

**DAFFODIL INTERNATIONAL UNIVERSITY**

**FYDP (Phase-I) Progress Report**

**Reporting Period- Fall 2024**

**Project Identification:**

|  |  |  |
| --- | --- | --- |
| **I. Project Title** | Student Mental Health Prediction using maching learning. | |
| **II. Group Members** | 1. Name: Atikul Islam Hemal Student ID:212-15-4222 2. Name: Student ID: | |
| **III. Supervisor** | Name: Mr.Md Assadurzzaman (MA)  Designation: Sr.Lecturer | |
| **IV. Co-Supervisor** | Name: Montasir Delwar Afnan (MDA)  Designation: Lecturer | |
| **V. Submission Date:** |  | |
| **VI. Certificate :** | “This is to certify that the final year design project work until Phase-I evaluation held on **\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**, titled as stated in *Sec. I*, executed by the students’ group mentioned in *Sec. II*, have been found satisfactory and every section of this report is reflecting the same.” | *(Signature of Supervisor & date)* |

**Project Insights**

|  |  |  |
| --- | --- | --- |
| **Thematic Area(s):**  ***[Just click the check box]*** | Artificial Intelligence and Machine Learning |  |
| Data Science and Analytics |  |
| Cybersecurity |  |
| Software Engineering and Development |  |
| Blockchain Technology |  |
| Internet of Things (IoT) |  |
| Computer Networks |  |
| Human-Computer Interaction (HCI) |  |
| Big Data Technologies |  |
| Computer Vision |  |
| Natural Language Processing (NLP) |  |
| Robotics |  |
| Game Development |  |
| Cloud Computing |  |
| Biomedical Computing |  |
| **Others *(please specify)*:** | |
| **Software packages, tools, and programming languages** | [list down the software, programming languages, and packages you have used to do your project] | |

**CO Description for FYDP-Phase-I**

|  |  |  |
| --- | --- | --- |
| **CO** | **CO Descriptions** | **PO** |
| **CO1** |  | **PO1** |
| **CO2** |  | **PO2** |
| **CO3** |  | **PO4** |
| **CO4** |  | **PO11** |

***\*\* Follow the guideline of COs from the FYDP Phase-I outline. Your CO have to map with the POs which is defined by the FYDP guideline.***

1. **Project Overview:**
   1. **Introduction**

- Mental health is an important component of general well-being because it enables people to manage life's problems, identify their potential, gain information effectively, work efficiently, and interact with their society. It has a significant impact on both personal and social growth by affecting decision-making, relationship-building, and societal contributions. Individual fulfillment, societal strength, and economic advancement all depend on mental health, which is a fundamental human right[1].Society frequently points out physical health and outward looks while ignoring the importance of mental health. However, maintaining excellent mental health is necessary for successfully navigating life from birth to maturity. Mental health issues can affect anybody, no matter age, gender, race, geography, or religious views.[2].In Bangladesh, mental health discussions are usually avoided due to judgment, since many people consider mental illness as a secret.. Those that are struggling are frequently ignored rather than helped, causing the problem. The Directorate General of Health Services (DGHS) reports that rising suicide rates reflect a developing mental health crisis. A countrywide study of higher education students indicated that 75% suffer from depression. Using a Beck Inventory (BDI-III), 21% expressed severe depression, 25% moderate to severe depression, and 26% mild to moderate symptoms, highlighting the critical need for mental health knowledge and assistance[3].Identifying the causes of mental health problems among students has been a challenge. These elements may be impacted by psychology, biology, or the environment. The familiarity of some symptoms and overlapping circumstances complicates diagnosis, frequently leading to physician errors.[5] A recent survey of 1,773 students revealed that 85.9% feel the internet contributes to their mental health problems. Students often face negative affects as a result of the internet's strong impact on their lives. According to the research, 72.2% of students face mental health issues, with 26.1% blaming the internet completely 59.8% considering it an important factor, while only 8.3% not attributing their problems to it.[4]

One of the many applications of AI nowadays is machine learning, particularly in the diagnosis of illnesses. Doctors can use it to evaluate large amounts of patient data and provide individualized therapies according to medical problems.[6] Additionally, by identifying hidden patterns in complex data, machine learning (ML) helps analysis of mental health. It identifies risk indicators, predicts conditions, and recommends appropriate solutions ML makes it possible for early detection and timely student help by automating these steps.[7]

