LONDON METROPOLITAN UNIVERSITY

Islington college
(इस्लिङ्टन कलेज)

Module Code & Module Title

CU6051NI Artificial Intelligence

Individual Coursework

Submission: Proposal

Academic Semester: Autumn Semester 2025

Credit: 15 credit semester long module

Student Name: Sanskriti Agrahari

London Met ID: 23048503

College ID: NP01AI4A230002

Assignment Due Date: 17/12/2025

Assignment Submission Date: 17/12/2025

Submitted To: Er. Roshan Shrestha

GitHub Link	https://github.com/03sans/AI_Coursework-

I confirm that I understand my coursework needs to be submitted online via MST Classroom under the relevant module page before the deadline for my assignment to be accepted and marked. I am fully aware that late submissions will be treated as non-submission and a mark of zero will be awarde

Table of Contents

Table of Figures

Table of Tables

1. Introduction

1.1 Overview of the AI Topic

This coursework is based on the Supervised Machine Learning to conduct a classification task. Supervised learning involves training models using labelled data where each input is associated with a known output, enabling the model to learn relationships between inputs and outputs to make predictions on new data (IBM, n.d.). Classification, a major type of supervised learning, involves assigning inputs into specific categories such as low, moderate, or high stress levels based on observed characteristics and patterns in the data (IBM, n.d.).

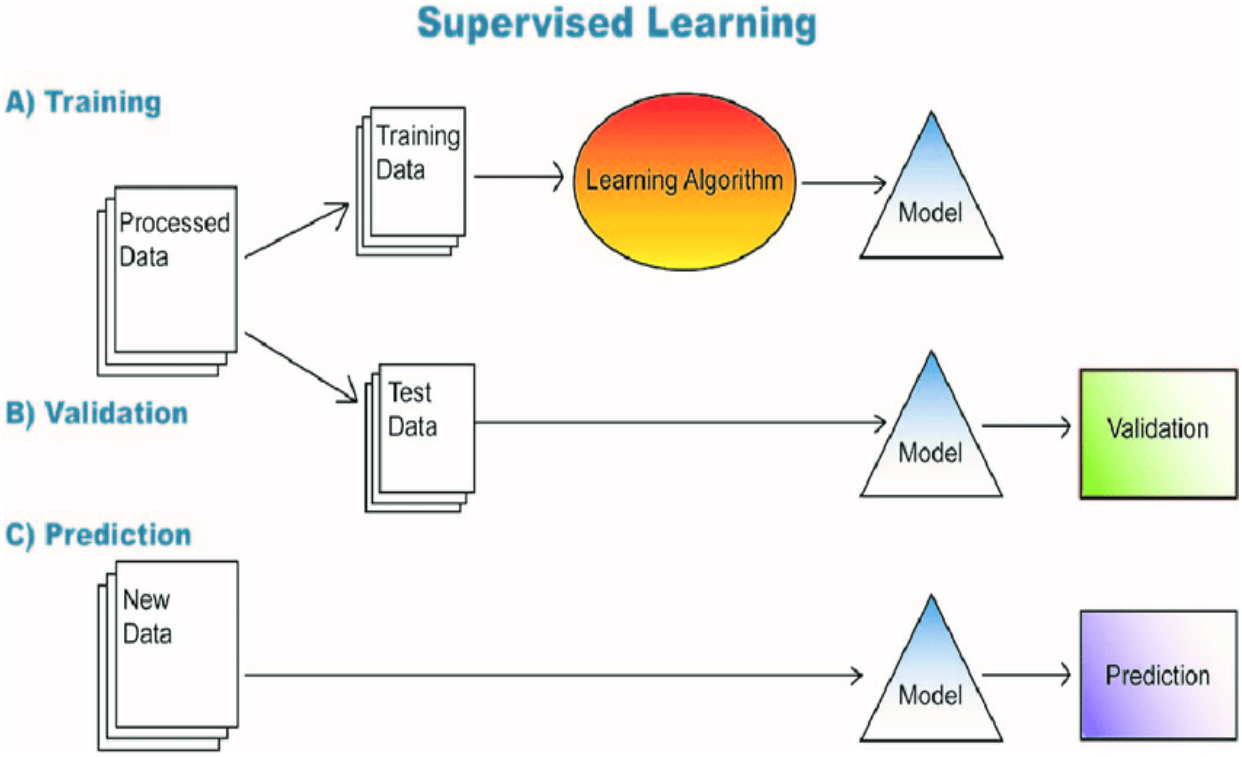

Figure 1 Graphic representation of supervised machine learning (Kanevsky, n.d.)

Supervised classification methods are widely used in predictive analytics to identify risk factors, predict student behaviour, and support the decision-making based on data trends in educational settings. Classification models have been found to help educational

institutions move from reactive approaches to proactive interventions by identifying potential issues at an early stage. In this study, core machine learning processes such as data preprocessing, feature encoding, model selection, training, and evaluation using appropriate performance metrics form the conceptual foundation for designing a student stress classification system.

1.2 Problem Domain: Student Stress Analysis

Stress in students has emerged as a serious concern in today's educational environments due to increasing in academic workload, pressure to perform well in exams, improper sleep, health problems, or bad lifestyle choices, the stress of students has become a major concern at educational institutes. The longer as student is exposed to these concerns causing high levels of stress, the more negatively it impacts their life along with their mental health, physical wellbeing and even their academic performance.

The traditional methods for detecting students who are experiencing excessing stress often rely on either self-reporting, sessions with counsellors or manually by their teachers or academic administrators, all of which are often subjective or may be irregular and delayed. Also, stress is likely influenced by multiple interrelated factors, making it difficult to accurately assess using single indicators or informal observation alone.

Therefore, there is a rapidly growing need for a very systematic and a data-driven approach to tackle this issue, which can analyse multiple causative factors simultaneously and provide early identification of students who are at the risk of high levels of stress.

1.2.1 Problem Statement

This coursework does the job of addressing the lack and need of an automated and data-driven system for identification of students experiencing high levels of stress. Existing approaches are often inconsistent, making early intervention difficult. In order to

automatically predict stress levels, this coursework suggests a supervised machine learning-based classification method that makes use of structured student survey data. The suggested method seeks to facilitate prompt and well-informed decision-making in educational settings by utilising a variety of academic, behavioural, and health-related characteristics.

