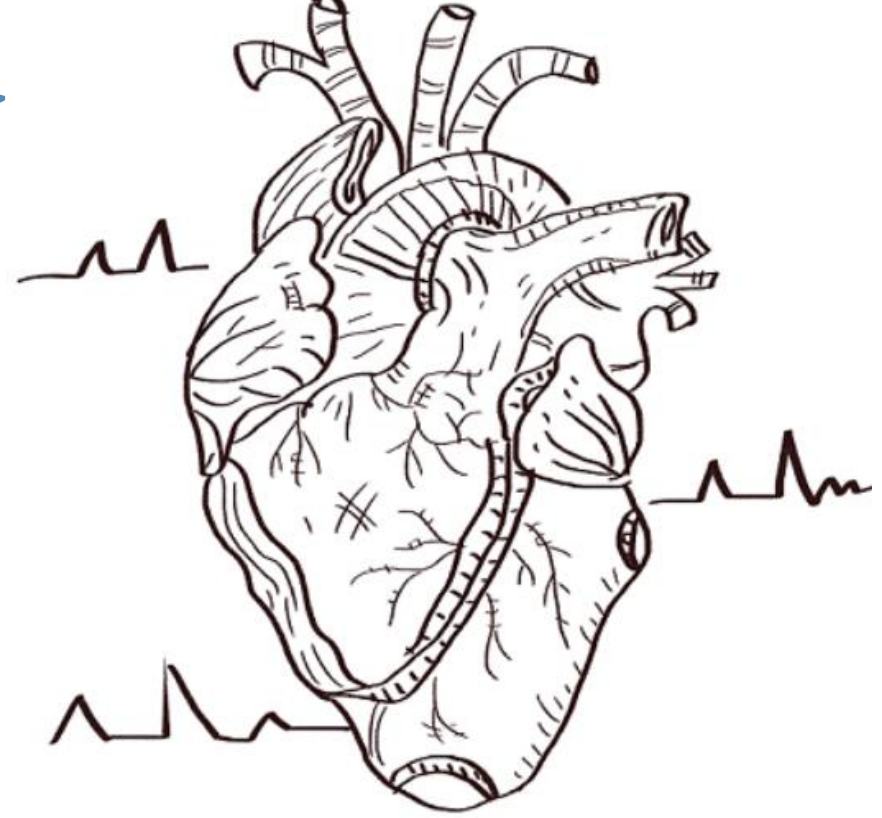# Heart Disease Prediction

GROUP-2

- PRIYANKA AKULA
- SOWJANYA PAMULAPATI
- SOMNA SATTOOR
- SHREYA CHIKATMARLA

Contents

Introduction

Heart disease is a significant health concern worldwide, and early detection plays a crucial role in improving patient outcomes.

In this project, we aim to utilize the power of visualizations and also develop a predictive model that can identify individuals at risk of developing heart disease.

By combining data analysis techniques with effective visual representations, we can enhance our understanding of the factors contributing to heart disease and provide valuable insights for healthcare professionals and patients alike.

Objectives

Investigating the relationship between specific attributes(such as sex, age, mental health, physical health)and the risk of heart disease.

Examining the impact of doing physical activity on heart disease risk.