* 1. **Background**

- Student mental health has been an increasingly important problem in recent years, with a rising number persons dealing with issues such as depression, anxiety, and stress during their studies. These mental health issues has an effect not only on students' well-being, but also on academic achievement, social interactions, and overall quality of life. According research, a significant number of students struggle with mental health issues, yet most do not seek help until their conditions gets worse. This delay in care stresses the important necessity for early detection and proactive support measures to treat mental health disorders before they worsen.Machine learning (ML) provides a viable approach for expecting mental health difficulties for youngsters by examining a variety of data that includes academic records, surveys, and behavioral patterns. Using ML models, researchers can identify complex structure in student behavior that indicate mental health issues, commonly before they become apparent with traditional methods. Support Vector Machines , Random Forests, and Neural Networks have all showed the ability to accurately forecast disorders such as anxiety and depression. However, barriers keep occurring, including concerns about dataset quality, data privacy, and the ethical use of algorithms for prediction.

Hanif Abdul Rahman et al. [17] submitted a study titled "Machine Learning-Based Prediction of Mental Well-Being Using Health Behavior Data from University Students," in Bioengineering (2023, 10(575)). The study used an online cross-sectional survey of 15,366 university students from 17 universities in seven ASEAN countries (Brunei, Indonesia, Malaysia, the Philippines, Singapore, Thailand, and Vietnam) to examine mental health and health-risk behaviors such as physical activity, diet, smoking, and alcohol consumption. To handle missing data, the dataset was preprocessed by multiple imputation (MICE) and balanced using the Synthetic Minority Oversampling Technique. Several machine learning classifiers were utilized, including GLM, KNN, NB, NNet, RPART, bagging, and boosting. Among these, Random Forest surpassed all other models with an accuracy of 92.1% on training data and 90.1% on test data, while Adaptive Boosting did well with an accuracy of 89.3%. Other models, such as KNN (77.5%), Naïve Bayes (72.3%) and Neural Networks (61.5%), performed relatively poorly. The study identified eleven main variables that impact mental well-being, including body mass index (BMI), number of sports activities per week, grade point average (GPA), sedentary hours, age, gender, salt intake, fruit and vegetable consumption, sleep hours, and physical activity levels. The authors highlight the success of machine learning for identifying mental health concerns and show for future research that incorporates deep learning models, natural language processing (NLP), and physiological sensing data such as EEG and ECG to improve prediction accuracy. They also highlight the significance of concerns in data privacy and security prior to large-scale deployment of such forecast algorithms.

Ayako Baba and Kyosuke Bunji [18] published an article titled "Prediction of Mental Health Disorders Using Annual Student Health Survey: Machine Learning Approach," in JMIR Mental Health (2023, 10: e42420). The study's goal was to anticipate Data from an annual health survey were used to assess mental health problems among students. done at a Japanese national university. The dataset contained replies from 3561 undergraduate students (out of 5690 polled) between 2020 and 2021. The study used logistic regression, elastic net, random forest, XGBoost, and LightGBM as machine learning models, with LightGBM coming as the best finisher.Logistic Regression (76.0%), Elastic Net (86.2%), Random Forest (83.3%), XGBoost (85.5%), and LightGBM (85.7%) provided the highest Matthews correlation coefficient (MCC) of 0.970 in predicting mental health problems in 2020 and 0.986 in predicting 2021 issues based on 2020 responses. The study examined the impact on various variables, like demographics, self-reported health conditions, and response time patterns. Although answering time did not have a significant effect on prediction accuracy, certain response behavior variables did. The most influential predictors of mental health problems were worry about campus life, future concerns, sleep length, academic stress, years of university study, and the number of same-sex students in a department.According to the research, "machine learning could be effectively used for early identification of mental health issues among students, enabling timely intervention through health service centers."Future enhancements were proposed, including the incorporation of "longitudinal data, more universities, and additional behavioral factors."

Dhanamma Jagli et al. [19]published their article, "Innovative Machine Learning Models for Student Mental Health Analysis," in Frontiers in Health Informatics (2024, Vol 13, Issue 3). The study explores the use of machine learning (ML) approaches to predict students' mental health difficulties based on behavioral and academic tendencies. The authors examine the effectiveness of multiple ML methods, including 'logistic regression, decision trees, support vector machines (SVM), and neural networks', in mental health classification. The study combines findings from three major studies: one on the effects of "online mobile gaming on mental health," another on the mental health of college students using machine learning models, and a third on "classification techniques for mental health analysis." SVM and neural networks beat other approaches in different circumstances, with SVM knowing gaming-related mental health concerns with a 91.68% accuracy. Logistic regression and decision trees had reasonable accuracy, however models that included extensive behavioral and academic data performed better. The paper focuses on issues such as data quality, privacy problems, and algorithmic fairness, emphasizing the importance of enhanced preprocessing, diversified datasets, and advanced ML approaches such as deep learning and ensemble methods. The authors conclude that machine learning has significant potential for early detection and personalized mental health interventions, advocating that future research prefer scalability, interdisciplinary collaboration, and ethical considerations to enhance the reliability of ML-based mental health support systems in educational settings