1.3 Aim of the Proposed Study

The aim of this study is to propose a supervised machine learning-based classification system that uses student survey data to predict stress levels. By analysing academic, behavioural, and health-related attributes, the system aims to achieve:

- The identification of students at high risk for high levels of stress at an early stage.
- To provide support for educators and administrations with insights that are data informed.
- To contribute to improving the well-being of students, ultimately helping them so they can excel academically.

The coursework focuses on conceptual system design, algorithm selection, and logical workflow representation rather than full system implementation.

1.4 Dataset Overview

This study uses a student stress survey dataset obtained from the Hugging Face dataset repository *(0xmarvel/student-stress-survey).* The dataset contains responses collected through a structured questionnaire designed to capture multiple factors related to student stress.

The dataset includes 1,000 student responses and approximately 39 attributes covering academic resources, study environment, learning habits, health conditions, sleep

patterns, lifestyle choices, and perceived stress indicators. One stress-related survey response is treated as the target variable, while the remaining attributes serve as independent features.

The column StudentStressSurvey_Id is used solely as a unique identifier and is excluded from model training, as it does not carry meaningful predictive information. The dataset is suitable for supervised machine learning because the independent variables represent real-world factors that logically influence student stress levels.

Attribute	Description
Dataset Source	Hugging Face (0xmarvel/student-stress-survey)
Number of Records	1,000 student responses
Number of Attributes	39 features
Data Type	Structured survey data
Feature Categories	Academic, behavioural, health, sleep, lifestyle, stress indicators
Target Variable	Stress-related survey response
Identifier Column	StudentStressSurvey_Id (excluded from training)
Suitability for ML	High features have logical influence on stress levels

Table 1 Dataset Overview

This dataset provides a reliable foundation for designing and analysing a supervised classification solution for student stress prediction.

2. Background

2.1 Student stress as an educational and wellbeing problem

Because stress has an impact on learning results, attendance, motivation, and long-term mental health, it is becoming more widely acknowledged as a significant issue in higher education. Global university students have a significant prevalence of mental health problems, including stress-related outcomes, according to extensive evidence syntheses, demonstrating that this is not a minor or regional problem (Paiva, et al., 2025). The necessity for early detection and support systems is also supported by empirical public health research that show significant percentages of students with moderate to high psychological distress and associated risk factors (Malebari, et al., 2024). From a problem-domain viewpoint, stress is influenced by many interrelated factors like academic pressure, sleep quality, lifestyle choices, health problems, and social-environmental context. Therefore, it is hard to make an exact identification of stress through manual observation or single-factor screening.

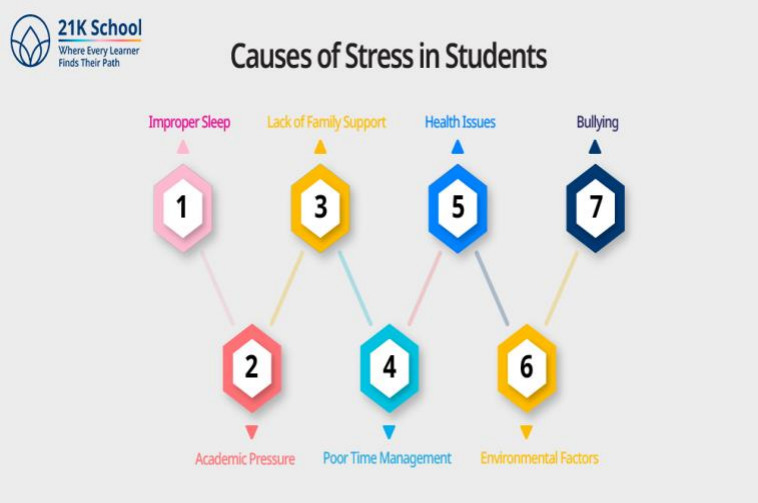

Figure 2 Common factors contributing to stress in students. (21kschools, 2025)

2.2 Research trend: educational data mining and learning analytics

Large volumes of student-related data (such as surveys, academic records, engagement patterns, and wellbeing indicators) have been produced by educational institutions over the past ten years. In order to better understand students and make better educational

decisions, educational data mining (EDM) and learning analytics concentrate on evaluating such data (Papadogiannis, et al., 2024). Supervised machine learning is frequently used in this setting for predictive tasks like identifying students who are at risk, predicting results, and assisting with focused interventions. Predictive models are utilised to identify warning indicators earlier than traditional techniques, allowing for prompt academic or pastoral care, according to a comprehensive review of at-risk student prediction studies (Li, et al., 2024). This encourages the application of comparable prediction techniques to stress-related consequences, where early warning is crucial.

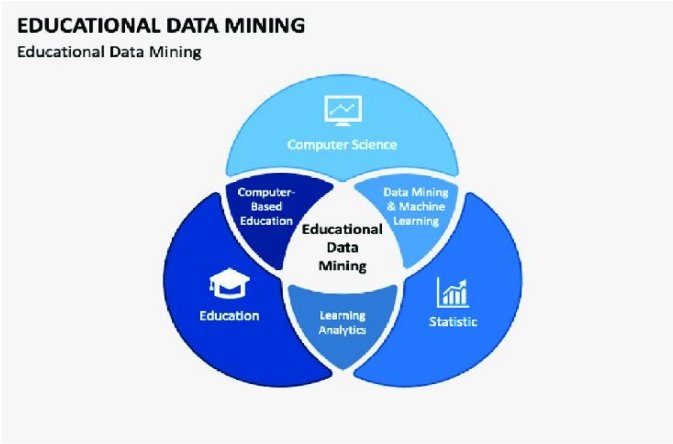

Figure 3 Relationship between educational data mining, learning analytics, and machine learning in educational contexts (Pallathadka, n.d.)

2.3 Machine learning for stress detection and prediction

The majority of machine learning research on stress prediction can be divided into two main streams: sensor-based detection using physiological signals, such as wearable devices, and timeseries modelling, and survey/behaviour-based prediction using self-reports and contextual variables. The ability of machine learning to identify or forecast stress episodes based on physiological inputs is reported in a recent systematic review on wearable-integrated stress prediction. It also emphasises the significance of careful feature engineering, validation, and generalisability across individuals and contexts.(Pataca, et al., 2025). While sensor-based methods may be unquestionably powerful, in most educational contexts cost, privacy, and availability make them hard to implement. Therefore, survey-based datasets such as the Student Stress Survey remain very practical in academic contexts due to their easily accessible multidimensional contributors: sleep, workload, health, habits, and perceived academic pressure.

