1. Define Problem Statement

1.1 Background

Cardiovascular diseases (CVD) are a significant global health concern, contributing to high mortality rates. Analyzing datasets related to heart diseases can offer valuable insights into the factors influencing cardiovascular conditions, facilitating improved prevention and treatment strategies.

1.2 Problem Statement

The primary objective of this project is to conduct a comprehensive analysis of a heart disease dataset, aiming to identify patterns, risk factors, and potential predictors of cardiovascular diseases. The insights derived from this analysis will be instrumental in developing strategies for early detection and prevention.

1.3 Objectives

Identify key features in the heart disease dataset.

Explore correlations and patterns within the data.

Develop predictive models for early detection.

Provide actionable insights for healthcare professionals.

2. Project Plan and Product Backlog

2.1 Project Plan

2.1.1 Project Scope

The scope of the project encompasses the analysis of specific variables within the heart disease dataset. Potential challenges and limitations are identified to ensure a realistic and achievable analysis.

2.1.2 Milestones

**Milestone 1:** Data Collection and Preprocessing,Define data collection sources.

Clean and preprocess the dataset (handle missing values, outliers, etc.).

**Milestone 2:** Exploratory Data Analysis (EDA),Perform descriptive statistics.

Visualize relationships between variables.Identify potential predictors of heart diseases.

**Milestone 3:** Model Development and Training,Select appropriate machine learning models.

Split the dataset into training and testing sets.Train and validate models.

**Milestone 4:** Results Interpretation and Documentation,Analyze model performance.

Document findings and insights.Prepare a summary report for stakeholders.

2.1.3 Timeline

**Week 1-2:** Data collection and preprocessing.

**Week 3-4:** Exploratory data analysis.

**Week 5-6:** Model development and training.

**Week 7-8:** Results interpretation and documentation.

2.2 Product Backlog

2.2.1 Data Collection

**Task 1:** Identify and obtain the heart disease dataset.

**Task 2:** Clean and preprocess the dataset (handle missing values, outliers, etc.).

2.2.2 Exploratory Data Analysis (EDA)

**Task 1:** Perform descriptive statistics on the dataset.

**Task 2:** Visualize relationships between variables.

**Task 3:** Identify potential predictors of heart diseases.

2.2.3 Model Development

**Task 1:** Select appropriate machine learning models.

**Task 2:** Split the dataset into training and testing sets.

**Task 3:** Train and validate models.

2.2.4 Results Interpretation and Documentation

**Task 1:** Analyze model performance.

**Task 2:** Document findings and insights.

**Task 3:** Prepare a summary report for stakeholders.