Barnali Sahu et al. [20] published their article, "Mental Health Prediction in Students Using Data Mining Techniques," in The Journal of Open Bioinformatics (2023, Vol. 16, e187503622307140). This study aims to predict mental health problems such as depression, relationships, and anxiety disorders among university students in Odisha, India. An online survey was conducted across three universities, producing 109 responses containing socio-demographic information as well as responses from the GAD-7 (General Anxiety Disorder), PHQ-9 (Patient Health Questionnaire), and the Social Connectedness Scale. Correlation analysis and regression approaches were used to assess relationships between parameters and predict mental health results. Results showed an elevated risk of depression (61.9%), with early-year students having higher rates. Social relationship and depression were found to be negatively related, but anxiety and depression were found to be positive correlated. Also, the study found a relationship among depression and academic achievement, with higher levels of depression being linked to lower academic performance. With an accuracy of 92.3% (training), 90.18% (testing), and 84.35% (validation), logistic regression with hyperparameter adjustment fared better than the other models evaluated, outperforming linear regression (70.03%). The results stress the value of early mental health interventions and recommend that colleges and universities place preventative measures and policies into effect to help students' mental health. Larger datasets, more behavioral variables, and sophisticated machine learning methods like deep learning could all be used in future research to enhance model performance.

Mohamed Drira, Sana Ben Hassine, Michael Zhang, and Steven Smith's [26]paper provides a systematic literature review (SLR) on the use of machine learning (ML) methods in student mental health (SMH) research, with a focus on ethical considerations such as privacy, transparency, interpretability, and fairness. The authors followed the PRISMA standards and analyzed papers published between 2015 and 2024, gathering data from six major academic databases, including PsycInfo, PubMed, IEEE Xplore, ACM Digital utilized ML techniques were supervised learning models like Support Vector Machines (SVM), logistic regression, and random forests, with deep learning models like CNNs and RNNs also used in some circumstances. The survey found that 67% of publications diagnosing mental health problems such as depression and stress. The most widely logistic regression, and random forests, with deep learning models like CNNs and RNNs also used in some circumstances. These models' performance was frequently assessed using accuracy (51%), recall (38%), and precision (30%), with k-fold cross-validation being the most popular validation method. The study discovered substantial ethical difficulties, as 65% of the assessed publications failed to address privacy concerns, 59% used black-box models with limited interpretability, and only 18% revealed demographic information, raising concerns about fairness. However, transparency was rather adequately addressed, with an average reporting score of 8/10 and a replicability score of 4/5. The authors underline the importance of increased ethical awareness in AI/ML-based SMH research, calling for better data privacy policies, the adoption of explainable AI techniques, and a balanced approach that blends high- performance ML models with interpretability and fairness concerns.

The paper "Student Mental Health Analysis Using Machine Learning Techniques" by Thejaswini M, Vivek A M, Shweta Y Bajantri, Shashank B S, and Shylaja B,[27 ] published in the International Journal of Research Publication and Reviews (Vol. 5, No. 12,December 2024) investigates the use of machine learning (ML) techniques for assessing student mental health. The study looks into the mental health profiles of university students and working professionals, including issues like academic pressure, financial stress, work pressure, and actions. The authors collected data from over 500 students and 2000 working professionals using surveys and socioeconomic questionnaires, which included validated mental health assessment measures such as the PHQ-9 and GAD-7. Various ML models, including cluster analysis, logistic regression, and random forest, were used to identify mental health concerns, and the system achieved 85% accuracy in detecting high-risk individuals. The study indicated that 52.15% of surveyed males and 47.85% of females reported depression, owing primarily to academic stress and job unhappiness. Furthermore, 60% of students expressed anxiety, with schoolwork, financial instability, and peer pressure being the most common stresses. The findings highlight the importance of early intervention measures, tailored mental health suggestions, and institutional support in reducing student stress. The authors propose future improvements, such as real-time mental health monitoring via wearable devices, deep learning approaches for increased accuracy, and cross-institutional cooperation to establish comprehensive mental health databases.

