, which according to the authors, could be significant factors influencing the outcomes.

2.3. Heart Rate Variability

The second part of the thesis focuses on a technique used to measure the mental stress of the students, which is based on a phenomenon called Heart Rate Variability (HRV). HRV analysis is a tool increasingly utilized for non-invasive analysis of the Autonomous Nervous System (ANS) in the human body. Its analysis and contextual application have gained importance due to its sensitivity to both physiological and psychological environmental factors. Altered HRV measurements are extensively utilized for monitoring the arrhythmic dysregulation of the Autonomous Nervous System. Additionally, HRV measurements are employed to monitor and assess sleep patterns, stress levels, drowsiness, and the effects of prolonged strenuous exercise training (Colom et al., 2010).

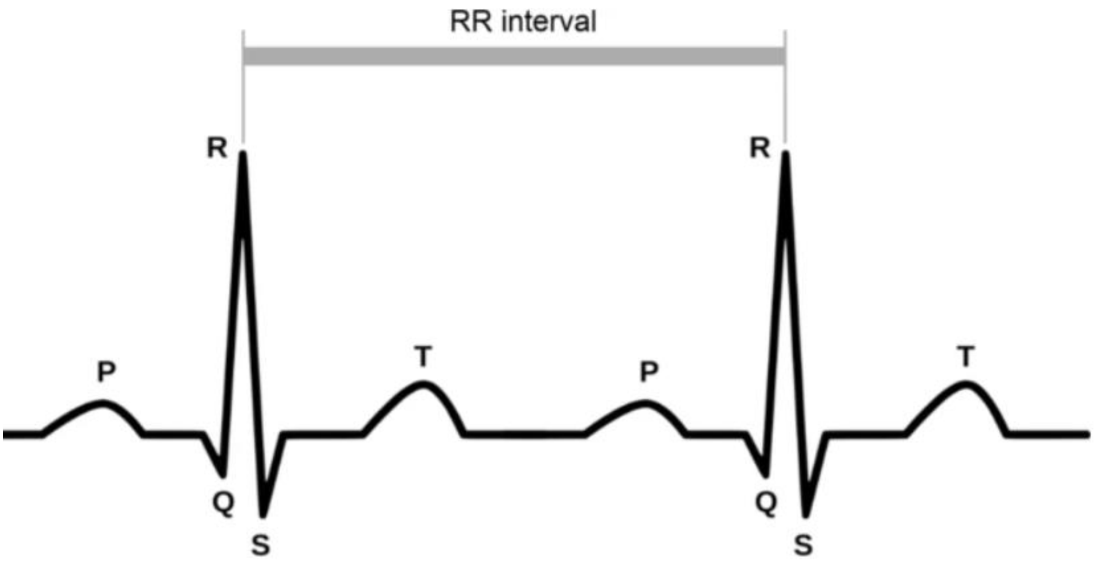

Fig. 6 RR interval representing the distance between two successive heartbeats

HRV refers to the irregular fluctuations in the duration between successive heartbeat intervals, which have been found to be influenced by stress in scientific literature. Specifically, the RR interval, representing the distance between two consecutive heartbeats (R waves of their QRS complexes), is directly associated with HRV, as shown in Figure 6 (Colom et al., 2010).

There are various approaches to reducing chronic stress and improving HRV. One potential method involves modulating the autonomic nervous system. During periods of stress, sympathetic nervous system responses become more dominant, triggering the "fight or flight" response. This heightened sympathetic activity leads to an increase in heart rate and a decrease in HRV. Conversely, the parasympathetic nervous system, responsible for the "rest and digest" state, becomes less active and restricted during stressful times (Welltory, 2023).

> **Low HRV:** " If the intervals in-between your heartbeats are relatively equal, then you are in a fight or flight state and your HRV is quite low."
>
> **High HRV**: "If the interval length variates and you are in a more relaxed state then your HRV is high. This is mostly associated with good recovery."

CHAPTER 3. METHODOLOGY

In this chapter project methodology that has been followed in the thesis has been elaborated.

3.1. Project methodology

The study was divided in three main phases.

The **first phase** consisted of a literature study. The study looked at previous related work in the form of research articles, surveys, journals, and e-books. This process was done to familiarize the reader on the current state-of-the-art ML techniques and to show a research gap, to justify the current research that was being conducted. Google Scholar was used to find these resources.

In the **second phase**, in order to establish a correlation between perceived mental stress and students´ academic performance a dataset was formed based on a survey that was conducted by asking 298 students to fill a survey form (attached in the Annex). All of the students were enrolled at Government College University, Lahore, and were studying in the first semester of their undergraduate degree. The stress levels of the students were measured using a Perceived Stress Scale (PSS) which is a classic stress assessment instrument. “The tool, while originally developed in 1983, remains a popular choice for helping us understand how different situations affect our feelings and our perceived stress. The questions in this scale ask about your feelings and thoughts during the last month. Individual scores on the PSS can range from 0 to 40 with higher scores indicating higher perceived stress.” (NH Dept. of Administrative Services, n.d.)

As per the Perceived Stress Scale (PSS) criteria, following are the thresholds defined (NH Dept. of Administrative Services, n.d.):

- Scores ranging from 0-13 would be considered low stress.
- Scores ranging from 14-26 would be considered moderate stress.
- Scores ranging from 27-40 would be considered high perceived stress.

The survey form consisted of five direct questions i.e., asking the students to specify their age, gender, self-study hours, number of times they skipped their school day and grades in the last three graded activities. The stress and cognitive performance levels of the students were gauged through the two psychological assessment scales (Perceived Stress Scale and Cognitive assessment scale, as attached in the appendix).

The **third phase** was an experimental phase. In this phase, deep analysis was performed with the help of a Bluetooth enabled Apple smart watch. A total of four students were considered for this experimental stage. Informed consent was obtained from each of the four students. The age group of the four students varied between 20 – 23 years.

The experimental phase was conducted during the mid-semester exams of the students. In the experimental phase, the perceived mental stress and Heart Rate Variability (HRV) was measured with the help of the Perceived Stress Scale (PSS) and Apple smart watch, respectively. The students were asked to wear the Apple watch during their exams.

3.1.1. SWELL Dataset

The model was trained using various parameters that were derived from the Heart Rate Variability. These HRV parameters were extracted from an original dataset, known as Swell dataset, that was used to train the model. The SWELL dataset consists of HRV indices computed from the multimodal SWELL knowledge work dataset for research on stress and user modelling (Hazer-Rau et al., 2020).

The original dataset consists of data that was captured using the following means (Hazer-Rau et al., 2020);

- Computer interactions, via a computer logging tool
- Facial expressions, via a webcam
- Body postures, via a Kinect 3D camera
- Physiology (ECG and skin conductance), via body sensors

As the third phase of this research work revolves around the idea of predicting mental stress from HRV (gathered through a smartwatch) of the students, hence, the model was developed using only the features related to the HRV (in Time Domain) that are as following (Shaffer & Ginsberg, 2017);

- **RR:** Interval between two heartbeats
- **MEAN_RR:** mean of RR-interval
- **MEDIAN_RR:** median of RR-interval
- **SDRR:** standard deviation of the RR-intervals
- **RMSSD**: It’s a measure for how much variation there exists in the heart rate. In a healthy heart, there is a natural variation, which is due to a balance between the sympathetic nervous system (SNS) and parasympathetic parts (PSNS) of the Autonomous Nervous System (Singh et al., 2018). If your body experiences stress, then the sympathetic system will activate, to prepare for fight or flight behaviour, and your heartrate will increase.
- **SDSD:** the standard deviation of the differences between successive NN intervals
- **HR:** Heart Rate
- **pNN25:** The number of pairs of successive RR-intervals that differ by more than 25 ms *(normal RR-intervals are often called NN-intervals)*
- **Output condition:** Stress/no stress

Table 1. HRV parameters calculations (Kim et al., 2022)

Parameter	Unit	Formula	Description
Time-domain HRV parameters			
MeanRRI	ms	$t\frac{\sum_{i=1}^{N} RR_i}{N}$	Mean variability of interbeat interval
RMSSD	ms	$\sqrt{\frac{\sum_{i=1}^{N-1}(RR_{i+1}-RR_i)^2}{N-1}}$	Reflects parasympathetic activity
pNN5	%	$\frac{Count((RR_{i+1}-RR_i)>5\text{ ms})}{N-1} \times 100$	Reflects parasympathetic activity

Table 1, shows the expressions used to calculate the HRV parameters in time domain. Once the HRV data of the four students was extracted from the Apple smart watch, these expressions were used to calculate the HRV parameters and fed to the trained (with SWELL dataset) ML model to evaluate the performance of the model.

3.1.1. Using Apple smart watch to measure HRV

Different wearable watches and gadgets have been studied and evaluated for HR estimation measurement. For a total of one minute of effective granularity, the Apple Watch has the best performance estimation (Hernando et al., 2018). This app stores the raw RR values, with a precision of centi-seconds, in the user's Personal Health Record, accessible to be exported in XML format using Apple's Health App (Hernando et al., 2018).