Develop a machine learning model that can predict the likelihood of a heart attack based on different attributes from dataset.
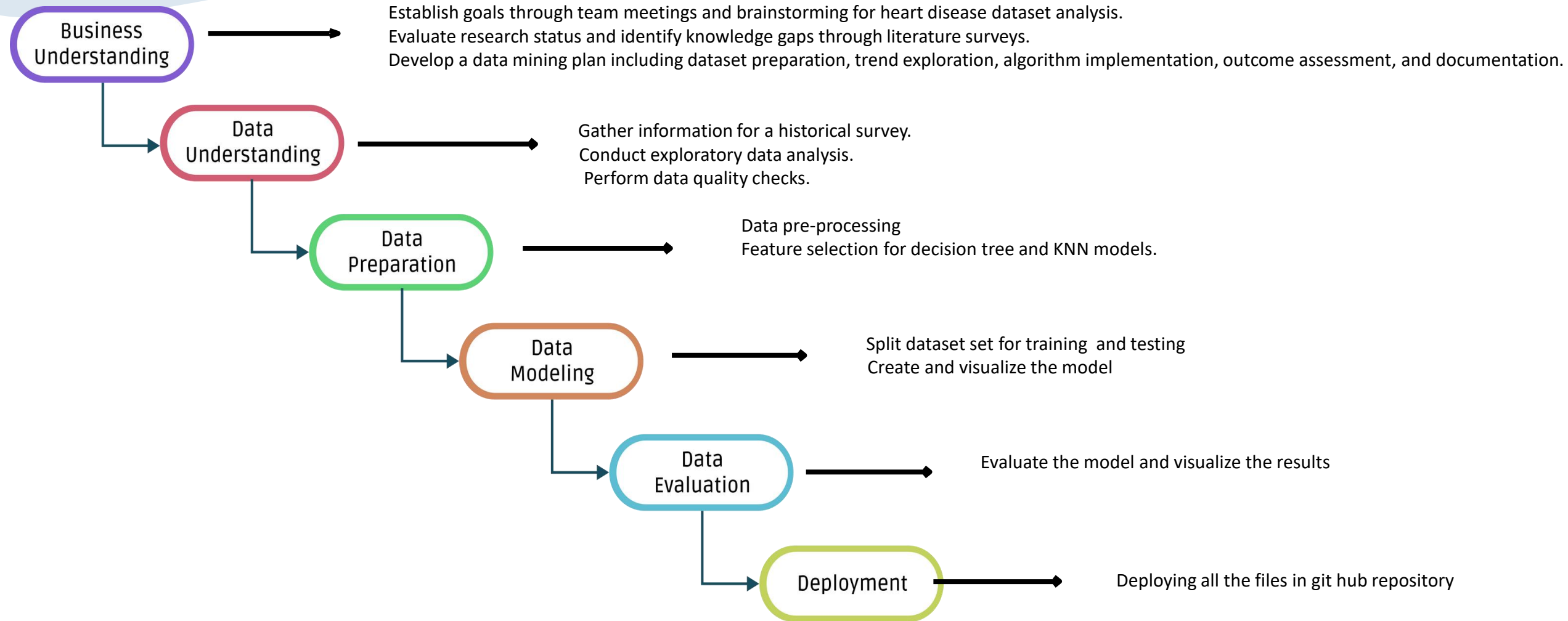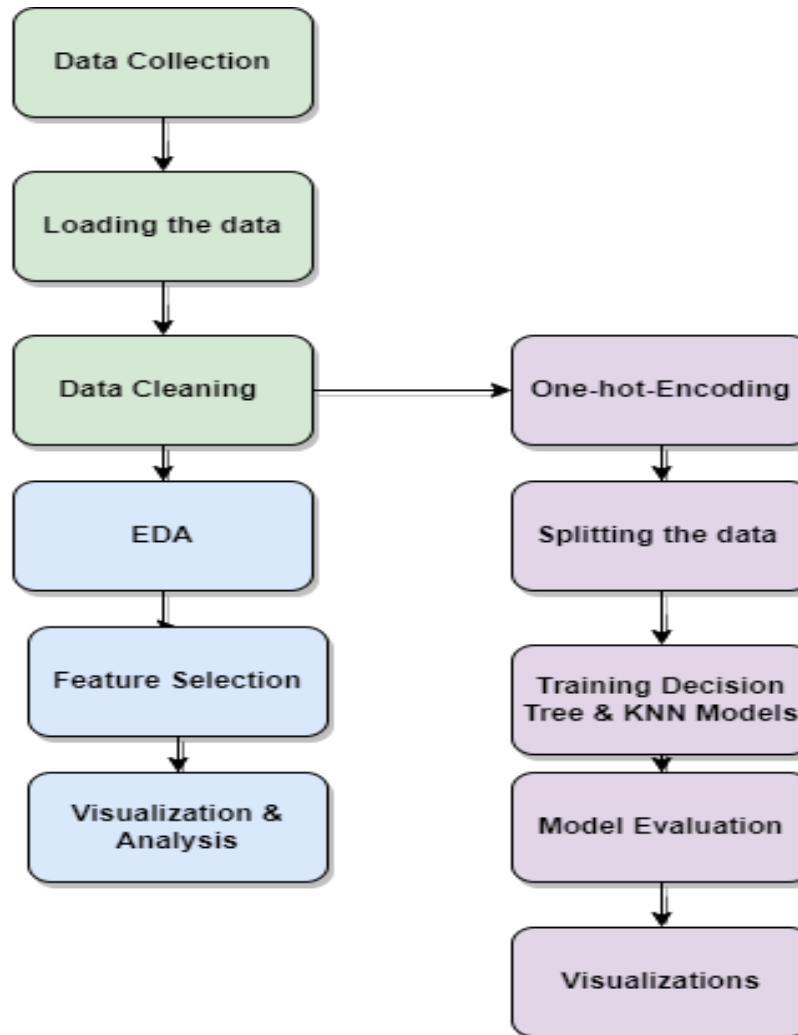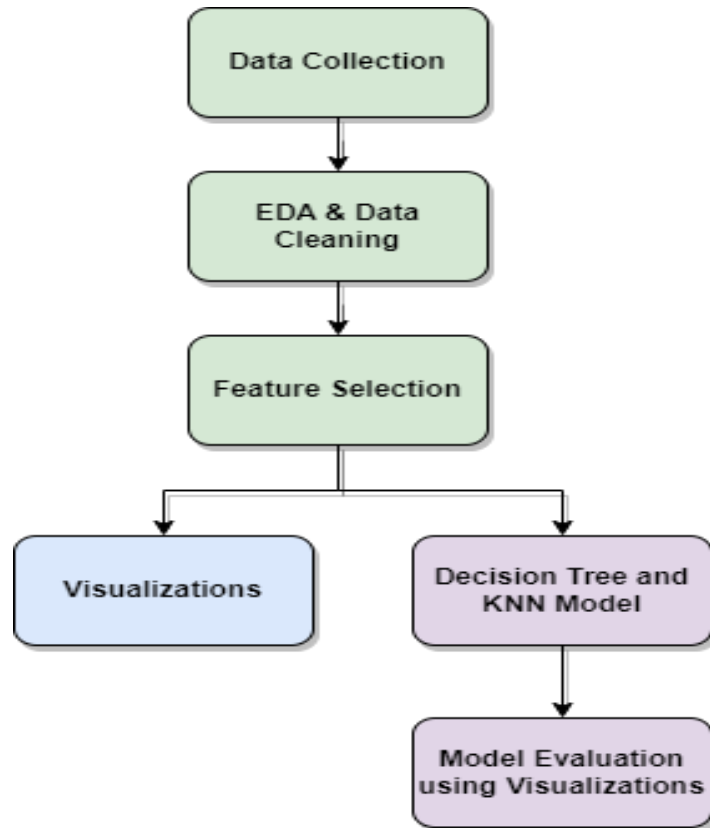
Motivations