2.4 Existing work in the student-stress prediction domain

Recent studies propose supervised learning systems that classify student stress using multi-factor inputs and produce categorical outputs such as low/medium/high stress. For example, a context-aware machine learning framework (using survey-based datasets) demonstrates that stress classification can be improved by combining diverse student-context variables and evaluating multiple model families rather than relying on a single algorithm (Ovi, et al., 2025). Similar to this, models such as Logistic Regression, Decision Trees, Random Forest, SVM, and Neural Networks are frequently compared in applied studies that develop stress-level predictors because they manage mixed-feature patterns and offer robust baselines for classification performance (Yeler & Sürücü, 2025). Cleaning and validation, category encoding, feature scaling (if necessary), train/test splitting, model training, and evaluation using measures like accuracy, precision, recall, F1-score, and confusion matrices are common components of reported processes found in this literature.

2.5 Explainability, ethics, and practical deployment considerations

In applications pertaining to student wellbeing, trust is a significant problem. Clear explanations of predictions are often needed by stakeholders (teachers, counsellors, and students), especially when the forecasts may result in interventions. Interpretability strategies (such feature importance and model-agnostic explanations) may improve transparency and encourage safer decision-making, according to current research on explainable AI in mental health predictive modelling (Tariq, et al., 2025). Because student stress data is sensitive, systems must put privacy, informed consent, and careful framing first so that forecasts help students rather than stigmatise or label them. These moral issues are also very important. The idea of a "conceptual solution" that emphasises explainability, responsible use, and human-in-the-loop decision-making is strengthened by this.

2.6 Review of Related Research Work

There has been a growing body of research that have investigated the implementation of machine learning techniques in order to analyse stress in students, their mental wellbeing, and even academic risk using survey-based and educational datasets. These papers offer crucial information about suitable datasets, algorithmic approaches, and expected outcomes for stress prediction and early risk identification in educational environments.

Research 1: (Paiva, et al., 2025)

- **Dataset:** Meta-analysis of global university student mental health and wellbeing survey studies
- **Algorithm:** A large-scale statistical analysis combined with machine learning-based synthesis methods
- **Key Findings:** According to this study, university students in a variety of geographical locations and educational settings frequently experience psychological and stress-related symptoms. The results show that rather than being a singular or localised issue, student stress is a pervasive and ongoing problem. In order to address stress before it has long-term negative effects on academic performance or health, the authors emphasised the significance of early detection mechanisms and data-driven support systems.

Research 2: (Malebari, et al., 2024)

- **Dataset:** Public health datasets focusing on student psychological distress, lifestyle factors, and academic pressure
- **Algorithm:** Predictive statistical modelling and risk-factor analysis

- **Key Results:** Strong correlations between students' academic workload, sleep disturbances, and moderate-to-high psychological distress were found in the study. The findings, which show that stress is impacted by a mix of behavioural, academic, and lifestyle factors rather than a single cause, support the use of multi-factor datasets for stress assessment.

Research 3: (Li, et al., 2024)

- **Dataset:** At-risk student datasets collected from educational institutions, including engagement and wellbeing indicators
- **Algorithms:** Supervised machine learning classifiers
- **Key Results:** The authors showed that supervised machine learning models outperformed conventional screening methods in identifying early warning signs of student risk. According to their findings, predictive models can help students receive timely academic and pastoral interventions by identifying risk patterns prior to serious academic decline or mental health issues.

Research 4: (Ovi, et al., 2025)

- **Dataset:** Survey-based student stress datasets incorporating academic, behavioural, and contextual variables
- **Algorithms:** Logistic Regression, Support Vector Machine (SVM), and Random Forest
- **Key Results:** Combining various contextual features greatly increased the accuracy of stress classification, according to this study's comparative analysis of several supervised learning algorithms. The findings emphasise how crucial feature diversity and algorithm comparison are when creating stress prediction systems based on survey data.

Research 5: (Yeler & Sürücü, 2025)

- **Dataset:** Student performance and stress-related datasets
- **Algorithms:** Decision Trees, Random Forest, and Neural Networks
- **Key Results:** The study found that tree-based model, Random Forest, achieved strong predictive performance while maintaining a high level of interpretability. This balance between accuracy and transparency makes such models especially suitable for educational decision-support systems, where understanding model behaviour is important for trust and ethical use.

All things considered, these studies demonstrate the efficacy of supervised machine learning techniques for predicting student stress, especially when employing survey-based datasets with multi-dimensional features. The design decisions in this coursework, such as the choice of dataset, algorithm, and evaluation method, are directly influenced by the results of previous research.

2.7 Summary of research gap and justification for this coursework

Overall, research shows strong potential for machine learning to classify or predict stress, and educational analytics research supports using supervised models for early risk detection (Papadogiannis, et al., 2024) (Li, et al., 2024). However, gaps remain in building practical, survey-based stress prediction pipelines that handle mixed categorical/numerical features robustly, clearly defining labels from survey instruments in a consistent way, and presenting solutions with transparent logic suitable for real academic settings (Tariq, et al., 2025). Therefore, this coursework proposes a supervised classification approach using the Student Stress Survey dataset to develop a detailed conceptual solution, including algorithm selection rationale, pseudocode, and diagrammatic representation, aligned to realistic educational needs.

3. Proposed Solution

3.1 Overview of the Proposed Solution

The proposed solution is a supervised machine learning-based classification system designed to predict student stress levels using survey data. The system analyses multiple factors related to students' academic workload, study habits, health, sleep patterns, lifestyle, and perceived stress indicators. One stress-related survey question is treated as the target variable, while the remaining attributes are used as input features.

The solution follows a standard machine learning pipeline consisting of data preprocessing, feature encoding, model training, evaluation, and prediction. During training, supervised classification algorithms would learn patterns that could link student attributes to different stress levels. The trained model is then evaluated using appropriate classification metrics to ensure reliable performance.

The final output of this proposed system is to be a classified and interpretable stress level (Low, Moderate, or High). The output can be used to support early identification of students experiencing high levels of stress, who may require academic or wellbeing support, enabling more informed and timely decision-making in educational settings.

3.2 Proposed Approach to Solving the Problem

The proposed solution will follow a very structured supervised machine learning workflow. Each and every stage is important and contribute to transform the student survey data into an extremely reliable and interpretable stress-level classifications. The aim of implementing all these stages is to ensure quality of data, have a meaningful learning experience, and have a useful output.