Maryam Sadiq [28 ]wrote the research paper Predicting Students' Mental Disorders Using Machine Learning for her graduate capstone at Rochester Institute of Technology (RIT) Dubai in 2022. The thesis attempted to use machine learning approaches to predict mental health issues in students, specifically stress, anxiety, and depression. The author conducted the study using the Depression Anxiety Stress Scales (DASS) dataset. This dataset was gathered from several internet sources between 2017 and 2019 and downloaded via Kaggle, a well-known repository for machine learning data. The dataset had 68 columns, which included demographic information and survey results. Before starting model training, Exploratory Data Analysis (EDA) was used to find missing values and anomalies in the data. Machine learning models used in this study included K- Nearest Neighbors (KNN), Naïve Bayes, Random Forest, and XGBoost. The dataset was divided among 70 percent training and 30 percent testing. Gradient Boosting (a boosting ensemble method) outperformed the other models across all categories. It had an accuracy of 92% for depression, 90% for stress, and 88% for anxiety prediction. Other models, including Random Forest, Naïve Bayes, and Decision Tree, showed lesser accuracy. The model's performance was evaluated based on accuracy, F1-score, recall, and precision.The study concluded that machine learning models, particularly ensemble methods such as Gradient Boosting, may accurately predict students' mental health status. The author proposed that future research may look at new artificial intelligence techniques, build continuous feedback loops, and create automated mental health assessment systems.

Ujunwa Madububambachu[29], Augustine Ukpebor, and Urenna Ihezue wrote the thesis titled "Machine Learning Techniques to Predict Mental Health Diagnoses: A Systematic Literature Review". The study examines multiple machine learning algorithms for predicting mental health issues in college students. The reviewed studies used structured and unstructured data from clinical records, electronic health records, neuroimaging data (fMRI, sMRI), social media (Twitter, Reddit), self-reported surveys, and physiological signals, with sample sizes ranging from 50 to more than 2.5 million data points. The study focuses on several machine learning models employed in mental health prediction, with Convolutional Neural Networks (CNN) achieving the best accuracy, particularly for bipolar disorder (99.75%). Other models tested were Support Vector Machine (SVM), Random Forest (RF), Logistic Regression (LR), Decision Trees (DT), Multi-Layer Perceptron (MLP), and Extreme Learning Machine (ELM). The study's drawbacks include a lack of consistent, high-quality datasets, issues with model interpretability, ethical concerns about data privacy, and biases in self-reported mental health data. Furthermore, the paper emphasizes the need of using longitudinal data to strengthen models and capture the temporal dynamics of mental health disorders. Despite these challenges, the study highlights machine learning's potential to improve early detection and intervention tactics in mental health diagnosis

Table has been created from the summary of the literature review:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Authors | Dataset | Methods Used | Results | Limitations |
| Hanif Abdul Rahman et al. (2023) | 15,366 university students from 17 ASEAN countries | GLM, KNN, NB, NN, RF, RPART, BAG, BOOST | Random Forest (92.1%), Adaptive Boosting (89.3%), KNN (77.5%), Naïve Bayes (72.3%), and Neural Networks (61.5%) | Limited to survey-based data; lacks physiological data integration |
| Ayako Baba & Kyosuke Bunji (2023 | 3,561 students from a national university in Japan | LR, EN, RF, XGB, LGBM | LightGBM (85.7% accuracy), followed by Elastic Net (86.2%), Random Forest (83.3%), and XGBoost (85.5%) | No inclusion of physiological or behavioral data like social media usage |
| Dhanamma Jagli et al. (2024) | Reviewed three previous studies on mental health | Logistic Regression, Decision Trees, SVM, Neural Networks | SVM achieved the highest accuracy (91.68%) in predicting gaming-related mental health issues | Study is a review; lacks experimental validation with new datasets |
| Barnali Sahu et al. (2023) | 109 university students from 3 colleges in Odisha, India | Logistic Regression, Linear Regression, Correlation Analysis, Chi-squared test | Logistic Regression achieved the highest accuracy (92.3% training, 90.18% testing, 84.35% validation). Depression prevalence: 61.9% | Small dataset; lacks generalizability beyond Odisha; limited feature set |
| Mohamed Drira, Sana Ben Hassine, Michael Zhang, Steven Smith | Six databases, surveys, sensors, social media | SVM, logistic regression, random forests, CNNs, RNNs | 67% used ML for prediction, 36% for diagnosis. Ethical concerns: privacy, black-box models, fairness. | 65% ignored privacy, 59% used black-box models, 18% lacked demographic data. |
| Thejaswini M, Vivek A M, Shweta Y Bajantri, Shashank B S, Shylaja B | 500 students, 2000 professionals, surveys, PHQ-9, GAD-7 | Cluster analysis, logistic regression, random forest | 85% accuracy, 52.15% males and 47.85% females had depression, 60% students had anxiety | No real-time monitoring or deep learning techniques |
| Maryam Sadiq | Depression Anxiety Stress Scales (DASS) dataset , collected from Kaggle | K-Nearest Neighbors (KNN), Naïve Bayes, Random Forest, XGBoost, Gradient Boosting | Gradient Boosting achieved highest accuracy: 92% for depression, 90% for stress, and 88% for anxiety | Limited dataset diversity, potential bias in model predictions, need for real-time implementation |
| Madububambachu, Ukpebor, Ihezue | Clinical records, neuroimaging , social media, surveys, physiological signals | CNN, SVM, RF, LR, DT, MLP, ELM | CNN (99.75% for bipolar), RF & CNN (99% for PTSD), ELM (AUC 0.8757 for ADHD) | Data quality issues, model interpretability, privacy concerns, bias, lack of longitudinal data |