The motivation is using visualizations to better understand heart disease and create a predictive algorithm to detect at-risk individuals.

This can enhance heart disease prevention and management, improving patient outcomes and healthcare expenditures.

Hybrid Crisp-DM &Waterfall Model

Business Understanding

Establish goals through team meetings and brainstorming for heart disease dataset analysis.
Evaluate research status and identify knowledge gaps through literature surveys.
Develop a data mining plan including dataset preparation, trend exploration, algorithm implementation, outcome assessment, and documentation.

Data Understanding

Gather information for a historical survey.
Conduct exploratory data analysis.
Perform data quality checks.

Data Preparation

Data pre-processing
Feature selection for decision tree and KNN models.

Data Modeling

Split dataset set for training and testing
Create and visualize the model

Data Evaluation

Evaluate the model and visualize the results

Deployment

Deploying all the files in git hub repository

Project Workflow

Dataset Description

The Heart Disease Prediction dataset was taken from Kaggle. It includes information from a study conducted on patients with suspected heart disease.

The dataset contains 18 columns of patient data, including age, BMI, blood pressure, cholesterol levels, heart disease, diabetic and other pertinent health indicators.
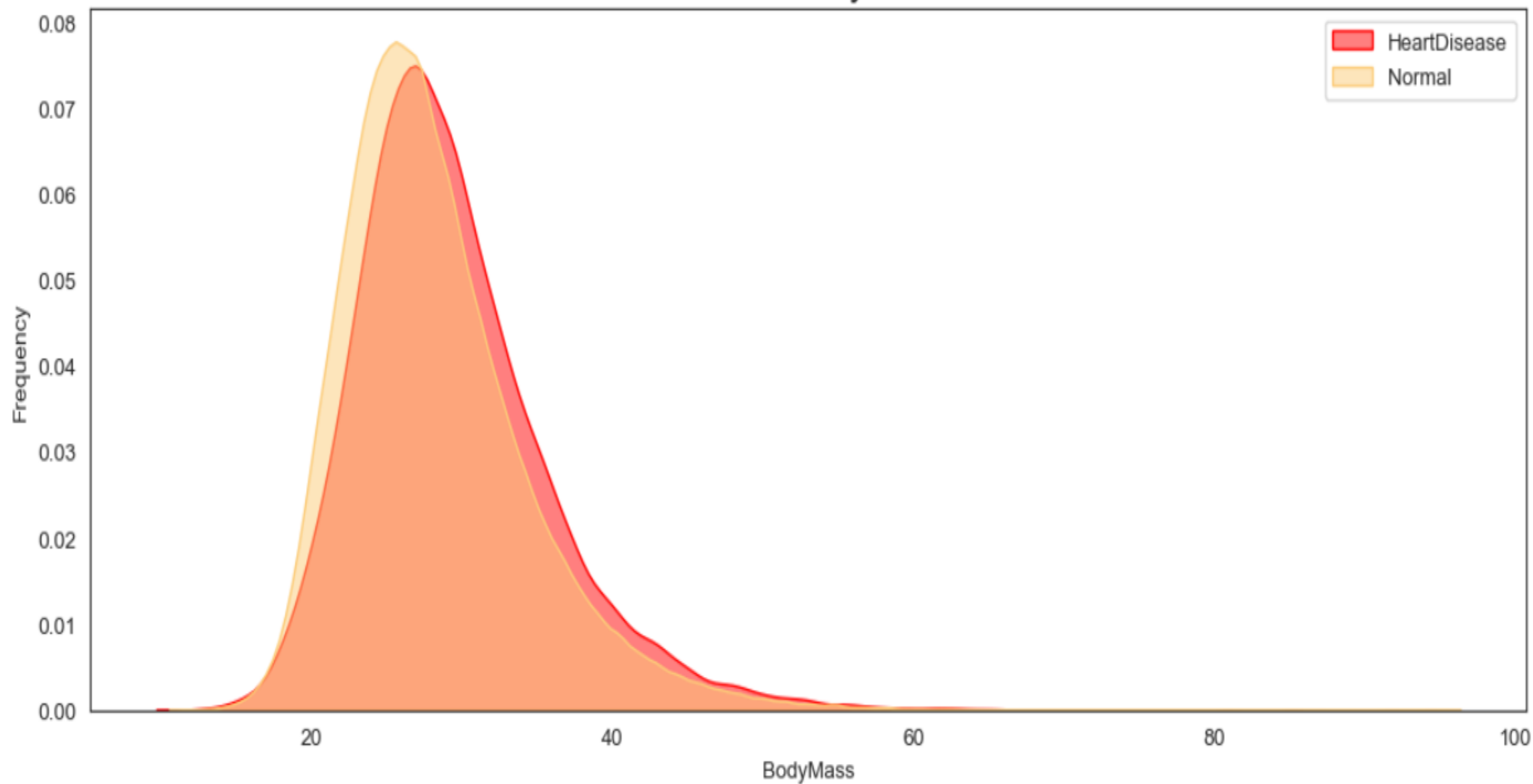
There are 319796 rows in the dataset, each of which represents a different patient. The target variable is heart disease which represents weather a patient having heart disease or not.
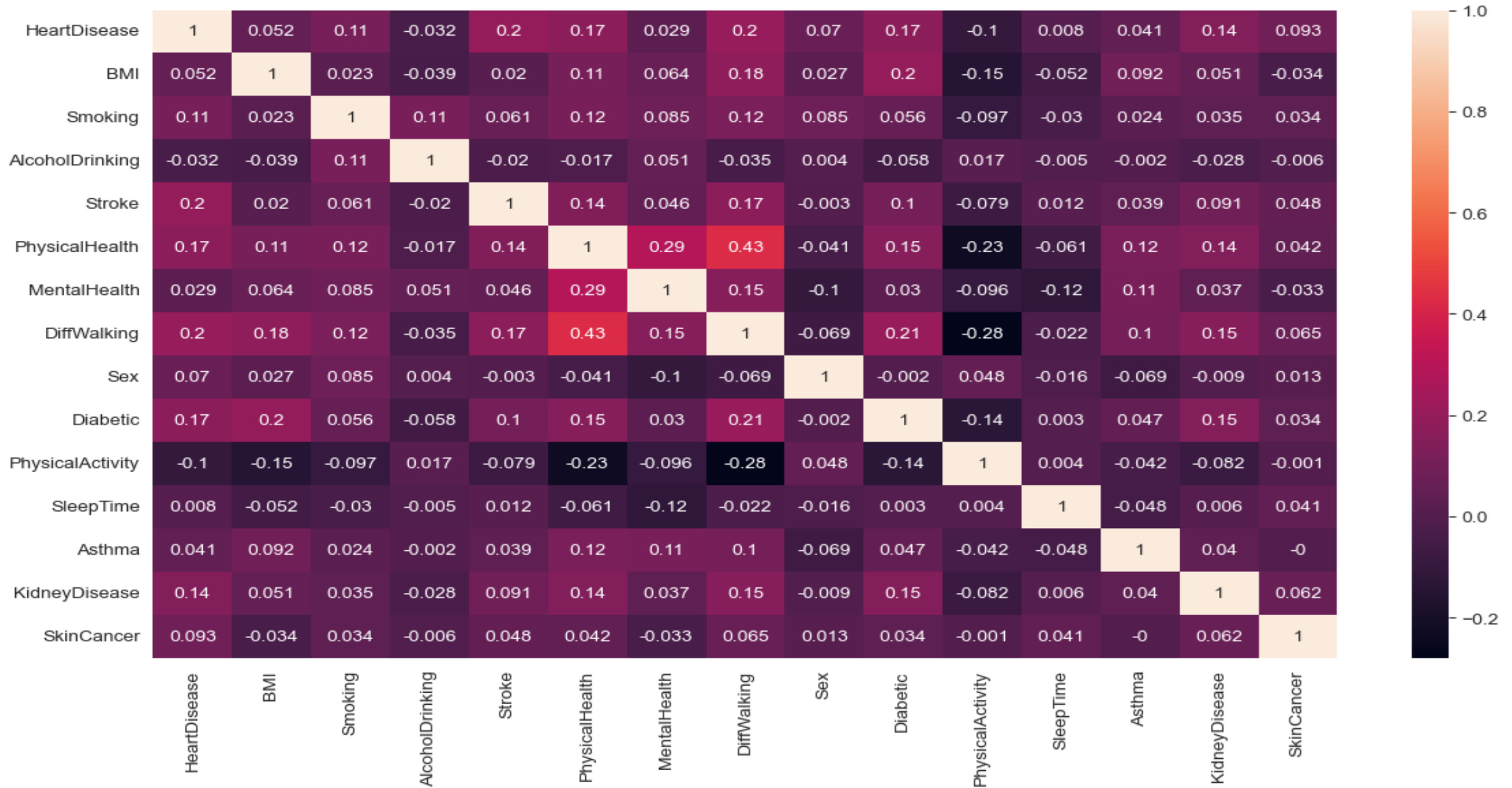
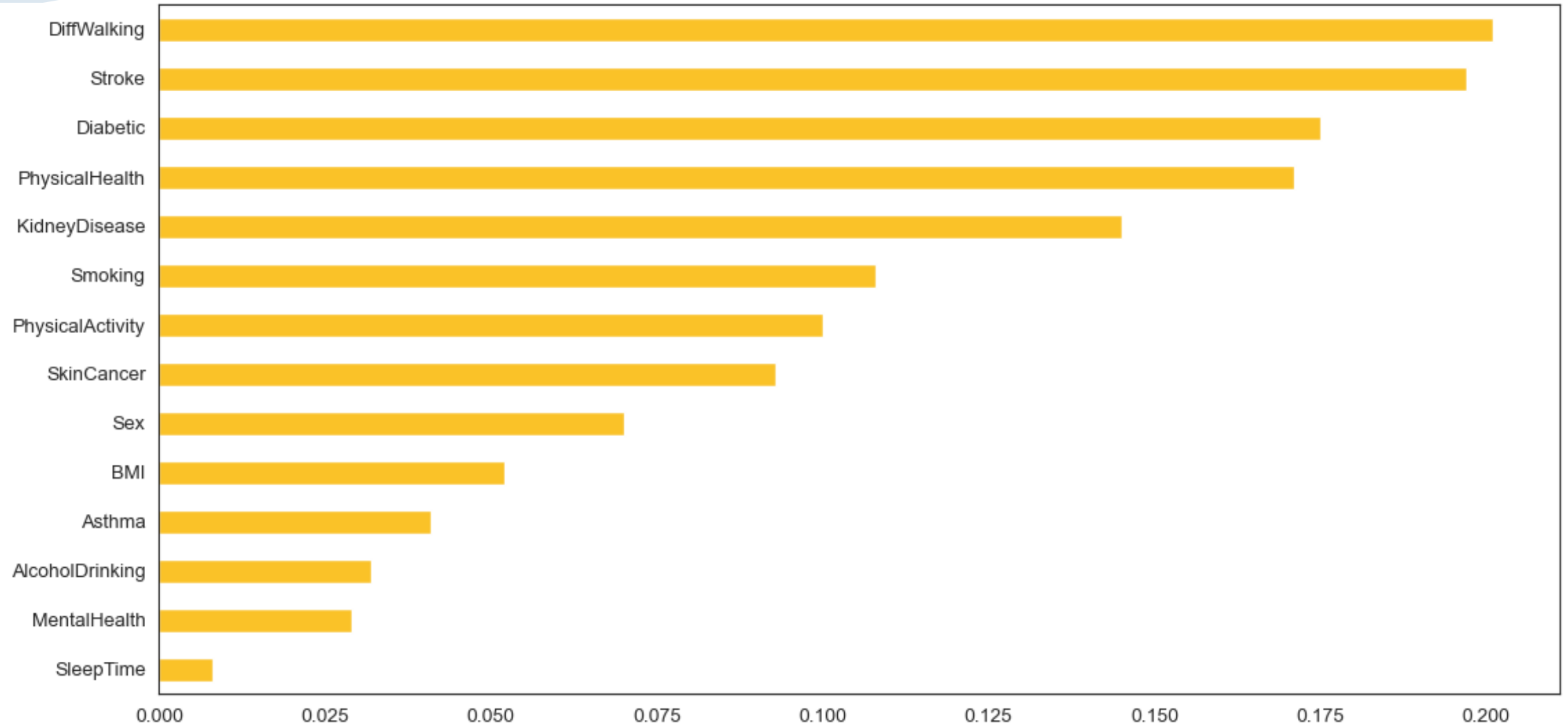# Visualizations