Stage 1: Data Selection

The very stage involves the selection of appropriate data for analysis. For this project, student responses are obtained from an existing structured survey dataset containing academic, behavioural, health, and lifestyle-related attributes. A stress-related survey question is selected as the classification label. The selected features represent real-world factors such as academic pressure, sleep quality, and health conditions that have a direct influence on student stress levels. This is to ensure that the dataset is suitable for supervised learning, where meaningful relationships exist between inputs and outputs.

Stage 2: Data Preprocessing

The next step is the data pre-processing step which is crucial as it prepares the raw dataset for effective model training. Datasets may contain attributes or columns that are not directly suitable for machine learning and therefore such attributes require transformation.

In this project, the following preprocessing operation are to be applied:

- **Removal of irrelevant identifiers:** Attributes like surveyIDs are removed because they do not provide any predictive information and would most like introduce noise in the system.
- **Handling categorical responses:** The categorical responses in the survey will be converted into numerical representations using proper encoding techniques to allow the algorithms to process them effectively.
- **Handling numerical features:** All the numerical and Likert-scale features will be retained in numeric form but according to requirement, normalization or scaling can be applied to ensure that some features with ranges on the larger end do not dominate the learning process of the algorithms.
- **Dataset Splitting:** The final processed dataset will then be divided into training and testing sets, allowing the evaluation and learning by models to be performed on separate subsets of data.

Stage 3: Model Training

In this stage of implementation, a supervised classification model is trained using the preprocessed training dataset. The model then learns patterns and relationships between the input features and the labels of stress. Through this learning process, the model develops decision boundaries that allow it to classify students into different stress categories being low, moderate or high.

In this project, different supervised classification algorithms will be applied, which will be explained below, during training to capture both linear and non-linear relationships between features and stress levels. The objective of this stage is to produce a trained model that generalises well to unseen data rather than memorising the training instances.

Stage 4: Model Evaluation

After the training is complete, in this stage, the model's performance is accessed with the help of the testing dataset that was split during the pre-processing stage. The evaluation is a very essential stage to determine exactly how accurate the trained model predicts levels of stress for data that it has not seen during training the model.

In this stage, standard classification metrics such as the following are used to measure the performance of the model:

- The accuracy metric is used to access overall how correct the predictions are.
- The precision metric is used to evaluate how accurate the predicted stress classes are.
- The recall metric is used to measure how well the model identifies any stressed student.
- The confusion matrix analysis is used to determine the misclassification patterns across stress levels.

Stage 5: Prediction and Output

This is the final stage of the implementation. In this stage, the trained as well as the evaluated model is used to predict stress level for new student survey inputs. On the basis of the patterns that the model has learned, the system then assigns an interpretable stress category, such as Low, Moderate, or High, for each student. This stage represents the practical outcome of the proposed solution.

3.3 AI Algorithms Used

The solution uses Supervised Learning Classification algorithms to predict the stress levels in the students based on the structured survey data. For this problem, supervised classification is the most appropriate because the dataset contains labelled instances, where a stress-related survey response serves as the target variable. These selected algorithms are known to be widely used in educational and wellbeing analytics, as also seen during the research, due to them being highly reliable, interpretable, most importantly suitable for structured data.

- **Logistic Regression:** Logistic Regression is employed as a baseline classifier. Logistic Regression calculates the probability of a particular student belonging to a certain category of stress based on a weighted combination of inputs. Although it is a simple model, it is useful because it is easy to interpret, allowing one to see directly how a variety of factors influence a stress level prediction.
  With regards to the study conducted, Logistic Regression will be used as a benchmark model by which more complicated algorithms can be measured. It's a clear way of deciding, making it ideal for educational institutions.

- **Decision Tree Classifier:** The Decision Tree Classifier is a model for decision-making process in a tree form, where interior nodes are decisions made from features, and leaf nodes are outcomes related to levels of stress. The Decision

Tree Classifier is a suitable model for this problem because it can address non-linear relationships found in variables like the attributes of students and their levels of stress.

Decision Trees are highly interpretable because the sequence of decision that led the classifier to a particular classification outcome is easy to visualise as well as explain. This makes them ideal for survey data, where the relationship between variables such as sleep quality, academic workload, or personal health may not necessarily be linear.

- **Random Forest Classifier:** Random Forest, on the other hand, is an ensemble method where several decision trees are used. Decision trees are built on different samples of the data and predictors, and the results are combined using a majority rule of voting.

  This helps to combat the problem of overfitting, which often tends to be common within Decision Trees. Random Forest models work best on survey-based datasets with complex feature correlations and offer relevant additional insights about important variables that tend to impact stress levels within students.

Suitability of the selected algorithms for this problem

Among these, Decision Tree or Random Forest classifiers are particularly suitable because:

- They handle mixed numerical and categorical features effectively.
- They also capture complex and non-linear relationship between different factors.
- They provide interpretable outputs, which is extremely important for ethical and practical use, especially in education context.
- They perform well on survey-based datasets with multiple interacting variables.

Other supervised classification algorithms such as Support Vector Machine (SVM) or gradient boosting models may also be explored for comparative analysis, depending on

performance outcomes. However, the final algorithm selection will be based on comparative evaluation results, balancing prediction performance with interpretability.

Algorithm	Key Characteristics	Relevance to This Study
Logistic Regression	Simple, probabilistic, highly interpretable	Provides baseline performance and transparency
Decision Tree	Rule-based, non-linear, interpretable	Captures complex decision logic in survey data
Random Forest	Ensemble of trees, robust, reduced overfitting	Improves accuracy and handles feature interactions

Table 2 Summary table of Selected Algorithms

3.4 Pseudocode and Diagrammatic Representation

3.4.1 State Transition Representation

Figure 4 presents the state transition representation of the system, highlighting the transformation from raw survey data to predicted stress categories.