3.1.2. Measuring stress threshold level of each individual

According to Dr Andrea Dinardo, different individuals have different threshold levels as shown in figure 7. She further elaborates that "Thresholds are more individualistic while tipping points are more universal. Thresholds vary from person to person (e.g Type A vs Type B), situation to situation (e.g Work or Personal Reasons), and are based on individual strengths and challenges (Dinardo, 2016).

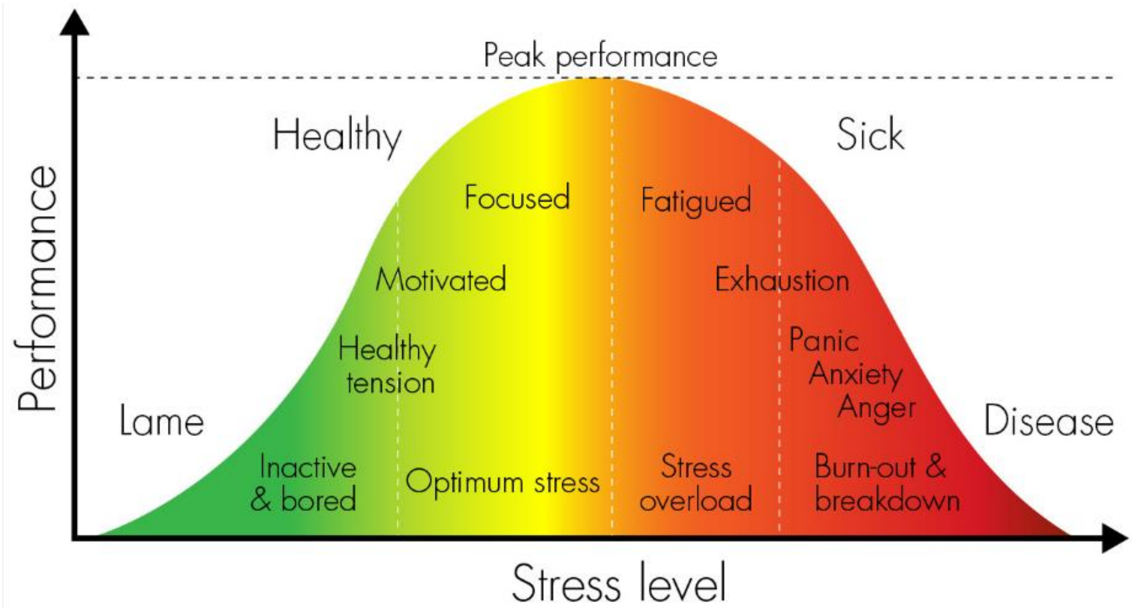

Fig. 7 How thresholds affect an individual stress response (Dinardo, 2016)

To establish a baseline stress measurement indicator, four different students underwent a pre-experimental phase in which research methods were employed. This phase consisted of two consecutive 20-minute stages: a relaxation stage and a stress stage, separated by a 1-minute break. In the relaxation stage, the students were instructed to listen to pleasant orchestral music. During the stress stage, they were required to complete an online version of the Stroop test, an attentional test that involves identifying the ink colour of words while disregarding their literal meaning (Hernando et al., 2018).

3.2. ML model performance measures

In this thesis, the typical parameters that have been used in evaluating the performance of a classification ML model based are as following;

3.2.2. Confusion Matrix

The confusion matrix is a summary of the classification model's performance in predicting outcomes across different classes. Vertical axis represents the predicted labels by the model and the horizontal axis represents the actual labels in the dataset.

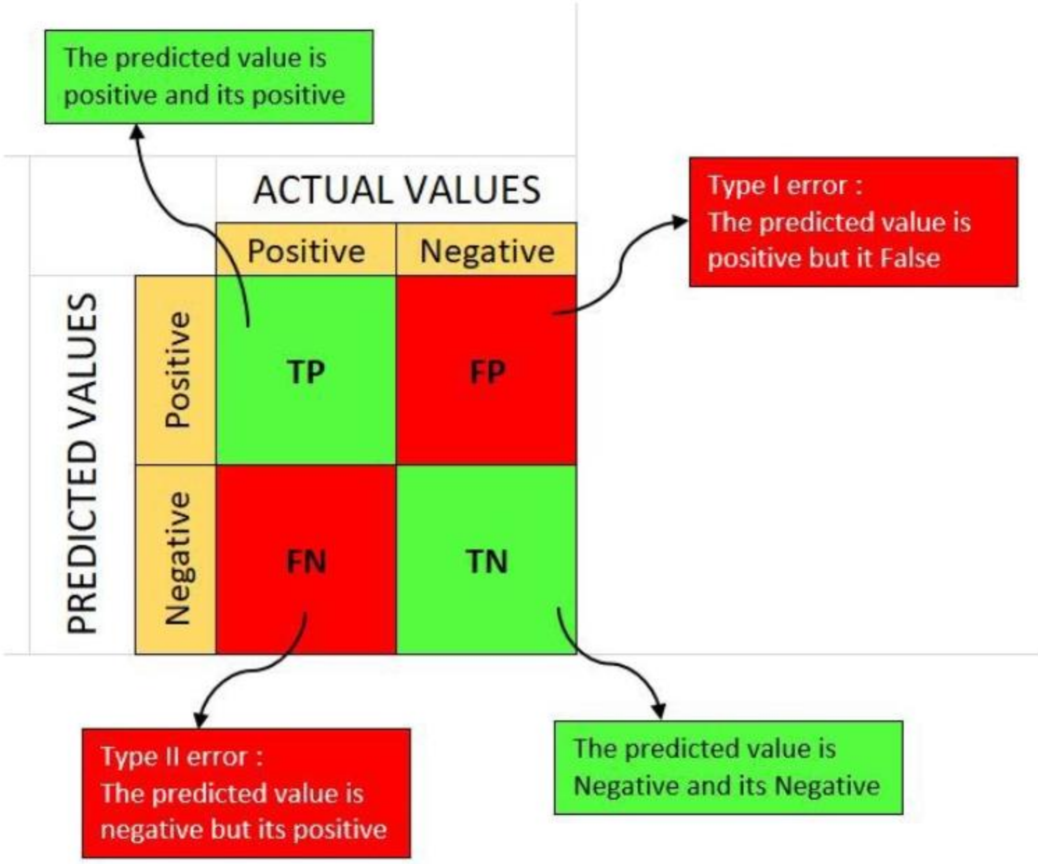

Fig. 8. Confusion matrix consisting of True Positives (TP), False Positives (FP), False Negatives (FN), Ture Negatives (TN) (Suresh, 2021)

3.2.3. Accuracy

Accuracy of an algorithm is represented as the ratio of correctly classified (TP+TN) instances to the total number of instances (TP+TN+FP+FN) (Singh et al., 2021).

$$Accuracy = \frac{(TP + TN)}{(TP + FP + FN + TN)}$$

3.2.4. Precision

Precision is represented as the ratio of correctly classified instances with positive class (TP) to the total instances predicted to have the positive class (TP+FP) (Singh et al., 2021).

$$Precision = \frac{TP}{TP + FP}$$

3.2.5. Recall

Recall metric is defined as the ratio of correctly classified instances with positive (TP) to total number of instances who have actually positive class(Singh et al., 2021).

$$\mathrm{Recall} = \frac{TP}{TP + FN}$$

3.2.6. F1 score

F1 score is also known as the F Measure. The F1 score states the equilibrium between the precision and the recall (Singh et al., 2021).

$$F1Score = \frac{2 * precision * recall}{precision + recall}$$

3.3. Ethical and societal considerations

Informed consent was obtained from all the participants while conducting this research work. None of the participants were minor or underage. The survey as well as the details of the participating four students in this research work is kept anonymous for privacy reasons. The purpose of the research was restated to all of the participants. Moreover, while conducting the survey, special permission was taken from the departmental director of the university.

This research will assist students and their teachers/parents to develop a better understanding of determining the impact of mental stress in an academic environment. Teachers/parents would be able to provide extra support to their students that would give these students, irrespective of their gender, a fair chance to achieve success.

CHAPTER 4. DATASET FEATURES AND ANALYSIS

The dataset was compiled by gathering the responses of 298 students. It has seven distinct features which are as discussed in this chapter.

4.1. PSS-Score

Figure 9 depicts the perceived stress of the students. This feature is based on score obtained from a Perceived Stress Scale (PSS) questionnaire.

As mentioned in the chapter 3, following the criteria given by Perceived Stress Scale, scores ranging from 0-13, 14 – 26 and 27 – 46 were considered as low stress level, medium stress level and high stress level, respectively. In the dataset, the three levels of stress i.e., low stress, moderate stress and high stress, were codified in three numerical classes i.e., 0, 1 and 2, respectively.

