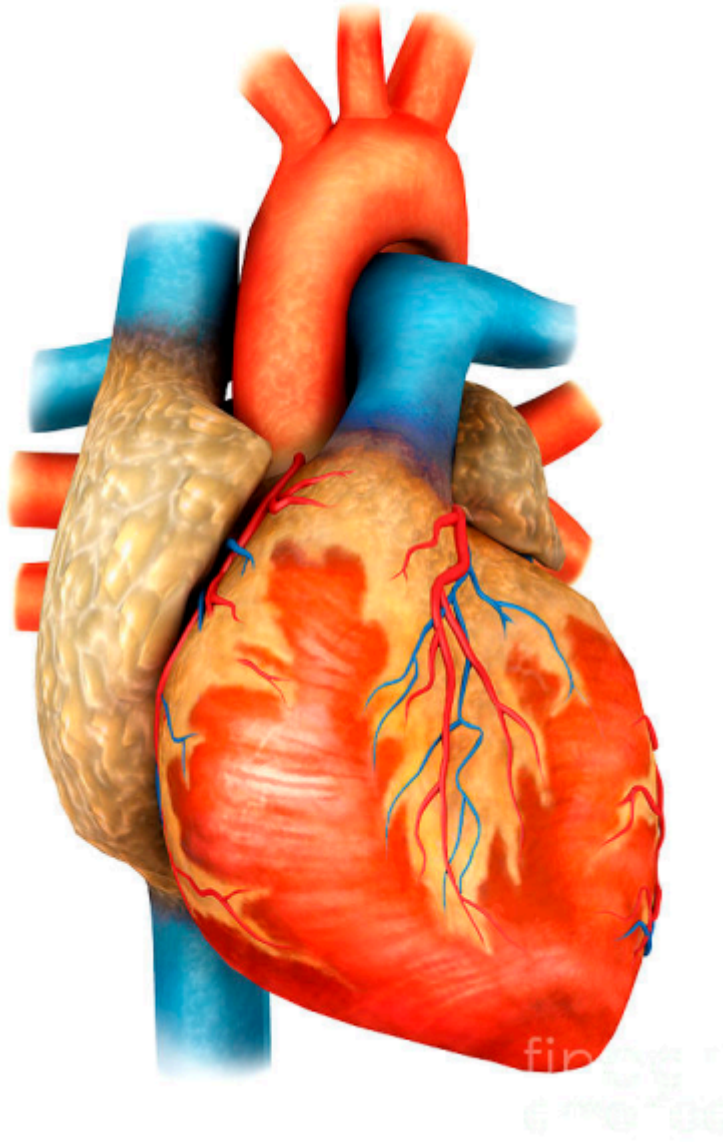# Project Proposal

Heart Disease Prediction with Logistic Regression

Team Members: Emu Negussie ,Sewnet Gebremichael, Meku Gela

Introduction

Heart disease is a leading cause of mortality worldwide, and early detection plays a crucial role in preventing adverse outcomes. Logistic regression is a widely used supervised learning algorithm for binary classification tasks that provides probabilistic predictions and interpretable results. It is a promising tool for heart disease prediction.This project aims to develop and evaluate a logistic regression model to predict the likelihood of heart disease based on various risk factors.

Objectives

1. To get and preprocess a comprehensive dataset containing relevant features for heart disease prediction.
2. To develop a logistic regression model to predict the likelihood of heart disease.
3. To evaluate the performance of the logistic regression model using appropriate metrics.
4. To compare the results with existing predictive models for heart disease.

Dataset

For this project we will use a dataset callecd heart failure clinical_records dataset from kaggle web site that includes features such as age, gender, blood pressure, cholesterol levels, and other relevant variables for heart disease prediction.

Analysis of the dataset is split into two parts.

1. **Exploratory Analyses**: In the exploratory analyses part, it is first ensured that there are no missing values in the data. Then, summary statistics of all the variables are analyzed. Bivariate analyses between the dependent and independent variables are conducted and plotted.
2. **Predictive Modeling**: In the predictive analysis part, the dataset is first split into train and test sets such that a random 20% of the data is captured in the test set and the rest are used to train the model.

Methodology

1. **Importing Libraries:** We first import the various libraries such as scikit-learn and NumPy for predict heart failure based on selected features.

2. **Data Preprocessing:** Perform data cleaning,handling missing values,dealing with outliers, feature selection, and normalization to prepare the dataset for model development.

3. **Exploratory Data Analysis(EDA):** Figure out any pairwise association or variable distributions.Chart and visualization,correlation factors and factor plots.

4. **Develop the predictive ML Model:**

   - Split the data in to a training set and test set with an 80-20 split.
   - Build a logistic regression model. using tools like Python and libraries such as scikit-learn and NumPy to predict heart failure based on selected features.

5. **Model Evaluation:**

   - Assess the performance of the logistic regression model using metrics such as accuracy, precision, recall,f1-score and support.

- Interpreting the coefficients of the logistic regression model to understand the impact of predictor variables on the outcome.

6. **Comparison:** Compare the results of the logistic regression model with other machine learning algorithms commonly used for heart disease prediction if relevant.

Timeline

- **Day 1-2:** Data collection and preprocessing.
- **Day 3-4:** Model development and evaluation.
- **Day 5:** Comparison with existing models and finalizing results.
- **Day 6:** Report writing and documentation.

Expected Outcome

The project is expected to deliver a well-performing logistic regression model for heart disease prediction, demonstrating its efficacy in the healthcare domain. The results will be documented in a research report outlining the methodology, findings, and implications for future research.