Green University of Bangladesh
Department of Computer Science and Engineering(CSE)

Faculty of Sciences and Engineering
Semester: (Spring, Year: 2026), B.Sc. in CSE (Day)

Project Proposal

Course Title: Machine Learning Lab
Course Code: CSE 412 Section: 231_D4

Student Details

Name	ID
Anisur Rahaman Maruf	222902078
Sakibul Hasan	222902083

Submission Date: 03/03/2025
Course Teacher's Name: Mayeesha Farjana

[For Teachers use only: Don't Write Anything inside this box]

Lab Report Status

Marks: …… **Signature:.....**

Comments:..... **Date:.....**

1. Project Title

Heart Disease Prediction System Using Machine Learning

2. Project Overview

This project is a data-driven healthcare application designed to predict the likelihood of heart disease in patients. By analyzing clinical parameters like age, blood pressure, cholesterol, and heart rate, the system identifies whether a patient is at "Risk" or "Healthy." It serves as a decision-support tool for medical professionals to provide early warnings and potentially save lives.

3. Problem Statement & Significance

Heart disease is one of the leading causes of mortality worldwide. Early detection is often difficult due to the complex interaction of various risk factors. This project helps by:

- **Early Detection:** Identifying high-risk patients before the onset of severe symptoms.
- **Accuracy:** Reducing human error in diagnosis by using trained computational models.
- **Efficiency:** Providing instant results based on patient data, allowing doctors to prioritize emergency cases.

4. Dataset Description

The project utilizes a structured dataset (CSV format) containing 14 key clinical attributes, including:

- **Demographics:** Age, Sex.
- **Clinical Measurements:** Resting blood pressure (**`trestbps`**), Serum cholesterol (**`chol`**), Maximum heart rate (**`thalach`**).
- **Diagnostic Indicators:** Chest pain type (**`cp`**), ST depression (**`oldpeak`**), and Number of major vessels (**`ca`**).
- **Target:** A binary label where **1** indicates Heart Disease (Risk) and `0` indicates No Disease (Healthy).

5. Methodology & Algorithms

To ensure the highest reliability, this project implements and compares four robust Machine Learning algorithms:

1. **Logistic Regression:** Used for binary classification to establish a baseline probability.
2. **K-Nearest Neighbors (KNN):** To classify patients based on their similarity to existing data points.
3. **Support Vector Machine (SVM):** To create an optimal hyperplane that separates healthy and high-risk patients.
4. **Decision Tree:** To visualize the decision-making rules based on clinical thresholds.

6. Implementation Workflow

- **Data Preprocessing:** Handling categorical variables through One-Hot Encoding and normalizing numerical features using **Standard Scaling**.
- **Model Training:** Training the algorithms on 80% of the dataset while keeping 20% for independent testing.
- **User Interface:** A custom interface where users can input real-time patient data to get an instant prediction.

7. Expected Final Results and Performance Evaluation

The final outcome of this research will be a comparative diagnostic dashboard that evaluates patient health through four distinct analytical lenses. Beyond just predicting "Healthy" or "Risk," the project will deliver:

- **Comparative Reliability:** By displaying results from **Logistic Regression, KNN, SVM, and Decision Trees** side-by-side, the system provides a cross-verified result. For instance, if all four models predict "Risk," the diagnostic confidence is significantly higher than a single-model prediction.
- **Performance Metrics:** We will evaluate the models using a **Confusion Matrix**, focusing on **Recall (Sensitivity)**. In medical science, missing a "Risk" case (False Negative) is more dangerous than a false alarm. Our goal is to maximize the Recall rate to ensure every potential heart patient is flagged.
- **Feature Significance Analysis:** The system will identify which clinical factors (e.g., Maximum Heart Rate or Chest Pain Type) contributed most to a specific patient's risk profile, providing interpretability to the medical results.

- **Real-time Decision Support:** The final system will allow healthcare providers to input live clinical data and receive an instantaneous risk assessment, significantly reducing the waiting time for preliminary diagnosis.

8. Conclusion and Future Scope

This project demonstrates the powerful integration of machine learning into the healthcare domain. By leveraging a multi-algorithm approach, we have created a robust tool that can handle the complexity of clinical data and translate it into actionable medical insights.

Conclusion: The proposed system effectively bridges the gap between raw patient data and early clinical intervention. While traditional methods rely on manual interpretation of test results, our automated system provides a standardized, unbiased, and data-driven assessment. This not only assists junior doctors in their decision-making process but also acts as a primary screening tool in remote areas where specialists are unavailable.

Future Scope: To further enhance the project, we plan to:

1. **Deployment:** Convert the model into a web-based application (using Flask or Streamlit) or a mobile app for easier accessibility.
2. **Expanded Dataset:** Incorporate more diverse datasets to improve the model's generalization across different ethnicities and age groups.
3. **Advanced Algorithms:** Explore Deep Learning (Neural Networks) and Ensemble methods like Random Forest to further increase prediction accuracy.