Distribution of Body Mass Index

Distribution of people having heart disease for different ages
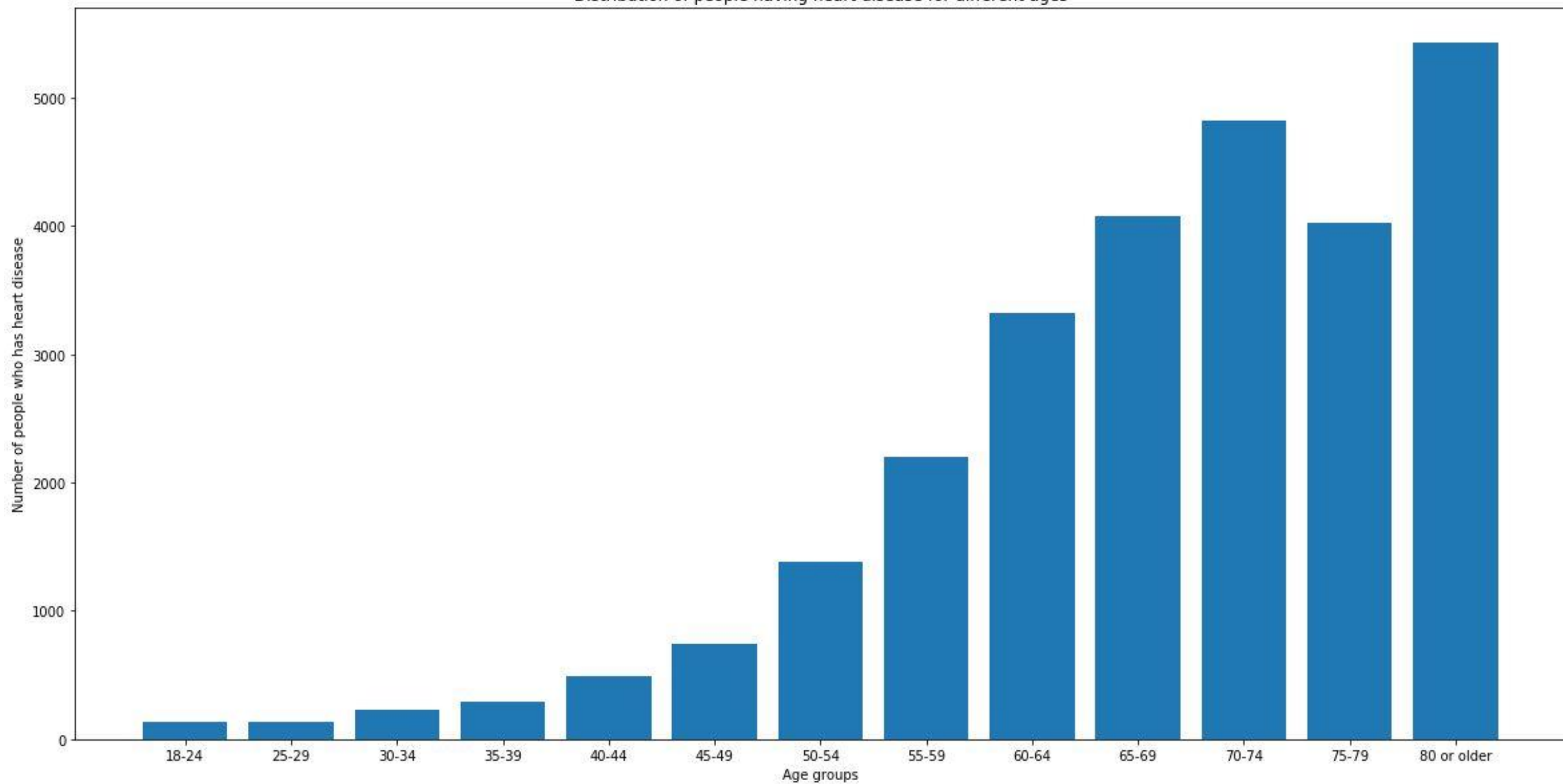
Correlation between all set of features

Distribution of correlation between the features

Algorithms

The data is spilt into train and test sets with a ratio of 70:30.
Train set is used for both models while test set is used for calculating evaluation metrics.