1. **Objectives:**
2. Collect and preprocess student mental health data to ensure accuracy, consistency, and validity to machine learning models.
3. Train and evaluate a number of machine learning models (such as Random Forest, SVM, and Neural Networks) to find the best successful prediction method.
4. To examine the connection between academic stress, lack of money, and social support on the mental health of Bangladeshi students, with the goal of finding critical variables that lead to depression and anxiety in this population.[8]
5. **Methodology:**
   1. **Prototype Design**

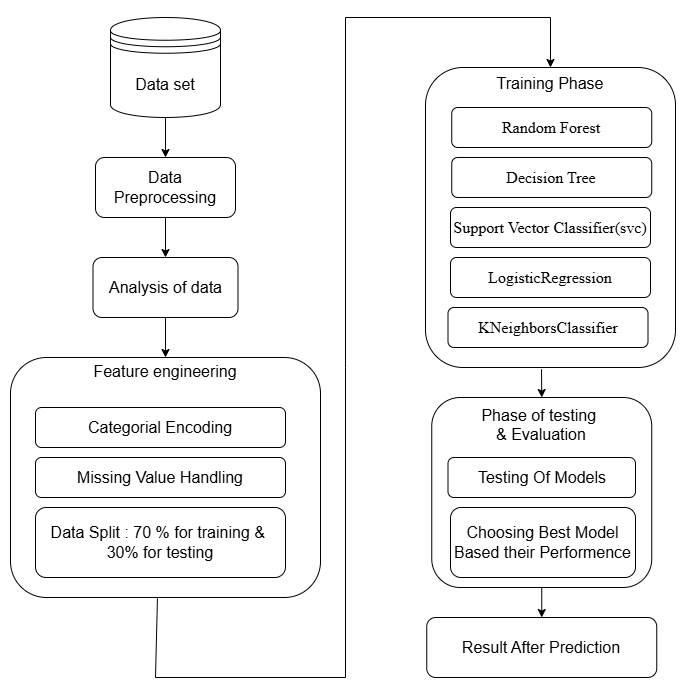


Fig 3.1.1: Data Flow Diagram

* 1. **Data Collection**

-The data for this study was collected using a Google Spreadsheet titled "Student Mental Health". A standard survey was developed to gather via information about students' demographics, academic performance, financial stress, sleep patterns, and overall mental health. The spreadsheet The URL was given with Daffodil International University students, allowing them to access and complete their input directly. This method enabled an easy and quick gathering of data processes while conserving accuracy and completeness.The answers were slowly recorded in a Google Spreadsheet, leading to an ordered and structured style. After obtaining the data, the spreadsheet was exported as a “CSV file”  for further processing and analysis. The dataset includes important variables such as gender, age, CGPA, academic pressure, financial stress, sleep duration, and overall student mental health status, giving a good foundation for using machine learning techniques to predict mental health issues among students.

* 1. **Analysis Techniques**

- For Analysis data first we have to apply data preprocessing technique. Data preparation is a critical stage in the machine learning pipeline. It entails converting raw data to a format that is more suitable for analysis and modeling. The main goal of preprocessing is to clean and organize data, removing inconsistencies and unnecessary data. Common tasks include dealing with values that are missing, encoding categorical variables, scaling numerical features, and separating the data into training and testing groups. Proper preprocessing improves the performance and accuracy of machine learning models.Preprocessing Techniques Used :

1. **Label Encoding**:Label Encoding is a technique used to convert categorical data into numerical values by assigning a unique integer to each category. It is commonly used for machine learning models that require numerical input.[12]
2. **Feature Selection**: Feature selection is an important step in machine learning and involves selecting a subset of relevant features from the original feature set in an attempt to reduce the feature space and boost the model's performance by reducing processing power. It's a significant phase in machine learning, particularly while working with large quantities of data..[13]