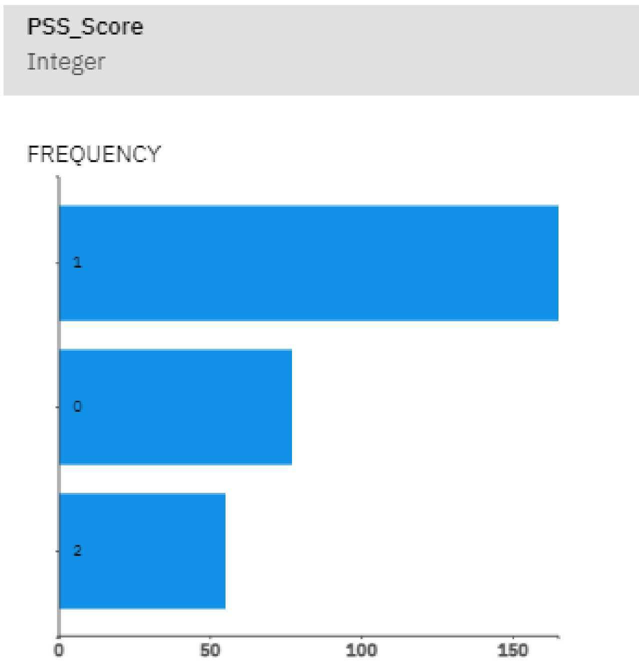

Fig. 9 PSS score calculated for all the 298 students. Among these, 70 students had no stress, 171 students had medium stress level and 57 students had high stress level

4.2. Cognitive performance

Cognitive ability refers to the capacity of the human brain to process, store and retrieve information. It also refers to innate functions of the brain which include attention, memory and reasoning ability. According to Sternberg and Sternberg (2009), it is the essential psychological component for people to successfully complete an activity.

In order to measure the cognitive performance level of the students, a cognitive performance assessment scale was utilized. This scale is based on 25 questions. “The most straightforward way to score the scale is simply to add up the ratings of the 25 individual questions, yielding a score from 0-100. Scores on the scale predict episodes of absent-mindedness in both the laboratory and everyday life, including slow performance on focused attention tasks, traffic and work accidents, and forgetting to save one's data on the computer.” (UC Berkeley, 2011).

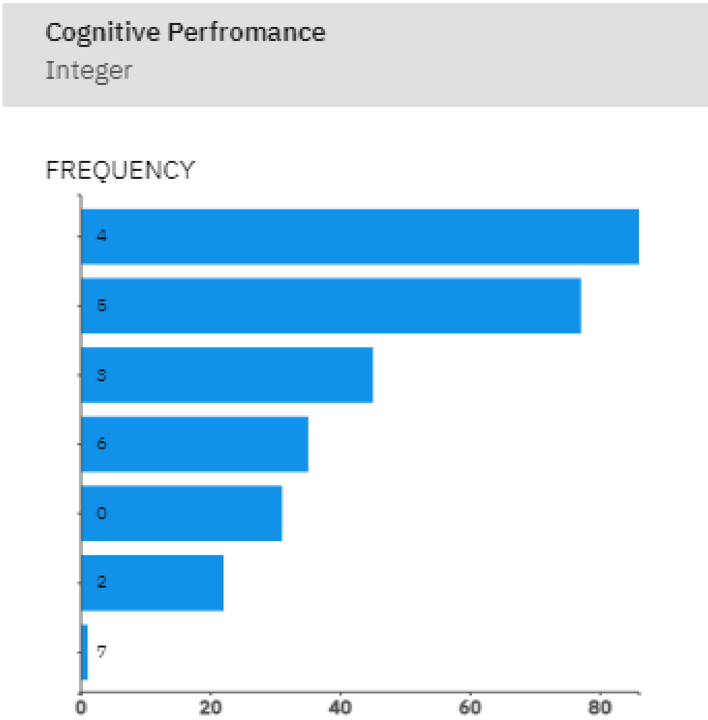

Fig. 10 Cognitive performance indicator

Figure 10 shows the cognitive performance score of the students. This score (0-100) has been codified ranging from a scale of 0 -10, with 0 being the lowest (worst) score and 10 being the highest (best). On vertical axis we can observe the cognitive performance score of the students whereas on the horizontal axis the number of students.

4.3. Gender

Both male and female students were considered in the survey. Male participants were assigned a “0” coded value, whereas, the female participants were represented by “1” as shown in figure 11.

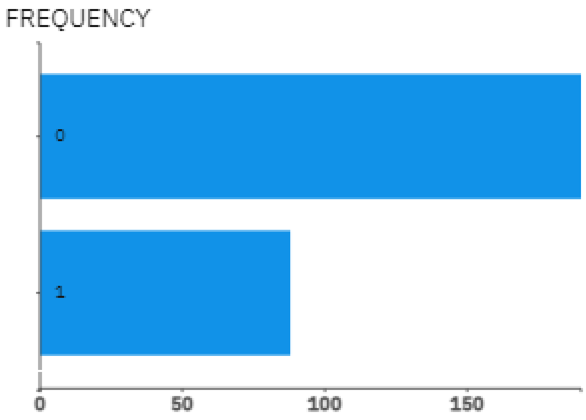

Fig. 11 Gender classification of the students

4.4. Age

This feature represents the age group of the students taking into account only the number of years. The age of the students was not codified.

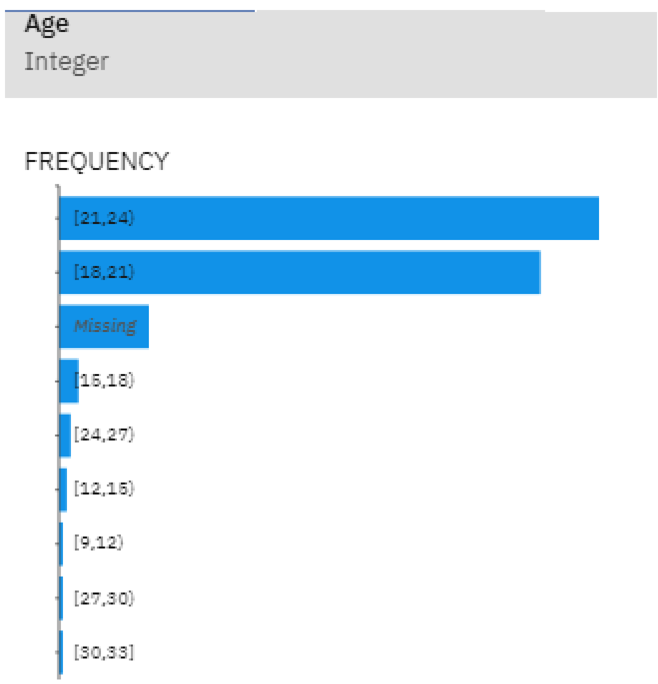

Fig. 12 Age of students in the survey

As shown in the figure 12, frequently the students falls in the age group of 18 to 21 years. There were also a small number of students who didn't specify their age, depicted as missing in the figure. The missing field were left blank in the dataset on which the model was trained.

4.5. Self-study hours

This feature depicts the total number of hours spent by any student on his/her self-study at home or on campus.
As shown in figure 13, approximately 120 students spent between 2 to 4 hours a day, studying by themselves doing homework and other academic related tasks. About 45 students preferred not to answer this question, depicted as missing in the figure. The missing field were left blank in the dataset on which the model was trained.

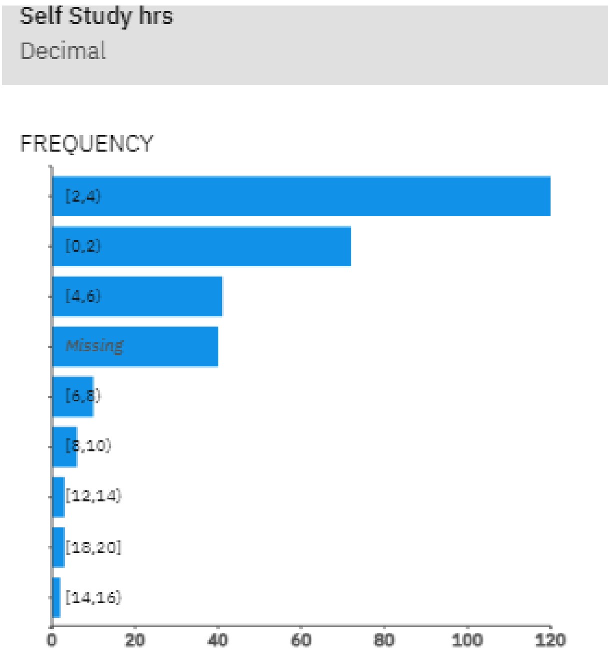

Fig. 13 Average number of hours spent by students, each day, on self-study

4.6. Number of absentees

This feature represents the number of days any student was absent from the class (i.e., absent for the whole working day) during the past three months in the on-going semester.