Decision Tree

- Since the data set is having both categorical and numerical values where the categorical columns are containing less unique values like yes, no and female, male, we are training decision tree classification model to classify data into yes or no.

- F1 Score = 0.244

- Accuracy = 0.86

K- nearest Neighbors Model

- Another classification model that we taught would work best for this data is KNN since there is good correlation between different features and heart diseases.

- F1 Score = 0.213

- Accuracy = 0.90

Model Evaluation

Comparing the models
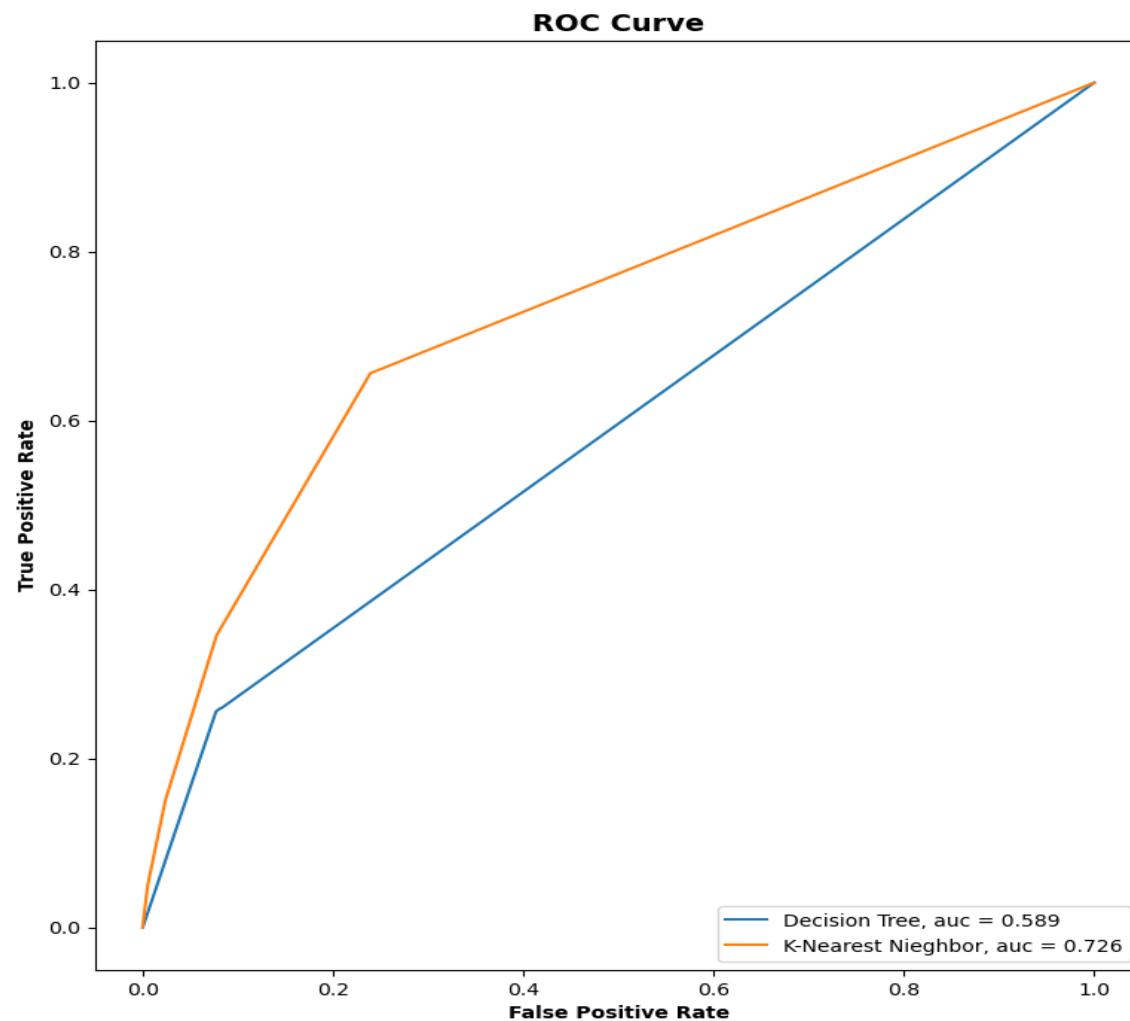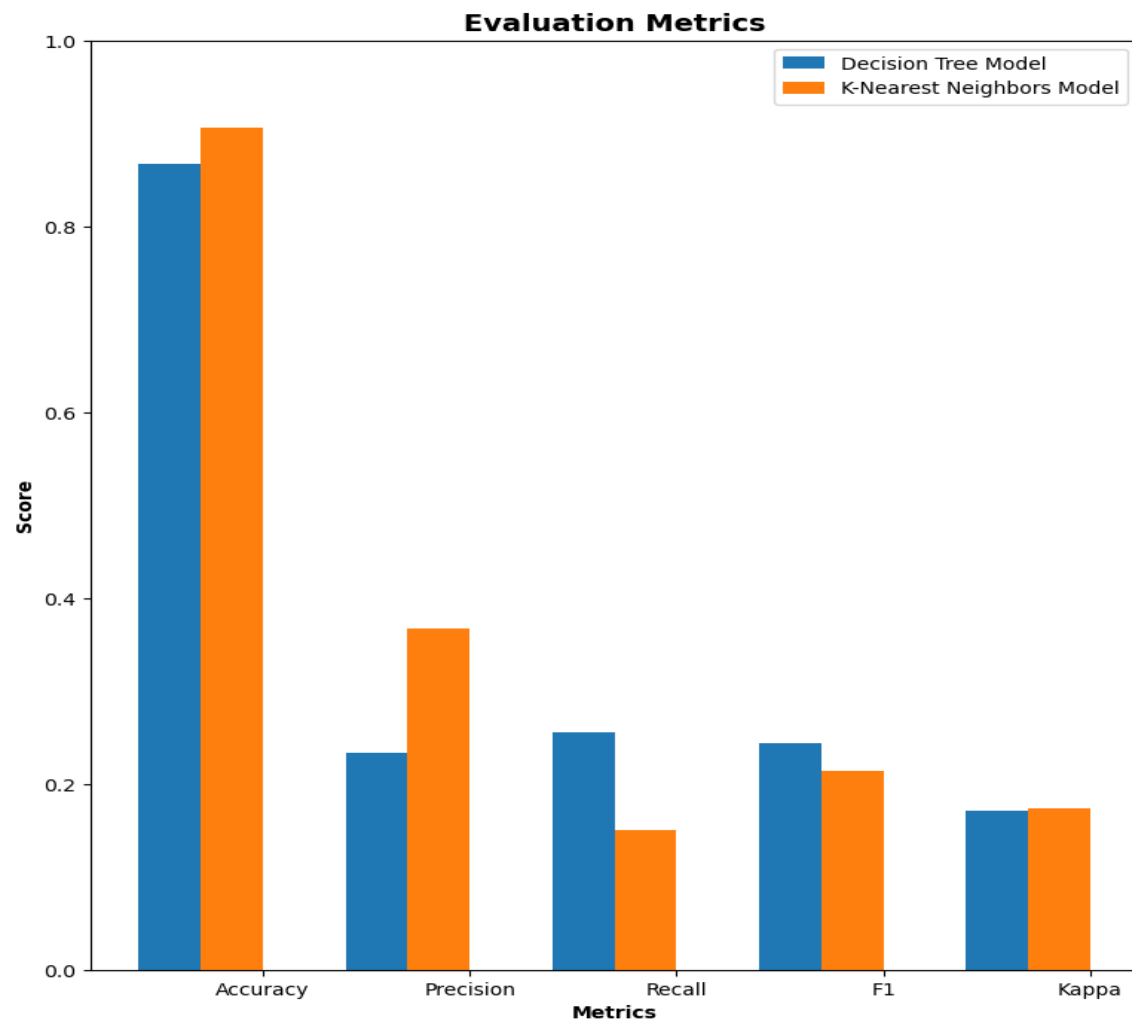
Conclusion

In conclusion, the heart disease study project yielded substantial insights regarding the incidence, causes, and mitigation strategies linked with heart disease.

It is possible to discover essential factors and develop efficient strategies for timely diagnosis and prevention using data analysis and statistical modeling.

It is critical to ensure the quality of data in order to generate reliable results.

In general, this also improves comprehension and results in the treatment of heart disease.

References

- https://github.com/Priyankaakula/DATA230-PROJECT
- https://www.kaggle.com/code/andls555/heart-disease-prediction/notebook
- https://link.springer.com/article/10.1007/s42979-020-00365-y

Thank you