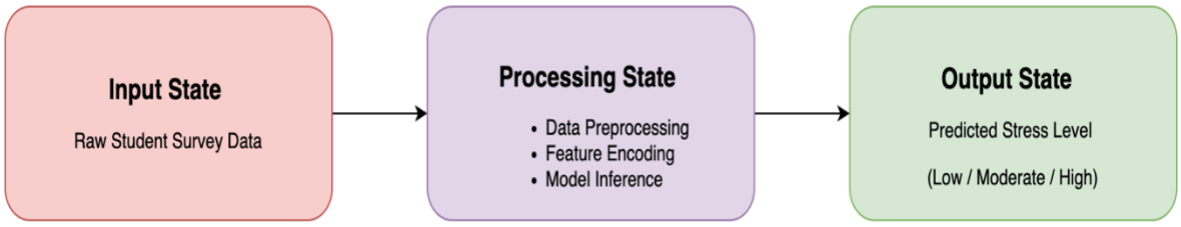

Figure 4 State Transition Representation

3.4.2 Pseudocode and Diagrammatic Representation for overall system

START

Import required libraries

Load student stress survey dataset

Remove irrelevant columns

Identify target variable

Select remaining columns as input features

Preprocess data

Encode categorical survey responses

Retain numerical and Likert-scale features

Apply scaling or normalisation if required

Split dataset into training set and testing set

Select supervised classification algorithm

(Logistic Regression / Decision Tree / Random Forest)

Train selected model using training data

Evaluate trained model using testing data

Compute accuracy, precision, recall

Generate confusion matrix

IF model performance is acceptable **THEN**

Save trained model

Use trained model to predict stress level for new student survey inputs

ELSE

Tune model parameters or change algorithm

Retrain model

END IF

Output predicted stress category (Low / Moderate / High)

END

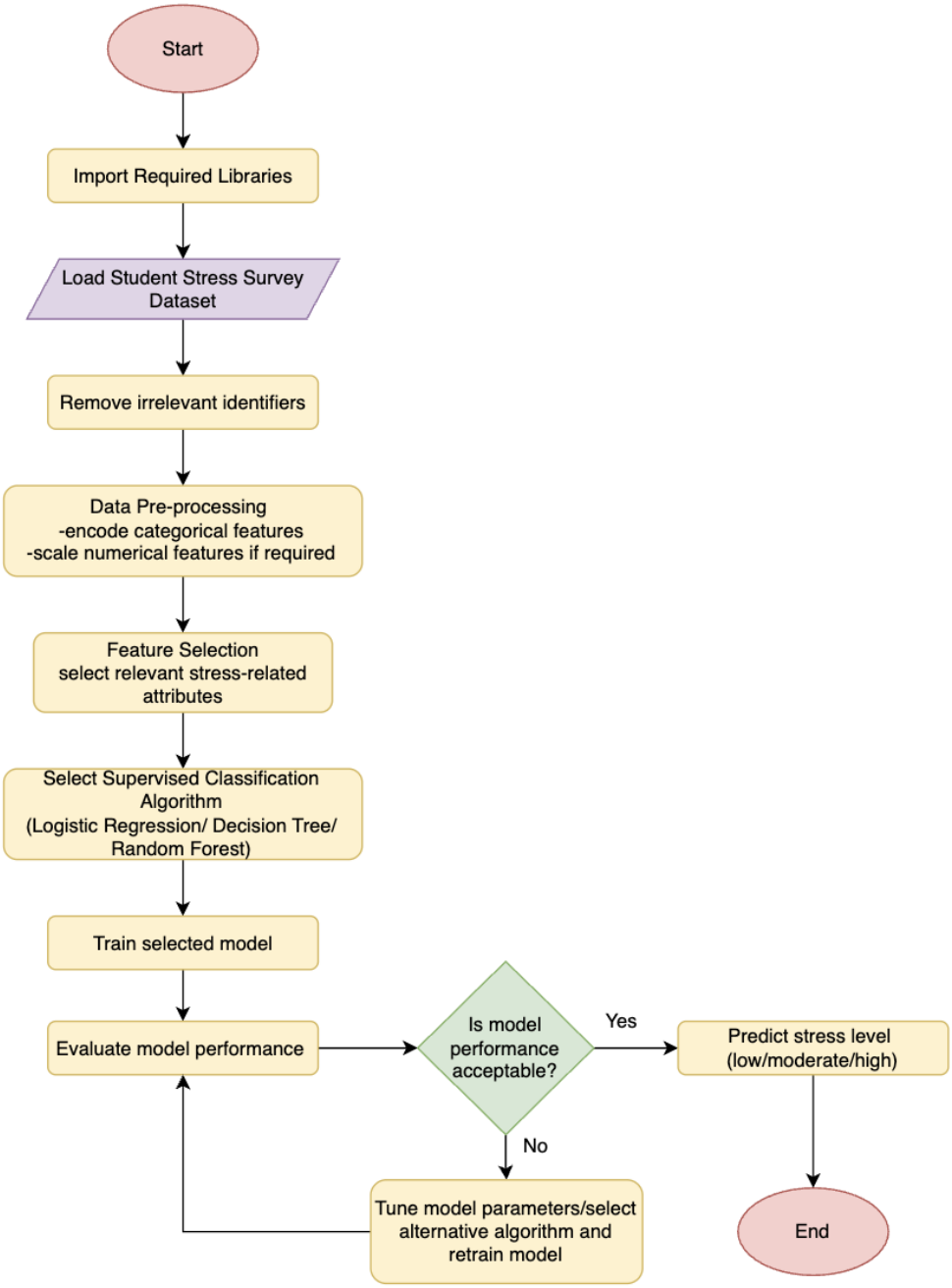

Figure 5 Flowchart of overall system

3.4.3 Pseudocode and Diagrammatic Representation for Logistic Regression Classifier

START

Import required libraries

Load pre-processed student stress dataset

Separate dataset into input features(X) and stress-level labels(Y)

Split data into training set and testing set

Initialize Logistic Regression model parameters (weights and bias)

