Software Design Document

for

Model to Diagnose Cardiac Attack

Version <1.0>

Prepared by

Group Members:

Devesh Gupta	**BT22GCS264**	**devesh.gupta22@st.niituniversity.in**
Guni Sharma	**BT22GCS070**	**guni.sharma22@st.niituniversity.in**
Raajit J Singh	**BT22GCS051**	**raajit.singh22@st.niituniversity.in**
Yash Upadhyay	**BT22GCS355**	**yash.upadhyay22@st.niituniversity.in**

Instructor: Manish Hurkat

Course: Capstone

Lab Section: C2

Date: 20 Feb'25

Table of Contents

1. Introduction

1.1 Purpose

This document outlines the design and implementation details for a cardiac attack prediction model using machine learning algorithms.

1.2 Scope

The project aims to predict cardiac attacks based on various health metrics using multiple machine learning models and integrate the model into a web application.

1.3 Definitions, Acronyms, and Abbreviations

SDD: Software Design Document

ML: Machine Learning

CSV: Comma-Separated Values

MSE: Mean Squared Error

2. System Overview

The system is designed to predict the likelihood of a cardiac attack using data preprocessing, feature selection, and multiple machine learning models.

3. Data Preprocessing

3.1 Data Loading

The dataset is loaded from a CSV file (heart_disease.csv).

3.2 Handling Missing Values

Missing values in the 'education' column are filled with 'Unknown'.

Missing values in the 'Gender' column are filled with the mode value.

Missing values in the 'glucose' column are forward-filled and backward-filled.

Remaining missing values are dropped.

3.3 Data Transformation

'Gender' is converted to binary values (Male: 1, Female: 0).

'Heart_stroke' is converted to binary values (Yes: 1, No: 0).

'prevalentStroke' is converted to binary values (No: 0, Yes: 1).

4. Model Training

4.1 Random Forest Regressor

The dataset is split into training and testing sets.

A Random Forest Regressor is trained on the training set.

The model's training accuracy, Mean Squared Error (MSE), and R-squared are calculated.

4.2 Logistic Regression

The Logistic Regression model is trained on the training set.

The model's training accuracy, MSE, and R-squared are calculated.

4.3 K-Nearest Neighbors (KNN) Classifier

The KNN Classifier is trained on the training set.

The model's accuracy, MSE, and R-squared are calculated.

4.4 Gaussian Naive Bayes

The Gaussian Naive Bayes model is trained on the training set.

The model's accuracy, MSE, and R-squared are calculated.

5. Model Evaluation

The performance of each model is evaluated using metrics such as training accuracy, MSE, and R-squared.

6. Integration and Deployment

The final model is selected based on performance metrics.

The model is integrated into a web application using Flask/Django.

The web application allows users to input health metrics and receive predictions for cardiac attacks.

7. Conclusion

This document outlines the design and implementation details for a cardiac attack prediction model. The project includes data preprocessing, model training, evaluation, and integration into a web application.