Phase-3

Student Name:Murshitha M

Register Number: 620123106070

Institution: AVS Engineering College

Department: ECE

Date of Submission:18-05-2025

Github Repository Link:https

://github.com/Murshitha06/Webproje

ct.git

---

1. Problem Statement

Early and accurate disease prediction is a critical aspect of improving patient outcomes and reducing healt

2. Abstract

This project leverages patient health records to develop predictive models capable of identifying potential

3. System Requirements

Hardware: 8GB RAM, Intel i5 or equivalent
Software: Python 3.9+, Jupyter/Colab, pandas, numpy, matplotlib, seaborn, scikit-learn, streamli

4. Objectives

- Predict disease likelihood

- Improve diagnosis efficiency

- Enhance decision-making in healthcare

- Deploy a real-time prediction tool

5. Flowchart of Project Workflow

[Insert your custom workflow image here]

6. Dataset Description

Source: Kaggle
Type: Public
Structure: ~10,000 records, 15 features

7. Data Preprocessing

- Missing values imputed
- Duplicates removed
- Features scaled and encoded

8. Exploratory Data Analysis (EDA)

- Correlation heatmaps
- Histograms and boxplots
- Key insight: Glucose levels and age correlated with disease

9. Feature Engineering

- Created BMI category
- Feature importance used for selection

10. Model Building

Models: Logistic Regression, Random Forest, XGBoost
Best: Random Forest

11. Model Building

Metrics: Accuracy 88%, F1 Score 0.86
Visuals: Confusion matrix, ROC curve

12. Model Building

Platform: Streamlit Cloud
Public Link: [Add Streamlit link]

13. Model Building

Available on GitHub repository.

14. Model Evaluation

- *Integrate with live EHR systems*

- *Multi-label disease prediction*

- *Incorporate wearable sensor data*

15. Team Members and Roles:

S.Thirulochine: Problem statement, EDA, model training, deployment
V.Sandhiya: Data preprocessing, visualizations, feature engineering
M.Murshitha: Abstract, documentation, evaluation & presentation