Train model using training data

Compute weighted sum of input features

Apply logistic (sigmoid) function

Estimate probability for each stress class

Optimise parameters to minimise classification loss

FOR each instance in testing set **DO**

Compute probability of stress classes

Assign stress category based on probability threshold

END FOR

Evaluate model using accuracy and confusion matrix

IF performance is satisfactory **THEN**

Save trained model

ELSE

Adjust model parameters and retrain

END IF

END

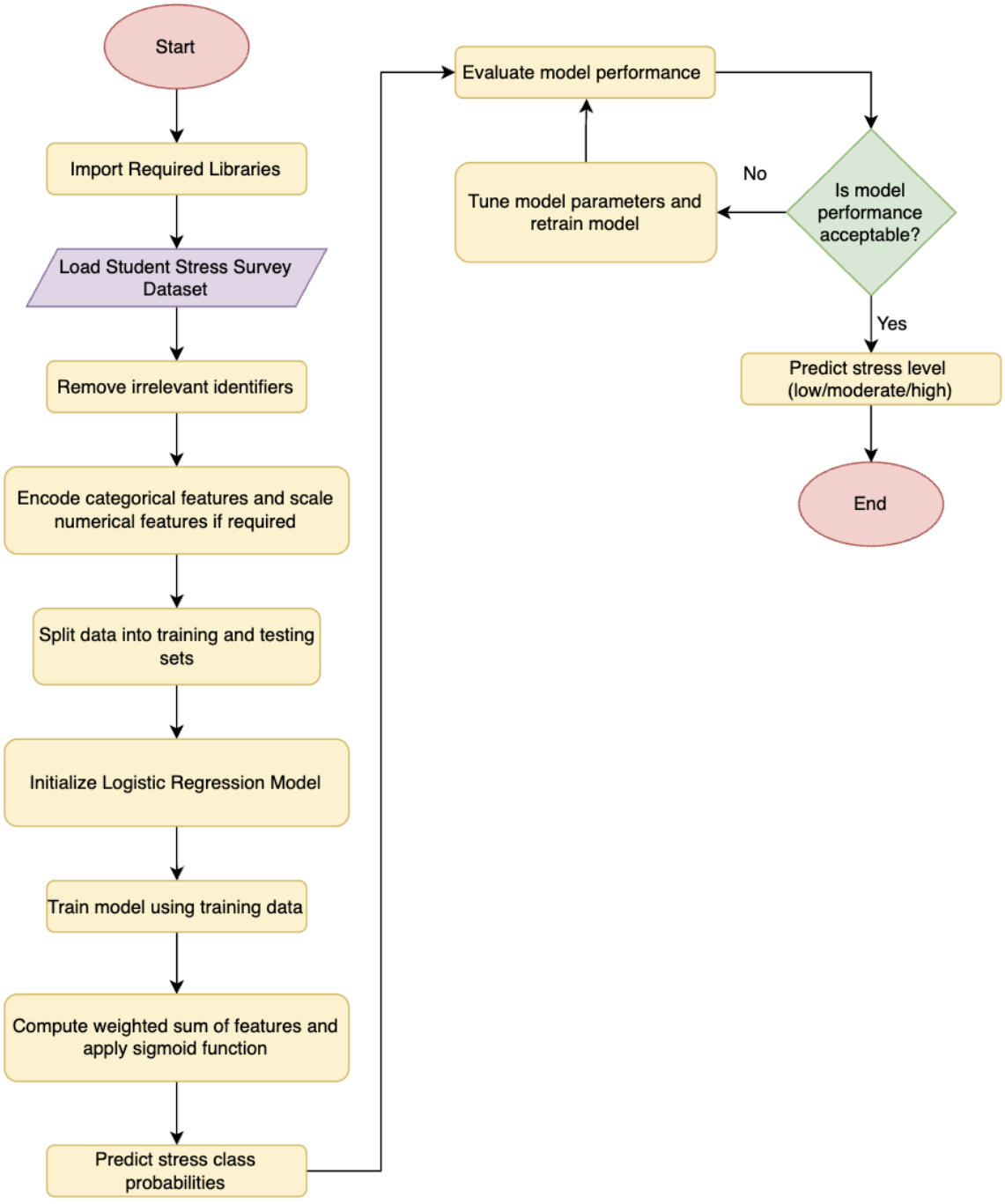

Figure 6 Flowchart of Logistic Regression Classifier

3.4.4 Pseudocode and Diagrammatic Representation for Decision Tree Classifier

START

Import required libraries

Load pre-processed student stress dataset

Separate dataset into input features (X) and stress-level labels (Y)