After using preprocessing we applied our pretrained models. Pretrained models are machine learning models that were previously trained on large data sets to find general patterns and characteristics. These models may be fine-tuned or used directly to specific tasks, significantly reducing training time and processing power. In the context of student mental health prediction. pretrained models used for this thesis are:

**1.Random Forest:** Random forests, also known as random decision forests, is a method of ensemble learning for classification, regression, and other tasks with a variety of decision trees during training.[9]

For a given input x, a Random Forest's prediction can be shown here:

ŷ = (1/M) \* Σ(T\_m(x))

Where ,

ŷ represents the predicted value (for regression) or class (for classification).

M represents the number of trees in the forest.

T\_m(x) represents the prediction of the m-th tree in the forest for input x. [21].

**2.Decision Tree:** A decision tree is a decision support recursive partitioning structure that utilizes a tree-like representation of decisions and their potential consequences, including chance event outcomes, resource costs, and utility. It is one method to show an algorithm using just conditional control statements.[10-11]

The prediction function of a decision tree is written as:

ŷ = f(x; θ)

Where:

The predicted value is shown as ŷ.

The feature input vector is denoted as x.

The decision tree's parameters, denoted by θ, include the learnt splitting rules and node values.

The function f shifts input x to prediction ŷ using the learnt tree structure.[22]

**3.Support Vector Classifier(svc):**Support vector machines (SVMs) are supervised learning algorithms used in machine learning for solving classification and regression problems. SVMs are highly efficient at solving binary classification problems, which require classifying data elements into two groups.[14]

The prediction function of a Support Vector Classifier is written as:

ŷ = sign(w ⋅ x + b)

Where:

ŷ is the predicted class label (+1 or -1).

x is the input vector of features.

w is the weight vector, representing the learned coefficients of the features.

b is the bias (or intercept) term.

sign() is the sign function, which returns +1 if the argument is positive and -1 if it's negative.[23]

**4.LogisticRegression:** Logistic regression is a supervised machine learning technique used in classification issues to predict either an instance belongs to a provided class or not. Logistic regression is a statistical procedure that analyzes the relationship between two data variables. [15]

A logistic regression's prediction function is written in the form of:

p(y = 1 | x; θ) = σ(θᵀx) = 1 / (1 + exp(-θᵀx))

Where,

p(y = 1 | x; θ): This represents the probability that the output y is 1, given the input x and the model parameters θ

x: This is the input vector of features.

 θ: This is the vector of model parameters (coefficients).

x: This is the input vector of features.

 θ: This is the vector of model parameters (coefficients).

θᵀ is the transpose of the parameter vector θ. σ(z): This is the sigmoid function, which is written as 1 / (1 + exp(-z)). It turns any real number to a value between 0 and 1, making it ideal for presenting probabilities.[24]

**4.KNeighborsClassifier:**KNeighborsClassifier is an algorithm that effectively categorizes data points using the patterns discovered in those points' close to data points, or neighbors.[16].K-Nearest Neighbors is also known as a lazy learner algorithm since it does not learn from the training set right away. Instead, it stores the dataset and performs an action on it when it comes time to classify.[17]

A KNeighborsClassifier prediction function is written in the following format:

ŷ = argmax\_{c} ∑\_{x\_i ∈ N\_k(x)} I(y\_i = c)

Where:

ŷ is the predicted class label for the input x.

x: is the input vector for which we want to predict the class.

N\_k(x): represents the set of the k-nearest neighbors of x in the training data. These neighbors are determined based on a distance metric (e.g., Euclidean distance).

y\_i: is the true class label of the i-th neighbor x\_i.

c: represents a possible class label.

I(condition): is the indicator function, which returns 1 if the condition is true and 0 if it's false.

argmax\_{c}: This means we choose the class c that maximizes the sum. In other words, we count how many neighbors belong to each class and choose the most frequent class.[25]

1. **Progress Achieved:**
   1. **Completed Tasks**

**•** Title Selection

• Reference Paper Collection

• Paper Review

• Data Collection

* 1. **Results Obtained**

- Provide a summary of any preliminary results or findings obtained.