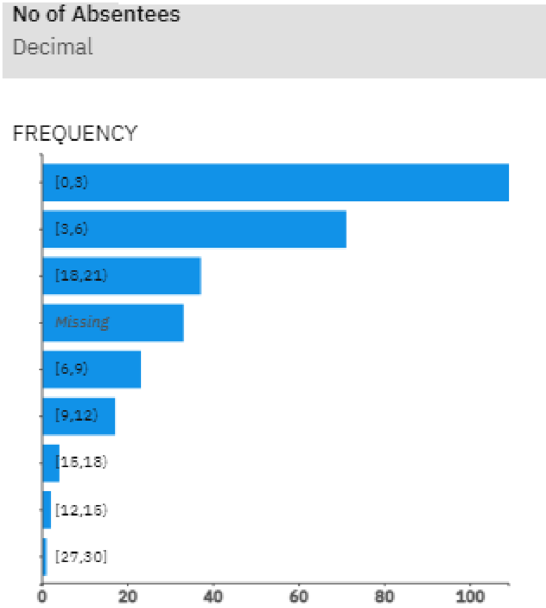

Fig. 14 Absence of students during the last three months

As depicted in figure 14, about 110 students skipped 0 to 3 lecture days, being the most frequent. About 38 students preferred not to answer this question, shown as missing.

4.7. Average Grade

The performance of the students has been analysed through this feature. Students were asked to mention their scores in terms of percentage in their last three graded class activities. Then the average of these three graded class activities was taken and classified in three coded values i.e., 0, 1 and 2, corresponding to percentage score falling in three categories i.e., 0-50 %, 51-70% and 71- 100 %, respectively, representing 0 the lowest and 2 the best academic performance.

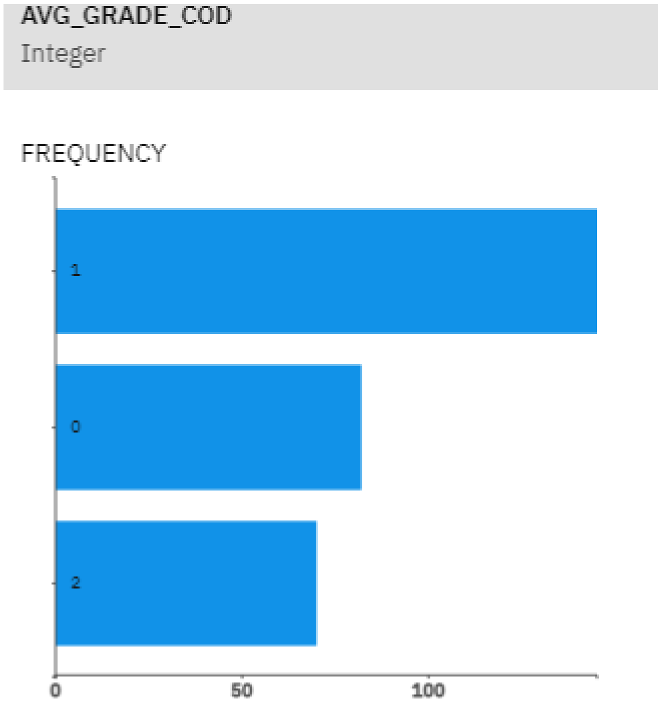

Fig. 15 Average grade of students participating in the survey

As shown in figure 15, 170 students had an normal performance in the last three graded activities with the score falling between 51-70%, thus, classified as 1. Around 66 students had poor performance with a score falling between 0-50 %, classified as 0. Approximately, 62 students had a good score with average grade between 71- 100 %.

CHAPTER 5. MODEL IMPLEMENTATION AND CONFIGURATION

In this study, IBM Watson Studio has been utilized to build and deploy machine learning models. Watson Studio empowers data scientists, developers and analysts to build, run and manage AI models, and optimize decisions anywhere on IBM Cloud Pak for Data.

The following steps are to be followed to develop a machine learning model on the IBM Watson Machine Learning Platform.

- In the first step, Watson Studio is deployed in the IBM cloud.
- Two resources i.e., IBM Cloud Object Storage and IBM Machine Learning service instance are created while being linked to the Watson studio space.
- Afterwards, the user creates a new project space in the Watson studio.
- Upload the data set in .csv format so the data can be fed to the ML model.
- Create the ML experiment that will automatically analyse the given tabular data and will generate the candidate model pipelines customized for the predictive modelling.
- Selection of output variable that we need the model to predict. As this study consisted of building two distinct models, thus, two out variables were considered, i.e., average grade and stress level. The output variable i.e., average grade has a multiclass with 0, 1 and 2, corresponding to percentage score falling in three categories i.e., 0-50 %, 51-70% and 71-100 %, respectively.
- 90 percent of the dataset was used to train the model whereas the 10% data was used to test the model accuracy.
- Choice of selecting algorithms either by the user or by the machine. IBM Watson machine gives the option to choose from eleven different algorithms which are appropriate for any kind of data and may include multiple variables under investigation.
- Run the experiment that will make the progress map available for review. The progress map includes several important steps including reading the data set, splitting the holdup data (10% of the total), reading the remaining 90% data, pre-processing, model selection, and then it moves to selected algorithms for further processing.
- Once the experiment completes, it will provide multiple tables showing the feature ranking, performance measures, confusion matrix, etc,.

CHAPTER 6. RESULTS

The dataset, compiled from both the survey based on the questionnaire and the Apple smart watch was analysed, separately, using IBM Watson ML experiment. The study was divided in two phases as described in the following sections

6.1. First Phase: Results Survey Dataset

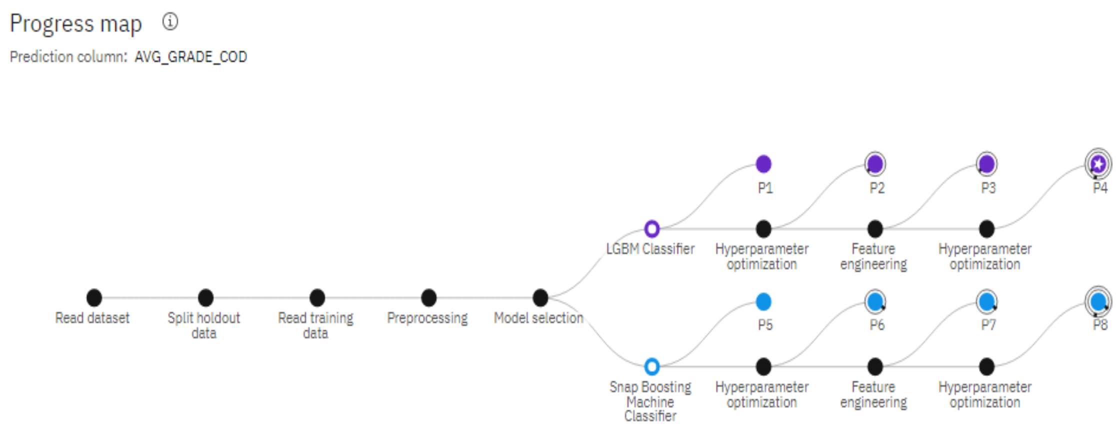

Fig. 16 Progress map showing different stages in IBM ML experiment. respectively.

Figure 16 shows the detailed progress map of the ML model building using IBM Watson studio. The experiment consists of various steps as shown in the figure. In the first step, the dataset is read by the machine. Then the dataset is split into training and testing data. Afterwards, the data is cleaned in the Pre-processing stage. In the last step, model that is best suited for the dataset, in terms of accuracy, is selected. The users have the option to set the numbers of models that they want to observe in the output. These models are fine-tuned by applying feature engineering and hyperparameter optimization.

The two models that gave the best performance measure were chosen, namely, LGBM classifier and Snap Boosting Machine Classifier, with the Cross Validation accuracy of 71.4% and 71.0%. Feature engineering and Hyperparameter optimization was applied to further improve the accuracy of the two models, thus generating different pipelines as shown in figure 16.

6.1.2. LGBM classifier feature ranking

Table 2 shows the ranking based on importance for various features of the dataset. As recommended by the IBM machine, any feature with the importance above 65% should be considered while the feature having importance below this threshold may be neglected for the optimum results.

Table 2. Feature ranking using LGBM classifier

Feature name	Transformation	Feature importance
No of Absentees	None	100.00%
PSS_Score	None	99.00%
Cognitive Perfromance	None	86.00%
Age	None	65.00%
NewFeature_1	log(No of Absentees)	52.00%
Self Study hrs	None	48.00%
NewFeature_0	log(Self Study hrs)	32.00%
NewFeature_3	round(No of Absentees)	23.00%
NewFeature_2	round(Self Study hrs)	13.00%
Gender	None	8.00%

As depicted in the Table 2, the four most important features using the LGBM classifier comes out to be, the total number of days any student skipped the class lectures, the perceived stress scale score, the cognitive performance and the age of the students. This shows us that stress level of the students, with a feature importance of 99%, has an important role in predicting the average grades of the students.

6.1.3. LGBM classifier model measures

Table 3, shows various model measures of the LGBM classifier. The accuracy of the LGBM classifier is 71.4%, impacted by the size of dataset and missing values in the dataset, could be improved further by increasing the dataset size.