Split data into training set and testing set

Create root node of the decision tree

WHILE stopping condition is not met **DO**

    Evaluate all features using impurity measure

    Select feature that best splits the data

    Partition dataset based on selected feature

    Create child nodes for each partition

END WHILE

Assign stress-level labels to leaf nodes based on majority class

FOR each instance in testing set **DO**

    Traverse decision tree from root to leaf

    Assign predicted stress level

END FOR

Evaluate model performance, calculate accuracy and generate confusion matrix

END

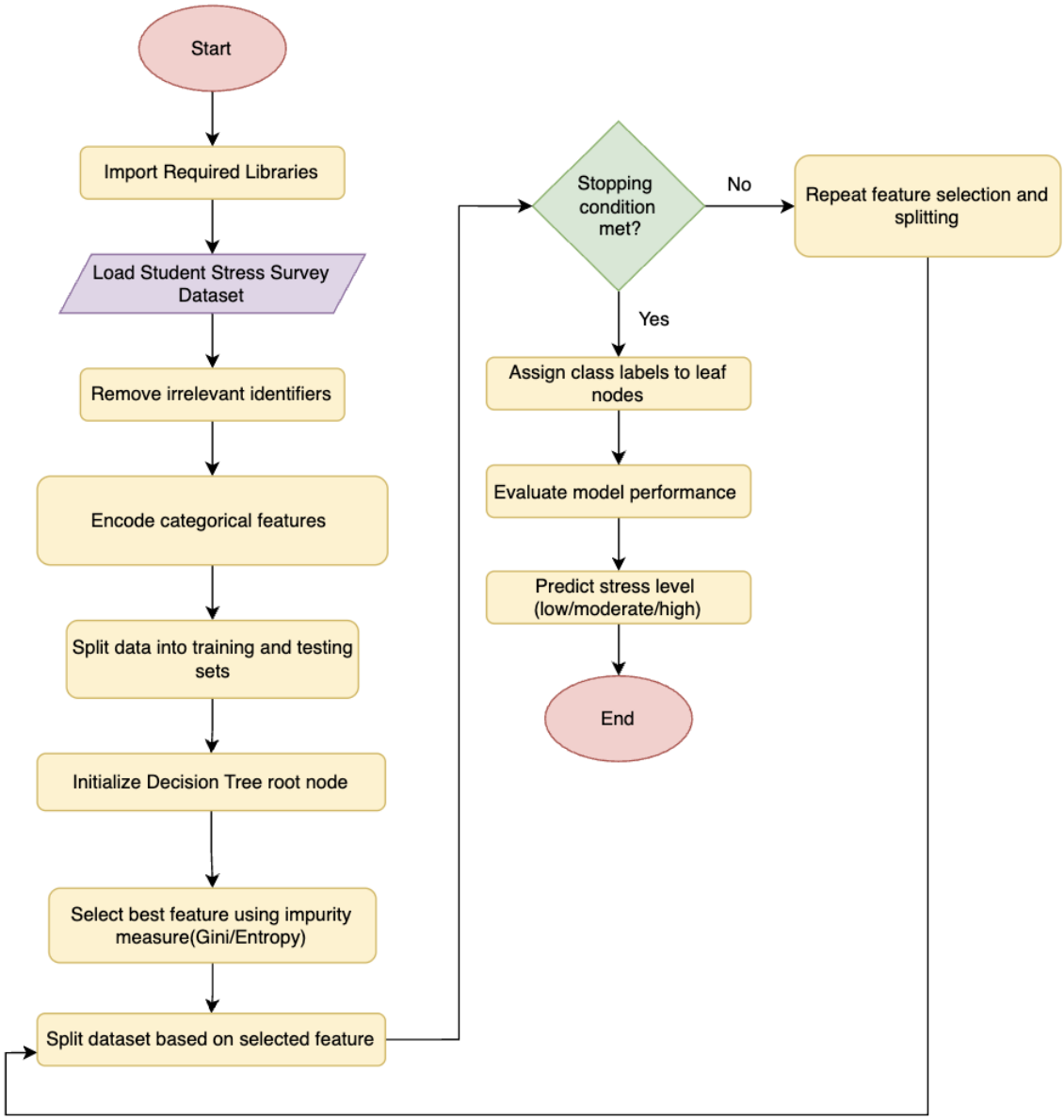

Figure 7 Flowchart of Decision Tree Classifier

3.4.5 Pseudocode and Diagrammatic Representation for Random Forest Classifier

START

Import required libraries

Load pre-processed student stress dataset

Separate dataset into input features (X) and stress-level labels (Y)

Split data into training set and testing set

Initialize number of decision trees in the forest

FOR each tree in the forest **DO**

Generate a random bootstrap sample from training data

Select a random subset of features

Train a decision tree on sampled data

END FOR

FOR each instance in testing set **DO**

Collect predicted stress level from all trees

Assign final stress level using majority voting

END FOR

Evaluate Random Forest using accuracy and confusion matrix

Use trained model to predict stress levels for new student data

END

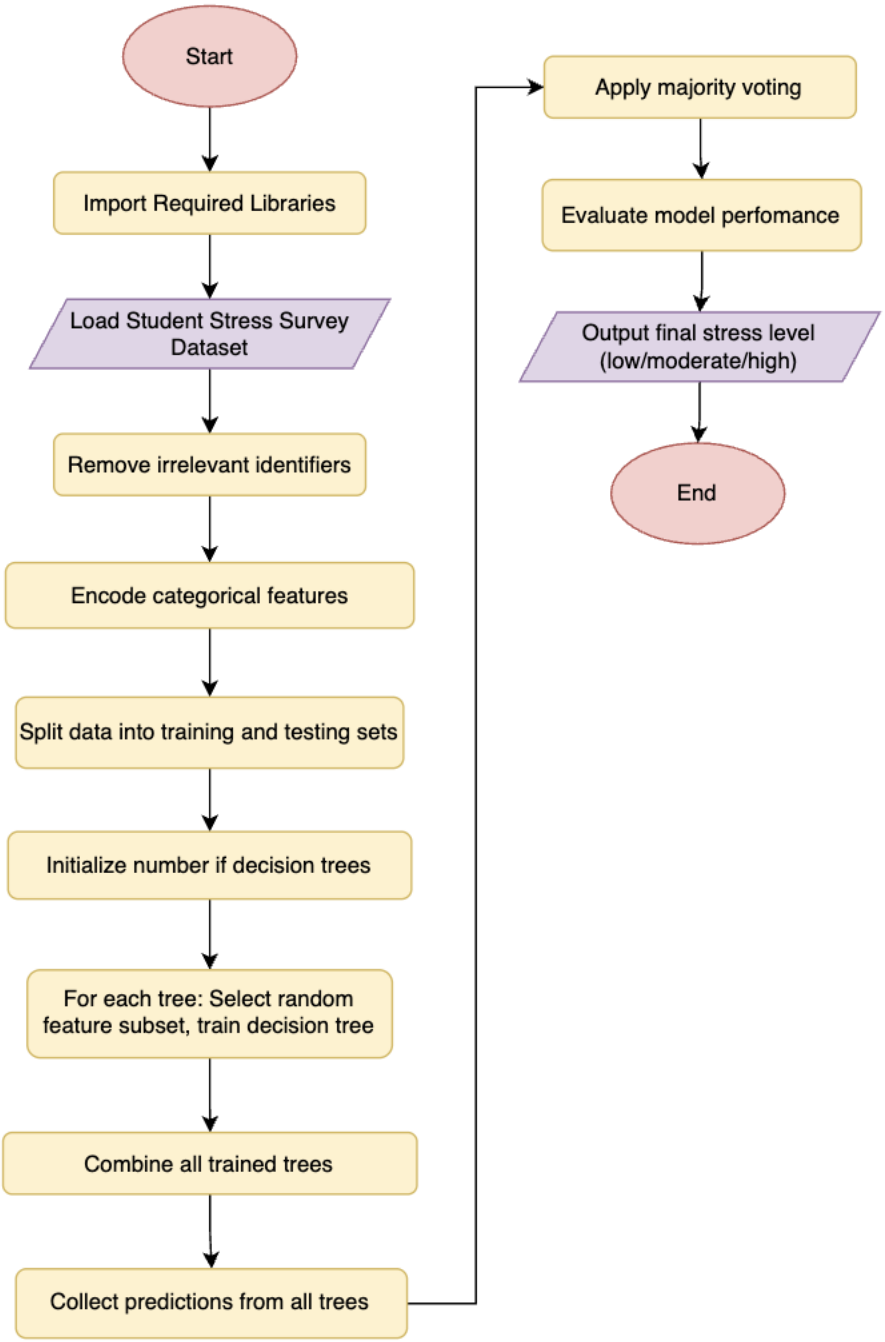

Figure 8 Flowchart of Random Forest Classifier

3.5 Summary

The proposed solution offers a fully thought-out conceptual approach towards applying supervised machine learning methodologies towards the problem of student stress identification using a survey-driven dataset. The approach is fully centred around a logical and structured machine learning solution, starting from the selection and preparation of data, through processing, training, and evaluation, and finally through arriving at a final student stress level classification. Every component of the solution is fully and carefully thought-out and aligned with clarity and educational dataset requirements.