1. **Challenges Faced:**

|  |  |  |
| --- | --- | --- |
| **S.No.** | **Issues and Challenges** | **Strategies or Plans** |
| 1 | Title Selection | I check the most popular websites to see the newest subjects before choosing a title. |
| 2 | Review Paper Collection | I used Sci Hub to collect them. |
| 3 | Dataset Collection | For me, it was a huge task. I gathered information by surveying students at my university. |
| 4 | Finding relevant elements for prediction was challenging due to the vast number of factors influencing stress and addiction. | To find the most crucial components and improve model accuracy, we used feature selection techniques. |

1. **Next Steps:**

- Outline the tasks and milestones planned for the next phase of the project.

|  |  |  |
| --- | --- | --- |
| **S.No.** | **Next Task** | **Estimate completion time**  **(MM-YY)** |
| 1 |  |  |
| 2 |  |  |
| 3 |  |  |
| 4 |  |  |

1. **Updated Timeline:**

- Provide an updated timeline, highlighting progress made and indicating any adjustments.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Tasks** | **Weeks** | | | | | | | | | | | | | | | | | |
|  | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 |
| Task-1 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Task-2 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Task-3 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Task-4 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |

|  |  |
| --- | --- |
| **Estimated Work Period** |  |
| **Actual Work Period** |  |

1. **Resources Utilized:**

- Detail the resources (materials, equipment, software, etc.) used during this reporting period.

1. **Project Management and Financial Analysis:**

* Discuss any estimated cost and project management of your project.

1. **Future Considerations:**

- Discuss any potential issues or considerations that may impact the project in the Next phase.

1. **Conclusion:**

- Summarize the key points discussed in the progress report.

**References**

1.World Health Organization, “Mental health,” *World Health Organization*, Jun. 17, 2022. https://www.who.int/news-room/fact-sheets/detail/mental-health-strengthening-our-response

2.S. BH *et al.*, “Mental Health Analysis of Employees using Machine Learning Techniques,” *2022 14th International Conference on COMmunication Systems & NETworkS (COMSNETS)*, Jan. 2022, doi: https://doi.org/10.1109/comsnets53615.2022.9668526.

3.The daily star report :  https://www.thedailystar.net/opinion/views/news/counselling-urgent-need-university-students-3294481

4. Dhaka Trubune Report:

<https://www.dhakatribune.com/bangladesh/313212/survey-86%25-of-students-suffer-from-mental-health>

5. Shahidul Islam Khan, A. Islam, Akther Hossen, Taiyeb Ibna Zahangir, and Abu, “Supporting the Treatment of Mental Diseases using Data Mining,” Oct. 2018, doi: https://doi.org/10.1109/iciset.2018.8745591.

6. A. Parekh, “Mental Health Prediction Using Machine Learning,” *Analytics Vidhya*, Jun. 10, 2022. https://www.analyticsvidhya.com/blog/2022/06/mental-health-prediction-using-machine-learning/

7.*Healthinformaticsjournal.com*, 2025. https://www.healthinformaticsjournal.com/index.php/IJMI/article/view/162/162 (accessed Feb. 19, 2025).

8.Muhammad Azizur Rahman and T. Kohli, “Mental health analysis of international students using machine learning techniques,” *PloS one*, vol. 19, no. 6, pp. e0304132–e0304132, Jun. 2024, doi: https://doi.org/10.1371/journal.pone.0304132.

9.“Random forest algorithm in machine learning,” *GeeksforGeeks*, Jul. 12, 2024. https://www.geeksforgeeks.org/random-forest-algorithm-in-machine-learning/

10.Wikipedia Contributors, “Decision tree,” *Wikipedia*, Sep. 28, 2019. https://en.wikipedia.org/wiki/Decision\_tree

11.Mahendramedapati, “Understanding Decision Tree Algorithm - Mahendramedapati - Medium,” *Medium*,Jun.20,2024. https://medium.com/@mahendramedapati.r469/understanding-decision-tree-algorithm-6f601425b24c

12.S. Kumar, “What is label encoding? Application of label encoder in machine learning and deep learning models.,” *Medium*, Jan. 12, 2024. https://medium.com/@sunnykumar1516/what-is-label-encoding-application-of-label-encoder-in-machine-learning-and-deep-learning-models-c593669483ed

13. “Feature Selection Techniques in Machine Learning,” *GeeksforGeeks*, Jan. 19, 2021. https://www.geeksforgeeks.org/feature-selection-techniques-in-machine-learning/

14]A. Sasidharan, “Support Vector Machine Algorithm,” *GeeksforGeeks*, Jan. 20, 2021. https://www.geeksforgeeks.org/support-vector-machine-algorithm/

15.GeeksforGeeks, “Understanding Logistic Regression,” *GeeksforGeeks*, May 09, 2024. https://www.geeksforgeeks.org/understanding-logistic-regression/

16.“Scikit Learn - KNeighborsClassifier,” *www.tutorialspoint.com*. https://www.tutorialspoint.com/scikit\_learn/scikit\_learn\_kneighbors\_classifier.htm

17. Hanif Abdul Rahman *et al.*, “Machine Learning-Based Prediction of Mental Well-Being Using Health Behavior Data from University Students,” *Bioengineering*, vol. 10, no. 5, pp. 575–575, May 2023, doi: <https://doi.org/10.3390/bioengineering10050575>.