Table 3. LGBM classifier model measures The Cross Validation accuracy score of the LGBM classifier comes out to be 71.4%

Measures	Cross validation score
Precision macro	0.708
Accuracy	0.714
Recall macro	0.735
Weighted precision	0.726
F1 macro	0.713
Weighted f1 measure	0.710
Weighted recall	0.714
Log loss	0.718

As depicted in table 3, we have a multi-class classification, consequently, IBM adopted the averaging methods for F1 score calculation, resulting in a set of different average scores i.e., weighted and macro F1 measures (Leung, 2022).

By taking the arithmetic mean of all the per-class F1 scores, macro F1 score is calculated. This method treats all classes equally regardless of their support values (Leung, 2022). Since, we have an imbalanced dataset, hence, the F1 macro score of the model should be given importance, which in our case comes out to be 0.713, signifying above average performance of the model.

6.1.4. LGBM classifier confusion matrix

The accuracy of the model is derived from the Confusion Matrix. Tables 4, 5 and 5 shows the confusion matrix of the stress levels of one class compared with the other two classes.

Table 4. Confusion matrix predicting “0” stress level (no stress) against the other two levels i.e., “1” (medium stress) and “2” (high stress)

Confusion matrix

View

0 (One v. Rest)

Observed	Predicted		
	0	Not 0	Percent correct
0	5	3	62.5%
Not 0	5	16	76.2%
Percent correct	50.0%	84.2%	72.4%

From table 4, it can be observed that the overall accuracy of predicting “0” stress level is 72.4%, with True Positives (i.e., individuals who had no stress and were correctly identified by the ML model) are equal to 5 and False Negatives (i.e., individuals who had stress and were incorrectly identified as not having stress by the ML model) are also equal to 5. Similarly, the True Negatives (i.e., individuals who had stress and were correctly identified by the ML model) are equal to 16 and False Positives (i.e., individuals who had no stress and were incorrectly identified as having stress by the ML model) are also equal to 3.

Table 5. Confusion matrix predicting "1" stress level (medium) against the other two levels i.e., "0" (no stress) and "2" (high stress)

Confusion matrix

View

1 (One v. Rest)

Observed	Predicted		
	1	Not 1	Percent correct
1	8	6	57.1%
Not 1	4	11	73.3%
Percent correct	66.7%	64.7%	65.5%

From table 5, it can be observed that the overall accuracy of predicting "1" stress level is 65.5%, compared with the "0" (no stress) and "2" (high stress) output classes.

Table 6. Confusion matrix predicting "1" stress level (medium) against the other two levels i.e., "0" (no stress) and "2" (high stress)

Confusion matrix

View

2 (One v. Rest)

Observed	Predicted		
	2	Not 2	Percent correct
2	4	3	57.1%
Not 2	3	19	86.4%
Percent correct	57.1%	86.4%	79.3%

Table 6 depicts that the overall accuracy of predicting "2" (high stress level) is 79.3%, compared with the "0" (no stress) and "1" (medium stress) output classes.

6.1.5. Snap Boosting Classifier Feature Ranking

The ranking based on importance of the features, while using Snap Boosting classifier, is shown in Table 7.

Table 7. Feature ranking using Snap Boosting classifier

Feature summary

High correlation

All features

Search feature or transformer names

Feature name	Transformation	Feature importance
Cognitive Perfromance	None	100.00%
PSS_Score	None	16.00%
Age	None	13.00%
NewFeature_0 Most improved	log(Self Study hrs)	8.00%
NewFeature_1	log(No of Absentees)	8.00%
No of Absentees	None	6.00%
Self Study hrs	None	5.00%

The most important features using the Snap Boosting classifier comes out to be, the cognitive performance of the students. The importance of the perceived stress in this case is only 16%, which implies that this models is not suitable for predicting the average grades of the students based on their stress level.

6.1.6. Snap Boosting Classifier Model Measures

Table 8 shows the performance parameters of the Snap Boosting model.

Table 8. Snap Boosting classifier model measures. The Cross Validation accuracy score comes out to be 71.0%

Measures	Cross validation score
Precision macro	0.708
Accuracy	0.710
Recall macro	0.713
Weighted precision	0.720
F1 macro	0.703
Weighted f1 measure	0.708
Weighted recall	0.710
Log loss	0.752

From table 8, it can be deduced that Snap Boosting model has more or less similar performance measures as compared to LGBM classifier model. But, this model is not recommended to predict the grades of the students based on their stress levels as it shows high correlation for cognitive performance, only.

6.1.7. Snap Boosting classifier confusion matrix

As depicted in table 9, the Hold-out accuracy of the Snap Boosting classifier is 69.0%.

Table 9. Confusion matrix of Snap Boosting classifier

Observed	Predicted			
	0	1	2	Percent correct
0	6	1	1	75.0%
1	1	11	3	73.3%
2	1	2	3	50.0%
Percent correct	75.0%	78.6%	42.9%	69.0%

6.2. Second phase: mental stress measurements based on HRV

Using the modified version of the SWELL dataset, a model was built and deployed on the IBM Watson studio. The top performing model in this case came out to be Decision Tree Classifier.

Table 10. Confusion matrix of Decision Tree classifier

Confusion matrix

Observed	Predicted		
	no stress	stress	Percent correct
no stress	3290	34	99.0%
stress	56	2775	98.0%
Percent correct	98.3%	98.8%	98.5%

Less correct — More correct

Table 10, depicts the confusion matrix of the Decision Tree classifier. It can be seen that the accuracy of the classifier comes out to be 98.5% which proves that the model is best suited for predicting the stress level of the individuals using HRV parameters.

Table 11. Model evaluation measures of the Decision Tree classifier

Model evaluation measure

Measures	Holdout score	Cross validation score
Accuracy	0.985	0.981
Area under ROC	0.985	0.980
Precision	0.983	0.982
Recall	0.990	0.982
F1	0.987	0.982
Average precision	0.535	0.974
Log loss	0.505	0.671

Table 11 shows the model evaluation measures in terms of performance. Both holdout and cross validation scores are shown. The holdout and cross validation accuracies are 98.5% and 98.1%, respectively. The model has a F1 scores of 0.987 and 0.982 corresponding to the holdout score and cross validation score, respectively, which signifies that the model has an excellent performance measures in terms of precision and recall, thus, resulting in a well-balanced model.

Table 12. Feature ranking of Decision Tree classifier model. The features of the dataset are ranked according to their corresponding correlation in predicting the

Feature summary

High correlation

All features

Search feature or transformer names

Feature name	Transformation	Feature importance
MEDIAN_RR	None	100.00%
pNN25	None	34.00%
SDRR_RMSSD	None	32.00%
SDRR	None	15.00%
SDSD	None	13.00%
MEAN_RR	None	3.00%
RMSSD	None	1.00%

As shown in table 12, the most important feature having the highest correlation in predicting the stress is median of RR intervals between two heartbeats.

6.2.1 Testing the Decision Tree classifier model with HRV parameters

After developing this model, the next step was to gather the data related to HRV of four students, in order to confirm the viability of using this model for HRV measured through the Apple smart watch.

In parallel to measuring the stress levels of the students with the Apple smart watch, the stress score of the students was also measured with the Perceived Stress Scale.

Table 13. the PSS score, cognitive assessment (CA) score, age and grades achieved i.e., performance in the exams of the four students

	Age	Grade level	CA score	PSS score		
				Sample 1	Sample 2	Sample 3
Student 1	22	1	29	14 (moderate stress)	09 (no stress)	13 (moderate stress)
Student 2	23	1	47	16 (moderate stress)	23 (high stress)	20 (moderate stress)
Student 3	20	1	39	19 (moderate stress)	16 (moderate stress)	20 (moderate stress)
Student 4	21	2	58	26 (high stress)	29 (high stress)	30 (high stress)

For each of the four students, three samples (during the examination time period) after every 3 days were taken with the help of PSS. From table 13, It can be observed that student 4 (an outlier) was constantly under high stress during the examination days, but irrespective of the high stress level, achieved an above average grades in the three exams. This may be due to the high cognitive performance score.

The HRV measurements were taken in two phases. In the first phase, the measurements were taken in order to determine the baseline threshold level of HRV parameters for the four students, whereas, the second phase was implemented during the mid-term examination days.

Table 14. HRV parameters measured during relaxing stage of the four students

Measurements during relax stage				
HRV Parameters	**Student 1**	**Student 2**	**Student 3**	**Student 4**
SDRR	140.97	90.37	62.76	517.53
MEAN_RR	885.15	881.75	809.62	923.28
MEDIAN_RR	853.76	893.4	811.18	617.79
RMSSD	15.55	15.72	19.21	9.96
SDRR_RMSSD	9.063	5.74	3.26	51.93
HR	69.49	68.8	74.56	81.34
pNN25	11.13	11.8	20.2	1.2

The HRV, during the relax stage, of the four students was recorded with the help of the Apple smart watch and the corresponding parameters were calculated as shown in table 14.