With the help of supervised classification algorithms like Logistic Regression, Decision Trees, and Random Forests, the solution shows how different methods of modelling can be harnessed to identify linear and non-linear relationships between factors related to stress. With the help of a variety of models being included for analysis, there also comes the benefit of comparative analysis to make informed decisions regarding the selection of the model based on demands for performance and interpretability. Detailed pseudocodes and flowcharts also help to make the system more transparent by explaining the working of the models.

On the whole, the proposed solution is an effective and scalable conceptual framework for the earliest detection of student stress in an educational setting. The importance of data-informed decision support in the identification of high-risk individuals is clearly illustrated, and the significance of sensible design principles, including clarity, explainability, and decision making, is underlined. As it is at the proposal level, there is considerable promise of successful implementation.

4. Conclusion

4.1 Analysis of the Work Done

This coursework is an account of the proposal of the use of a supervised machine learning-based solution for classifying student stress levels using structured survey data. All the work that has been done so far has been carried out in a very systematic manner, which begins with the analysis of the problem domain and then moving on to a detailed review of five existing research works accounting student stress, educational data mining, and prediction systems based in machine learning.

The key components of the work completed so far include the following:

- The identification of stress in students as a vital educational and wellbeing issue requiring early detection and prediction system.
- Thorough analysis and review of relevant academic papers to understand approaches, datasets, algorithms that already exist within the domain of student stress and risk prediction.
- The selection a suitable and relevant survey-based dataset that contains multi-dimensional factors like academic workload, health, sleep, lifestyle, and other indicators of stress.
- The design of a complete supervised machine learning pipeline that includes data selection and preprocessing, feature encoding and preparation, training of the model and its evaluation and finally the stress level prediction.
- The justification as to why a certain supervised classification algorithm was to be used.
- The development of pseudocode and diagrammatic representation to illustrate the system’s workflow and algorithm logic clearly.

This coursework is presented at a very conceptual level since the focus is on the proposal development and design of the system. So far, there has not been any implementation or evaluation of model performance. This is to be carried out in the next coursework.

4.2 Real-World Applicability of the Solution

A real-world challenge, an educational one was addressed in this report. The solution proposed offers how to tackle this challenge through a data driven approach for identifying students who may be experiencing stress in their life. There are many inconsistencies in the detection of stress in students through traditional methods such as self-reporting or teacher's observation. They can be delayed and also subjective.

The proposed system has many practical relevance in the real educational environments which is possible due to analysis of multiple contributing factors. The potential real-world relevancies are listed as follows:

- The early identification of students who are at risk of high levels of stress before any serious academic or mental health consequences occur in their life.
- The conscious support for counsellors and teachers to help them make accurate and informed decision relating to students.
- The reduction of depending on single-factor detection methods that are inconsistent and not viable in actual life.
- The immense improvement of student's wellbeing through supportive intervention that can be achieved in time.
- The contribution to better performance and engagement in regard to academics, ultimately enhancing the overall learning experience of students.

Since a survey-based data will be used to implement the solution, it makes these solutions much more practical and effective, as it does not require intrusive methods of data collection.

4.3 Further Work

In any project, there are always potentials for further works that can be implemented. While this coursework does the job of presenting a very relevant solution, there are still

some future developments and enhancements within this project that can be made. These works are listed below:

- In the future, there can be implementation of the proposed system with the help of real machine learning models to help evaluate performance empirically.
- There can be experiments with additional classification algorithms or ensemble techniques to help increase the accuracy of predictions being made.
- There could be use of cross-validation and class imbalance handling techniques to make the model strong.
- In the near future, there could be incorporation of longitudinal data to analyse trends in stress.
- The integration of explainable AI technique such as SHAP, could definitely be used to improve the transparency and trust in using the system.
- A web-based application could also be developed in the future that allows real-time assessment for several education institutions, making it scalable.

All these extensions could very well be applied in the future in order to make this system much more valuable, functional and usable.

Bibliography

IBM, n.d. *How supervised learning works.* [Online] Available at: https://www.ibm.com/think/topics/supervised-learning [Accessed 14 12 2025].

IBM, n.d. *What is classification in machine learning?.* [Online] Available at: https://www.ibm.com/think/topics/classification-machine-learning#684929709 [Accessed 14 12 2025].

Paiva, U. et al., 2025. Prevalence of mental disorder symptoms among university students: An umbrella review. *Neuroscience & Biobehavioral Reviews.*

Malebari, A. M. et al., 2024. Prevalence of depression and anxiety among university students in Jeddah, Saudi Arabia: exploring sociodemographic and associated factors. *Frontiers in Public Health,* Volume 12.

Papadogiannis, I., Wallace, M. & Karountzou, G., 2024. Educational Data Mining: A Foundational Overview. *MDPI.*

Li, K. C., Wong, B. T.-M. & Liu, M., 2024. A survey on predicting at-risk students through learning analytics. *International Journal of Innovation and Learning.*

Pataca, A. O. et al., 2025. Use of machine learning for predicting stress episodes based on wearable sensor data: A systematic review. *Computers in Biology and Medicine,* Volume 198.

Ovi, M. S. I., Hossain, J., Rahi, M. R. A. & Akter, F., 2025. Protecting Student Mental Health with a Context-Aware Machine Learning Framework for Stress Monitoring. *arxis.*

Yeler, K. & Sürücü, S., 2025. Classification of Student Stress Levels Using Machine Learning Methods. *IJANSER,* Volume 9.

Tariq, R. et al., 2025. Explainable artificial intelligence for predictive modeling of student stress in higher education. *National Library of Medicine.*

Kanevsky, J., n.d. *Research Gate.* [Online]
Available at: https://www.researchgate.net/figure/Graphic-representation-of-supervised-machine-learning-In-supervised-learning-original_fig1_301688300
[Accessed 15 12 2025].

21kschools, 2025. *Top 7 Causes of Stress in Students And How to Manage Them.* [Online]
Available at: https://www.21kschool.com/in/blog/causes-of-stress-in-students/
[Accessed 15 12 2025].

Pallathadka, H., n.d. *Machine learning and educational data mining.* [Online]
Available at: https://www.researchgate.net/figure/Machine-learning-and-Educational-Data-Mining_fig1_353611098
[Accessed 15 12 2025].