18. A. Baba and K. Bunji, “Prediction of Mental Health Problem Using Annual Student Health Survey: A Machine Learning Approach (Preprint),” *JMIR Mental Health*, Sep. 2022, doi: https://doi.org/10.2196/42420.

19. D. Jagli, M. Talakoti, R. Solanki, N. Naik, K. Jamdaade, and R. Naik, "Innovative machine learning models for student mental health analysis," *Frontiers in Health Informatics*, vol. 13, no. 3, pp. 1270-1281, 2024.

20.Sahu B, Kedia J, Ranjan V, Mahaptra B, Dehuri S. Mental Health Prediction in Students using Data Mining Techniques. Open Bioinform J, 2023; 16: e187503622307140. <http://dx.doi.org/10.2174/18750362-v16-230720-2022-19>

21.[1]L. Breiman, “Random Forests,” *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001, doi: <https://doi.org/10.1023/a:1010933404324>.

22. L. Breiman, J. H. Friedman, R. A. Olshen, and C. J. Stone, *Classification And Regression Trees*. Routledge, 2017. doi: <https://doi.org/10.1201/9781315139470>.

23.“Cortes, C. and Vapnik, V. (1995) Support-Vector Networks. Machine Learning, 20, 273-297. - References - Scientific Research Publishing,” *Scirp.org*, 2014. <https://www.scirp.org/reference/referencespapers?referenceid=1150668>

24.T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*. New York, NY: Springer New York, 2009. doi: <https://doi.org/10.1007/978-0-387-84858-7>.

‌25. “IEEE TRANSACTIONS.” Available: <https://isl.stanford.edu/~cover/papers/transIT/0021cove.pdf>

1. M. Drira, S. Ben Hassine, M. Zhang, and S. Smith, "Machine Learning Methods in Student Mental Health Research: An Ethics-Centered Systematic Literature Review," Applied Sciences, vol. 14, no. 24, p. 11738, Dec. 2024. doi: [10.3390/app142411738](https://doi.org/10.3390/app142411738).
2. T. M, V. A. M, S. Y. Bajantri, S. B. S, and S. B, "Student Mental Health Analysis Using Machine Learning Techniques," International Journal of Research Publication and Reviews, vol. 5, no. 12, pp. 5608-5615, Dec. 2024.
3. M. Sadiq, “Predicting Students’ Mental Disorders Using Machine learning,” *RIT Digital Institutional Repository*, 2022. https://repository.rit.edu/theses/11413/ (accessed Feb. 21, 2025).
4. **U. Madububambachu, A. Ukpebor, and U. Ihezue,** "Machine Learning Techniques to Predict Mental Health Diagnoses: A Systematic Literature Review," Clin. Pract. Epidemiol. Ment. Health, vol. 20, pp. e17450179315688, Jul. 2024, doi: 10.2174/0117450179315688240607052117.

‌

**Appendix**

- [Additional documentation, charts, graphs, or technical details]

**FINAL YEAR DESIGN PROJECT**

**PHASE-I PROGRESS REPORT**

This report, in the form of a template, has been specifically designed for BSc. students working on their Final Year Design Project (FYDP) at Computer Science and Engineering Department, Daffodil International University (DIU).

Every group of students is required to do the following:

1. Complete all the sections of this template
2. Get it certified by the assigned supervisor before one week of Phase-I evaluation presentations
3. Submit 01 photocopy to each of the following, on or before the day of Phase-I presentations:
   1. Supervisor
   2. Internal Evaluator
4. Submit original copy to FYDP committee on the day of Phase-I presentations.

**Note:**

1. Use English
2. There should be NO grammatical or spelling mistakes
3. Submission after due date will not be accepted
4. For more information, contact your Supervisor

|  |  |
| --- | --- |
| **Template prepared by:**  **FYDP Committee**  **Dept. of CSE, DIU** | **Template approved by:**  **Dr. Sheak Rashed Haider Noori**  **Professor and Head, Dept. of CSE, DIU** |

The students and faculty members of Computer Science and Engineering Department, Daffodil International University have full access rights to read and print this document without any prior notice to the Head and FYDP committee.

All rights are reserved.