Table 15. HRV parameters measured during stress stage

Measurements during stress stage				
HRV Parameters	**Student 1**	**Student 2**	**Student 3**	**Student 4**
SDRR	81.31	84.49	57.88	199.96
MEAN_RR	939.42	898.186	848.62	793.61
MEDIAN_RR	948.35	907.0	851.57	720.74
RMSSD	12.96	16.3	14.14	14.786
SDRR_RMSSD	6.27	5.182	4.093	13.523
HR	64.36	67.45	71.04	79.49
pNN25	5.6	13.06	8.13	8.26

Table 15 illustrates the HRV measurements calculated during the stress stage, of the four students.

Considering the results obtained, in tables 14 and 15, it can be observed that the median value of the RR interval increases during the stress stage of the students whereas the heart rate (HR) decreases.

After calculating the HRV parameters during both the relax and stress stages, these parameters were fed into the previously deployed model on IBM cloud and the model made predictions about the stress level of the students. As depicted in table 16, all the predictions were made correctly.

Table 16. Predictions made by the Decision Tress Classifier model, deployed on the IBM Cloud

Table view JSON view Show input data

	Prediction	Confidence	MEAN_RR	MEDIAN_RR	SDRR	RMSSD	SDRR_RMSSD	HR	pNN25
1	no stress	100%	885.15	853.76	140.97	15.55	9.063	69.49	11.13
2	no stress	100%	881.75	893.4	90.37	15.72	5.74	68.8	11.8
3	no stress	100%	809.62	811.18	62.76	19.21	3.26	74.56	20.2
4	no stress	100%	923.28	617.79	517.53	9.96	51.93	81.34	1.2
5	stress	100%	939.42	948.35	81.31	12.96	6.27	64.36	5.6
6	stress	100%	898.186	907.0	84.49	16.3	5.18	67.45	13.06
7	stress	100%	848.62	851.57	57.88	14.14	4.093	71.04	8.13
8	stress	100%	793.61	720.74	199.96	14.78	13.523	79.49	8.26

Similarly, the HRV recording was taken into account before and during the mid-term examination of the four students as shown in tables 17 and 18, respectively.

Table 17. HRV measurements during the examination days of the students (measurements taken before the start of the exam)

Measurements (before exam)				
HRV Parameters	**Student 1**	**Student 2**	**Student 3**	**Student 4**
SDRR	63.96	70.86	65.096	119.70
MEAN_RR	802.85	704.35	958.968	738.15
MEDIAN_RR	804.29	695.00	958.817	725.53
RMSSD	14.2348	20.419	11.787	13.936
SDRR_RMSSD	4.4938	3.470	5.522	8.589
HR	75.22	86.01	62.86	83.06
pNN25	6.8	23.26	4.0	6.2

Table 18. HRV measurements during the examination days of the students (measurements taken while the students were taking their exam)

Measurements during exam				
HRV Parameters	**Student 1**	**Student 2**	**Student 3**	**Student 4**
SDRR	385.751	124.941	141.86	32.84
MEAN_RR	1017.31	696.782	1028.88	792.47
MEDIAN_RR	970.72	710.4	1020.40	796.62
RMSSD	15.868	15.723	15.2304	7.556
SDRR_RMSSD	24.30	7.946	9.31422	4.347
HR	68.096	89.159	59.452	75.84
pNN25	10.466	8.06	10.4	0.4

After calculating the HRV parameters during both before and during exam stages, these parameters were fed into the previously deployed model on IBM cloud and the model made predictions about the presence of mental stress in the students.

Table 19. Predictions made by the Decision Tress Classifier

Table view JSON view Show input data

	Prediction	Confidence	MEAN_RR	MEDIAN_RR	SDRR	RMSSD	SDRR_RMSSD	HR	pNN25
1	no stress	100%	802.85	804.29	63.96	14.2348	4.49	75.22	6.8
2	no stress	100%	704.35	695.00	70.86	20.419	3.470	86.01	23.26
3	stress	100%	958.968	958.817	65.096	11.787	5.522	62.86	4.0
4	no stress	100%	738.15	725.53	119.70	13.936	8.589	83.06	6.2
5	stress	100%	1017.31	970.72	385.751	15.868	24.30	68.096	10.46
6	stress	100%	696.782	710.4	124.941	15.723	7.946	89.159	8.06
7	stress	100%	1028.88	1020.40	141.86	15.23	9.314	59.452	10.4
8	no stress	100%	792.47	796.62	32.84	7.556	4.347	75.84	0.4

From table 13, we observed that using the PSS scale, it was determined that student 4 was constantly under high stress during the examination days, but observing the prediction made in table 19, for student 4, the DT classifier model predicted “no stress” (even though the student was undertaking the exam).

The correlation between the physiological stress measurement and the subjective perception of stress could be further explored by including more test subjects under study.

CHAPTER 7. DISCUSSION

7.1. Analysis and discussion of results

The first objective of this research work was to determine the feasibility of using ML algorithms by establishing a correlation between students' academic performance and their mental stress level. To this extent this research seems to be effective as the two classifier ML models that are developed tends to achieve an accuracy of 71% approximately. The low accuracy is due small size of dataset. The accuracy could be enhanced, significantly, taking into account a larger dataset. Furthermore, when we consider the feature ranking of the two ML classifiers, it can be observed that Perceived Stress Score (PSS) feature plays a significant role in predicting the academic performance of the students, considering the LGBM classifier. The higher the PSS score, the lower the academic performance of the students.

The second research question was to explore the benefit of using IBM Watson ML platform. Implementing this research by utilizing IBM Watson ML cloud based platform proved to be a very viable and powerful solution as one can easily utilize the automatic experimental approach of the IBM ML without the need of extensive coding. In this regard, steps to build and implement the ML models using IBM auto AI experiment has also been mentioned in this thesis. Top performing ML model was chosen on the basis of accuracy in addition to various parameters like, F1 Score, Precision, Recall, etc.

The third research objective was to the implementation of a ML model to predict the mental stress using Heart Rate Variability. In this context, several research papers and reference sources were considered to determine the parameters of Heart Rate Variability (HRV). There are three main classifications of HRV parameters, namely, time domain, frequency domain and non-linear parameters. In this research only the time domain parameters are considered and a ML model has been developed utilizing the time domain parameters as features for the dataset. The ML model was trained and tested using the SWELL dataset.

The SWELL dataset was compiled by retrieving information through an experimental setup (Koldijk et al., 2014). The experimental setup included twenty five individuals, and they were monitored while performing usual work related tasks that included making presentations, report writing, browsing information and responding to e-mails (Koldijk et al., 2014). The working environment of these 25 individuals were controlled with interruptions, time pressure and various stresses (Koldijk et al., 2014). The data was compiled based on monitoring the facial expressions from camera recordings, computer logging, body movements from Kinect 3D sensor, and skin conductance from body sensors (Koldijk et al., 2014). As per the researchers, "the participants' subjective experience on work load, mental efforts and emotional stress was assessed with pre documented questionnaires as a central focus for ground truth. The resulting dataset on working behaviour and affect is a valuable contribution to several research fields, such as work psychology, user modelling and context aware systems." (Koldijk et al., 2014).

The original SWELL dataset was altered by taking into account only the HRV time domain parameters as mentioned in the previous section. The top performing model in our case came out to be Decision Tree Classifier with a Cross Validation accuracy of 98.1%.

After developing this model, the next step was to gather the data related to HRV of four students, in order to confirm the viability of using this model for HRV measured through Apple smart watch. As far as the question of how accurately the smart watch measures the HRV is concerned, the results of this analysis as described in the previous chapter, confirms the suitability of using Apple smart watch.

Furthermore, it was deduced that whenever the mean value of the RR interval increases, the heart rate decrease and consequently the stress level increases.
Additionally, it was observed that the physiological stress measurement can deviate from subjective perception of stress. In another study it was revealed that there might be a genetic moderation in the association between resting state HRV and perceived stress (Looser et al., 2023).

CHAPTER 8. CONCLUSIONS

This research work utilized a quantitative approach to analyse the impact of stress on heart rate variability (HRV) using machine learning techniques. The findings revealed a positive association between higher stress levels and increased HRV, consequently, degrading the academic performance of the students as proved from the first phase of this study. In this context, Snap Boosting model had more or less similar performance measures as compared to LGBM classifier model. But, this model is not recommended to predict the grades of the students based on their stress levels as it shows high correlation for only cognitive performance.

The stress level was determined by extracting features from HRV analysis, and a classification technique was employed using threshold values derived from the training dataset. Performance accuracy measures were employed to assess the outcomes. Consequently, this study suggests that stress influences HRV, thus establishing its potential as an objective tool for assessing stress in academic settings.

8.1. Future work

Following are some of the recommendations given in regard to the future scope of this study;

- The accuracy of the model, developed to predict the academic performance of the students, could be further enhanced if the size of the dataset is increased i.e., conducting survey on a larger scale.
- The dataset that was gathered with the help of a survey based on responses from the students, seems to be less efficient. Steps can be

taken to reduce the human error due to negligence of the students taking part in the survey.

