Abstract: Heart Disease Prediction Using Machine Learning

Heart disease is a leading cause of death globally, emphasizing the need for early and accurate diagnosis to improve patient outcomes. This project aims to develop a machine learning-based predictive model to identify the presence of heart disease in patients using various clinical parameters. By leveraging data science techniques, the goal is to assist healthcare professionals in making faster and more reliable decisions.

The problem statement centers around the challenge of detecting heart disease early using non-invasive data. Traditional diagnostic methods can be time-consuming and costly. Our solution uses machine learning algorithms to analyze patient data and predict the likelihood of heart disease. The dataset used in this study is the Cleveland Heart Disease dataset, which includes features such as age, sex, chest pain type, cholesterol levels, and more.

The tools and technologies employed in this project include Python, along with libraries such as Pandas, NumPy, Matplotlib, Seaborn, and Scikit-learn. For optional deployment and user interaction, Streamlit or Flask can be used to build a web-based application.

The project consists of the following sub-modules:

1. Data Collection and Preprocessing - Loading the dataset, handling missing values, and normalizing features.
2. Exploratory Data Analysis (EDA) - Visualizing feature correlations and distributions to understand data patterns.
3. Model Building and Training - Implementing machine learning algorithms such as Logistic Regression, Random Forest, and Support Vector Machine.

4. Model Evaluation - Measuring performance using metrics like accuracy, precision, recall, and confusion matrix.

5. Prediction Interface (optional) - Building a user-friendly interface for real-time heart disease prediction using user input.

The design flow of the project begins with raw data input, followed by data cleaning, exploratory analysis, model training, evaluation, and finally prediction output. The architecture ensures modularity and ease of understanding for both developers and users.

The expected outcome of this project is a trained model that can accurately predict whether a patient has heart disease based on input parameters, achieving an accuracy of approximately 85-90%. The solution offers a scalable and cost-effective tool to support early diagnosis and reduce the burden on healthcare systems.