- The correlation between the physiological stress measurement and the subjective perception of stress can be further explored by including more test subjects under study.
- A smart Learning Management System (LMS) can be designed that can give smart recommendations to the students on ways to improve their academic performance & to cope up with the mental stress suggestions like breathing exercises etc. Data mining could be further applied while considering solutions for students with different performance/stress levels. Figure 17 shows proposed structure of a smart LMS database.

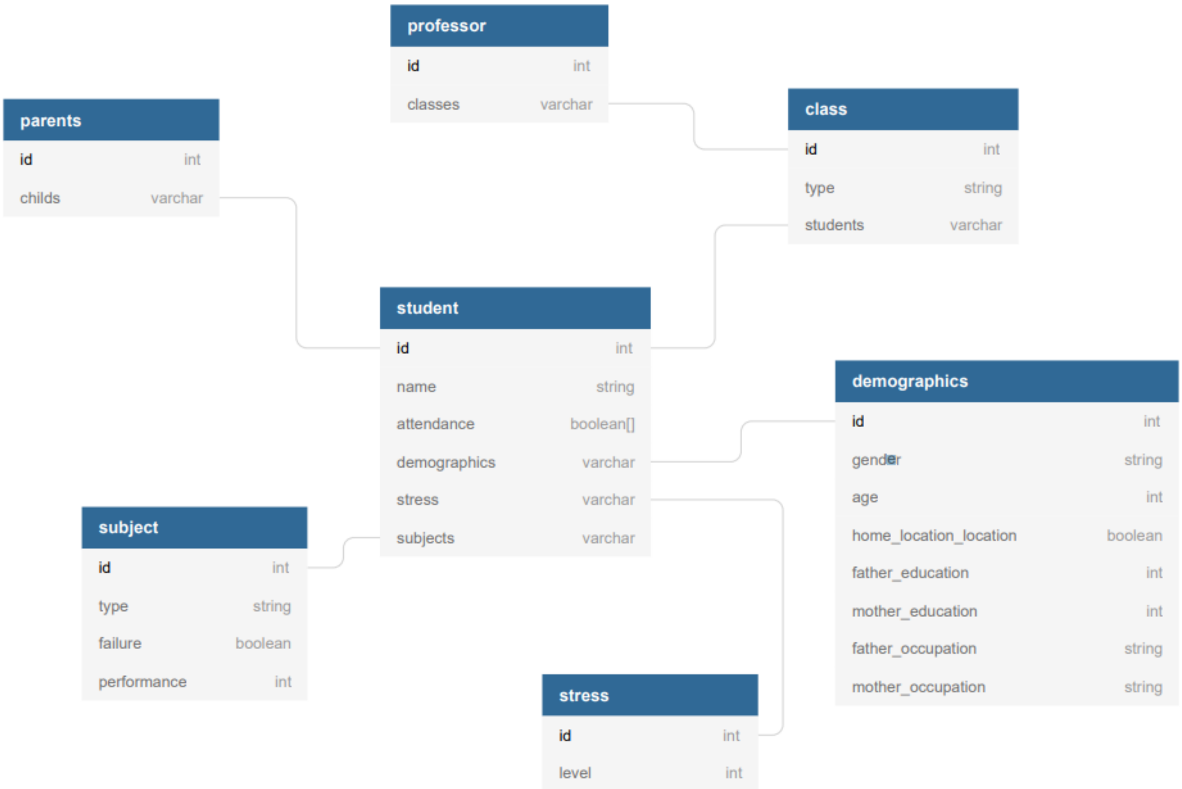

Fig. 17 Proposed Database structure for LMS

References

Analysis of Autonomic Nervous System. (n.d.). HRV - measuring parameter - ans analysis. Retrieved April 3, 2023, from https://www.ans-analysis.com/hrv/hrv-measuring-parameter.html

Colom, R., Karama, S., Jung, R. E., & Haier, R. J. (2010). *Human intelligence and brain networks*. Dialogues in clinical neuroscience. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3181994/

Deng, Y., Cherian, J., Khan, N. U. N., Kumari, K., Sial, M. S., Comite, U., Gavurova, B., & Popp, J. (2022). Family and Academic Stress and Their Impact on Students' Depression Level and Academic Performance. Frontiers in psychiatry, 13, 869337. https://doi.org/10.3389/fpsyt.2022.869337

Dinardo, D. A. (2016, October). *What's Your stress threshold?* Thriving Under Pressure. Retrieved from https://drandreadinardo.com/2016/10/05/whats-your-stress-threshold/

Hazer-Rau, D., Meudt, S., Daucher, A., Spohrs, J., Hoffmann, H., Schwenker, F., & Traue, H. C. (2020, April 17). *The UULMMAC database-a multimodal affective corpus for affective computing in human-computer interaction*. MDPI. Retrieved from https://www.mdpi.com/1424-8220/20/8/2308

Hernando, D., Roca, S., Sancho, J., Alesanco, Á., & Bailón, R. (2018, August 10). *Validation of the apple watch for heart rate variability measurements during relax and mental stress in healthy subjects*. MDPI. Retrieved from https://www.mdpi.com/1424-8220/18/8/2619/htm

IBM Watson. (n.d.). *IBM Watson Studio - Overview*. IBM. Retrieved April 3, 2023, from https://www.ibm.com/cloud/watson-studio

IBM. (2023). *What is a decision tree?*. Retrieved April 3, 2023, from https://www.ibm.com/topics/decision-trees#:~:text=A%20decision%20tree%20is%20a,internal%20nodes%20and%20leaf%20nodes

What is supervised learning? IBM. (n.d.). Retrieved April 6, 2023, from https://www.ibm.com/topics/supervised-learning

IBM. (2022). *What is boosting?* Retrieved from https://www.ibm.com/topics/boosting

Kiselica, M. S., Baker, S. B., Thomas, R. N., & Reedy, S. (1994). Effects of stress inoculation training on anxiety, stress, and academic performance among adolescents. *Journal of Counseling Psychology*, *41*(3), 335–342. https://doi.org/10.1037/0022-0167.41.3.335

Koldijk, S., Sappelli, M., Verberne, S., Neerincx, M. A., & Kraaij, W. (2014). The swell knowledge work dataset for stress and User Modeling Research. *Proceedings of the 16th International Conference on Multimodal Interaction*. https://doi.org/10.1145/2663204.2663257

Kim, H.-G., Cheon, E.-J., Bai, D.-S., Lee, Y. H., & Koo, B.-H. (2018, February 28). *Stress and heart rate variability: A meta-analysis and review of the literature*. Psychiatry Investigation. Retrieved from https://www.psychiatryinvestigation.org/journal/view.php?doi=10.30773%2Fpi.2017.08.17

Kurtis Pykes - The Difference Between Classification and Regression in Machine Learning, Published in *Towards Data Science, Aug, 7, 2021*

Kim, Y., Yoon, H. Y., Kwon, I. K., Youn, I., & Han, S. (2022). Heart rate variability as a potential indicator of cancer pain in a mouse model of peritoneal metastasis. *Sensors*, *22*(6), 2152. https://doi.org/10.3390/s22062152

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., & Liu, T.-Y. (2017, December 4). *LightGBM: A highly efficient gradient boosting decision tree*. ACM. https://dl.acm.org/doi/10.5555/3294996.3295074#sec-ref

Lin, X.-J., Zhang, C.-Y., Yang, S., Hsu, M.-L., Cheng, H., Chen, J., & Yu, H. (2020, June 3). *Stress and its association with academic performance among dental undergraduate students in Fujian, China: A cross-sectional online questionnaire survey - BMC Medical Education*. BioMed Central. Retrieved April 3, 2023, from https://bmcmededuc.biomedcentral.com/articles/10.1186/s12909-020-02095-4

Leung, K. (2022, September 13). *Micro, Macro & weighted averages of F1 score, clearly explained*. Medium. https://towardsdatascience.com/micro-macro-weighted-averages-of-f1-score-clearly-explained-b603420b292f#:~:text=The%20macro%2Daveraged%20F1%20score,regardless%20of%20their%20support%20values.&text=The%20value%20of%200.58%20we,score%20in%20our%20classification%20report

Looser, V. N., Ludyga, S., & Gerber, M. (2023). Does heart rate variability mediate the association between chronic stress, Cardiorespiratory Fitness, and working memory in young adults? *Scandinavian Journal of*

Medicine & Science in Sports, *33*(5), 609–618. https://doi.org/10.1111/sms.14308

Metabolic Meals. (2021, June 20). *A beginner's Guide to Heart Rate Variability (HRV) - metabolic meals*. Blog. Retrieved April 3, 2023, from https://blog.mymetabolicmeals.com/hrv-guide/

Microsoft Corporation. (n.d.). *LightGBM*. Features - LightGBM 3.3.5.99. Retrieved April 3, 2023, from https://lightgbm.readthedocs.io/en/latest/Features.html#references

NH Dept. of Administrative Services . (n.d.). *Perceived stress scale - new hampshire*. Retrieved April 3, 2023, from https://www.das.nh.gov/wellness/Docs/Percieved%20Stress%20Scale.pdf

Padilha, T. P. P., & Catrambone, R. (2021). USE OF THE TENSORFLOW FRAMEWORK TO SUPPORT EDUCATIONAL PROBLEMS: A SYSTEMATIC MAPPING. *International Conference on Education*. https://doi.org/10.17501/24246700.2021.7133

Rajendran, S., Chamundeswari, S., & Sinha, A. A. (2022). Predicting the academic performance of middle- and high-school students using machine learning algorithms. *Social Sciences & Humanities Open*, *6*(1), 100357. https://doi.org/10.1016/j.ssaho.2022.100357

Singh, P., Singh, N., Singh, K. K., & Singh, A. (2021). Diagnosing of disease using machine learning. *Machine Learning and the Internet of Medical Things in Healthcare*, 89–111. https://doi.org/10.1016/b978-0-12-821229-5.00003-3

Shi, Y., & Qu, S. (2022). The effect of cognitive ability on academic achievement: The mediating role of self-discipline and the moderating role of planning. *Frontiers in Psychology*, *13*. https://doi.org/10.3389/fpsyg.2022.1014655

Shaffer, F., & Ginsberg, J. P. (2017, September 28). *An overview of heart rate variability metrics and norms*. Frontiers in public health. Retrieved April 3, 2023, from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5624990/

Singh, N., Moneghetti, K. J., Christle, J. W., Hadley, D., Plews, D., & Froelicher, V. (2018). Heart Rate Variability: An Old Metric with New Meaning in the Era of using mHealth Technologies for Health and Exercise Training Guidance. Part One: Physiology and Methods. *Arrhythmia & electrophysiology review*, *7*(3), 193–198. https://doi.org/10.15420/aer.2018.27.2

Suresh, A. (2021, June 22). *What is a confusion matrix?*. Medium. https://medium.com/analytics-vidhya/what-is-a-confusion-matrix-d1c0f8feda5

Welltory, W. I. (2023, March 18). *Stress and HRV - what's The connection: Welltory app*. Welltory. Retrieved April 3, 2023, from https://welltory.com/stress-and-hrv/

Williams, S., Layard Horsfall, H., Funnell, J. P., Hanrahan, J. G., Khan, D. Z., Muirhead, W., Stoyanov, D., & Marcus, H. J. (2021). Artificial Intelligence in Brain Tumour Surgery-An Emerging Paradigm. *Cancers*, *13*(19), 5010. https://doi.org/10.3390/cancers13195010

Xu, J., Moon, K. H., & van der Schaar, M. (2017). A machine learning approach for tracking and predicting student performance in degree programs.

IEEE Journal of Selected Topics in Signal Processing, *11*(5), 742–753. https://doi.org/10.1109/jstsp.2017.2692560

UC Berkeley . (2011). *Cognitive failures questionnaire - university of California, Berkeley*. Open Computing Facility . Retrieved from https://www.ocf.berkeley.edu/~jfkihlstrom/ConsciousnessWeb/Meditation/CFQ.htm

Annex

Perceived Stress Scale

A more precise measure of personal stress can be determined by using a variety of instruments that have been designed to help measure individual stress levels. The first of these is called the **Perceived Stress Scale**.

The Perceived Stress Scale (PSS) is a classic stress assessment instrument. The tool, while originally developed in 1983, remains a popular choice for helping us understand how different situations affect our feelings and our perceived stress. The questions in this scale ask about your feelings and thoughts during the last month. In each case, you will be asked to indicate how often you felt or thought a certain way. Although some of the questions are similar, there are differences between them and you should treat each one as a separate question. The best approach is to answer fairly quickly. That is, don't try to count up the number of times you felt a particular way; rather indicate the alternative that seems like a reasonable estimate.

For each question choose from the following alternatives:
0 - never 1 - almost never 2 - sometimes 3 - fairly often 4 - very often

________ 1. In the last month, how often have you been upset because of something that happened unexpectedly?

________ 2. In the last month, how often have you felt that you were unable to control the important things in your life?

________ 3. In the last month, how often have you felt nervous and stressed?

________ 4. In the last month, how often have you felt confident about your ability to handle your personal problems?

________ 5. In the last month, how often have you felt that things were going your way?

________ 6. In the last month, how often have you found that you could not cope with all the things that you had to do?

________ 7. In the last month, how often have you been able to control irritations in your life?

________ 8. In the last month, how often have you felt that you were on top of things?

________ 9. In the last month, how often have you been angered because of things that happened that were outside of your control?

________ 10. In the last month, how often have you felt difficulties were piling up so high that you could not overcome them?

Figuring Your PSS Score

You can determine your PSS score by following these directions:

- First, reverse your scores for questions 4, 5, 7, and 8. On these 4 questions, change the scores like this:

  0 = 4, 1 = 3, 2 = 2, 3 = 1, 4 = 0.

- Now add up your scores for each item to get a total. **My total score is ___________.**

- Individual scores on the PSS can range from 0 to 40 with higher scores indicating higher perceived stress.
  - ► Scores ranging from 0-13 would be considered low stress.
  - ► Scores ranging from 14-26 would be considered moderate stress.
  - ► Scores ranging from 27-40 would be considered high perceived stress.

The Perceived Stress Scale is interesting and important because your perception of what is happening in your life is most important. Consider the idea that two individuals could have the exact same events and experiences in their lives for the past month. Depending on their perception, total score could put one of those individuals in the low stress category and the total score could put the second person in the high stress category.

***Disclaimer**: The scores on the following self-assessment do not reflect any particular diagnosis or course of treatment. They are meant as a tool to help assess your level of stress. If you have any further concerns about your current well being, you may contact EAP and talk confidentially to one of our specialists.*

State of New Hampshire
Employee Assistance Program

The Cognitive Assessment Questionnaire

The following questions are about minor mistakes, which everyone makes from time to time, but some of which happen more often than others. We want to know how often these things have happened to your in the past 6 months. Please circle the appropriate number

		Very often	Quite often	Occasion-ally	Very rarely	Never
1.	Do you read something and find you haven't been thinking about it and must read it again?	4	3	2	1	0
2.	Do you find you forget why you went from one part of the house to the other?	4	3	2	1	0
3.	Do you fail to notice signposts on the road?	4	3	2	1	0
4.	Do you find you confuse right and left when giving directions?	4	3	2	1	0
5.	Do you bump into people?	4	3	2	1	0
6.	Do you find you forget whether you've turned off a light or a fire or locked the door?	4	3	2	1	0
7.	Do you fail to listen to people's names when you are meeting them?	4	3	2	1	0
8.	Do you say something and realize afterwards that it might be taken as insulting?	4	3	2	1	0
9.	Do you fail to hear people speaking to you when you are doing something else?	4	3	2	1	0
10.	Do you lose your temper and regret it?	4	3	2	1	0
11.	Do you leave important letters unanswered for days?	4	3	2	1	0
12.	Do you find you forget which way to turn on a road you know well but rarely use?	4	3	2	1	0
13.	Do you fail to see what you want in a supermarket (although it's there)?	4	3	2	1	0
14.	Do you find yourself suddenly wondering whether you've used a word correctly?	4	3	2	1	0

		Very often	Quite often	Occasion-ally	Very rarely	Never
15.	Do you have trouble making up your mind?	4	3	2	1	0
16.	Do you find you forget appointments?	4	3	2	1	0
17.	Do you forget where you put something like a newspaper or a book?	4	3	2	1	0
18.	Do you find you accidentally throw away the thing you want and keep what you meant to throw away – as in the example of throwing away the matchbox and putting the used match in your pocket?	4	3	2	1	0
19.	Do you daydream when you ought to be listening to something?	4	3	2	1	0
20.	Do you find you forget people's names?	4	3	2	1	0
21.	Do you start doing one thing at home and get distracted into doing something else (unintentionally)?	4	3	2	1	0
22.	Do you find you can't quite remember something although it's "on the tip of your tongue"?	4	3	2	1	0
23.	Do you find you forget what you came to the shops to buy?	4	3	2	1	0
24.	Do you drop things?	4	3	2	1	0
25.	Do you find you can't think of anything to say?	4	3	2	1	0

Scoring the Scale

The CFQ was developed by Broadbent et al. (1982) -- yes, the same Broadbent who proposed the filter theory of attention -- to assess the frequency with which people experienced cognitive failures, such as absent-mindedness, in everyday life -- slips and errors of perception, memory, and motor functioning. The most straightforward way to score the scale is simply to sum up the ratings of the 25 individual items, yielding a score from 0-100.

Scores on the scale predict episodes of absent-mindedness in both the laboratory and everyday life, including slow performance on focused attention tasks, traffic and work accidents, and forgetting to save one's data on the computer.

Survey questionnaire

Note: This questionnaire is anonymous, therefore, please don't mention your name anywhere.

Please mark your gender below;

Male ________

Female ________

Please state your age (in years only) ________

On average, how many hours per day do you spend on homework/self-study? ________

Do you take private tuition? (Please tick in the concerned blank)

Yes ________

No ________

What were your marks (in percentage) in the last three class activities (including homework, exams, and quizzes?)

________ ________ ________

In the last three months, how many times were you absent from the school/college? ________