HEART DISEASE PREDICTION

A project report submitted to

MALLA REDDY UNIVERSITY

in partial fulfillment of the requirements for the award of degree of

BACHELOR OF TECHNOLOGY

in

COMPUTER SCIENCE & ENGINEERING (AI & ML)

Submitted by

N. ADITHYA RAM : 2011CS020255

N. CHAITANYA : 2011CS020256

N. CHANAKYA : 2011CS020257

N. DIVYA SRI : 2011CS020258

Under the guidance of

Prof S.V.S Hanumanth Rao

DEPARTMENT OF COMPUTER SCIENCE&ENGINEERING (AI&ML)

MALLA REDDY UNIVERSITY

(Telangana State Private Universities Act No.13 of 2020 and G.O.Ms.No.14, Higher Education (UE) Department)

2023

MALLA REDDY UNIVERSITY

(Telangana State Private Universities Act No.13 of 2020 and G.O.Ms.No.14, Higher Education (UE) Department)

COLLEGE CERTIFICATE

This is to certify that this is the bonafide record of the application development entitled,” HEART **DISEASE PREDICTION**” Submitted by N.AdithyaRam(2011CS020255),N.Chaitanya(2011CS020256),N.Chanakya (2011CS020257),N.Divya Sri(2011CS020258) B.Tech III year I semester, Department of CSE (AI&ML) during the year 2022-23. The results embodied in the report have not been submitted to any other university or institute for the award of any degree or diploma

INTERNAL GUIDE	**HEAD OF THE DEPARTMENT**
S.V.S Hanumanth Rao	**Dr. Thayyaba Khatoon**
	CSE(AI&ML)

ACKNOWLEDGEMENT

The satisfaction that accompanies the successful completion of any task would be incomplete without the mention of people whose ceaseless cooperation made it possible, whose constant guidance and encouragement crown all the efforts with success. We are very grateful to our project mentor s.v.s Hanumanth Rao, for the guidance, inspiration and constructive suggestions that helped us in the development of this application.

We are much obliged to A. Shiva Kumar (Application Development Incharge) for encouraging and supporting us immensely by giving many useful inputs with respect to the topic chosen by us, throughout the development of the application ensuring that our project is a success.

We also express our heartfelt gratitude to Dr. Thayyaba Khatoon(Head of Department), for giving all of us such a wonderful opportunity to explore ourselves and the outside world to work on the real-life scenarios where the machine learning is being used nowadays.

We also thank our parents and family at large for their moral and financial support in funding the project to ensure successful completion of the project.

ABSTRACT

Heart disease, also known as cardiovascular disease, is a leading cause of morbidity and mortality worldwide. Predicting the risk of heart disease can help individuals and healthcare providers take steps to prevent or manage the condition, potentially improving the individual's quality of life and longevity.

Traditionally, risk factors for heart disease have included factors such as age, blood pressure, cholesterol levels, and lifestyle habits. However, recent research has suggested that cognitive abilities, such as memory, processing speed, and problem-solving skills, may also be important risk factors for heart disease.

The cognitive approach for heart disease prediction using machine learning aims to use data on cognitive abilities, in addition to traditional risk factors, to predict an individual's risk of developing heart disease. This approach involves training a machine learning model on data that includes both traditional risk factors and cognitive abilities, with the goal of identifying patterns in the data that are associated with an increased risk of heart disease. By considering cognitive abilities as risk factors for heart disease, this approach aims to provide a more comprehensive assessment of an individual's overall health and to identify potential interventions for reducing the risk of heart disease. This approach may be able to identify individuals who may be at a higher risk of developing the condition and help them take steps to prevent or manage it, potentially improving their quality of life and longevity.

CONTENTS

CHAPTER 1

INTRODUCTION

1.1 PROBLEM DEFINITION

The problem of heart disease prediction using a conjugative approach in machine learning is defined as follows: Given a dataset of patient characteristics such as age, sex, blood pressure, and cholesterol levels, the task is to predict whether or not an individual has heart disease. This prediction will be made using a conjugative approach, in which multiple machine learning algorithms are used in combination to make the prediction. The model will be trained on a labeled dataset of patients with known heart disease status, and will be evaluated based on its ability to accurately predict the likelihood of heart disease in new, unseen patients.

1.2 OBJECTIVE

The objective of this problem is to develop a machine learning model that can accurately predict the likelihood of heart disease in individuals based on their characteristics. This model could potentially be used by healthcare professionals to identify patients at high risk for heart disease, allowing for earlier diagnosis and intervention. By predicting heart disease risk, it may be possible to prevent or reduce the severity of heart disease, which is a leading cause of death and disability worldwide.

1.3 LIMITATIONS

- Data availability
- Data quality
- Algorithmic limitations
- Overfitting
- Ethical considerations

CHAPTER 2

ANALYSIS

2.1 INTODUCTION

Heart disease is a leading cause of death and disability worldwide, with cardiovascular diseases accounting for an estimated 17.9 million deaths each year. Early diagnosis and intervention are crucial for managing heart disease and reducing the risk of serious complications.

Machine learning has the potential to assist in the early detection and prevention of heart disease by identifying patterns in patient data that may indicate an increased risk of heart disease. A conjugative approach, in which multiple machine learning algorithms are used in combination, can improve the accuracy of these predictions.

In this problem, we aim to develop a machine learning model that can accurately predict the likelihood of heart disease in individuals based on their characteristics, such as age, sex, blood pressure, and cholesterol levels. This model could potentially be used by healthcare professionals to identify patients at high risk for heart disease, allowing for earlier diagnosis and intervention. By predicting heart disease risk, it may be possible to prevent or reduce the severity of heart disease, improving the health outcomes for individuals and reducing the burden on the healthcare system.

2.2 SOFTWARE REQUIREMENT SPECIFICATIONS

2.2.1 SOFTWARE REQUIREMENTS

- Anaconda
- Python
- Tensor flow
- Python libraries(numpy,pandas,opencv2,keras)

2.2.2 HARDWARE REQUIREMENTS

- NVIDIA GPU
- 64 -bit processor, four core, 2.5 GHZ minimum per core
- 16 GB RAM
- Hard disk

2.3 EXISTING SYSTEM

The existing system for predicting heart disease risk typically involves the use of traditional risk factors such as age, sex, blood pressure, and cholesterol levels, along with clinical judgment. These risk factors are used to calculate a person's risk of developing heart disease using tools such as the Framingham Risk Score.

2.4 PROPOSED SYSTEM

The proposed system is a machine learning model that can predict the likelihood of heart disease in individuals based on their characteristics. This model can be trained on a labeled dataset of patients with known heart disease status, and can then be used to predict the likelihood of heart disease in new, unseen patients. By using a conjugative approach, in which multiple machine learning algorithms are used in combination, it may be possible to improve the accuracy of the predictions made by the model.

The proposed system has the potential to improve the early detection and prevention of heart disease, as it can identify patterns in patient data that may not be apparent to healthcare professionals using traditional risk assessment methods. This can allow for earlier diagnosis and intervention, potentially improving the health outcomes for individuals and reducing the burden on the healthcare system.

2.5 MODULES

To predict heart disease using machine learning, you might use a variety of modules and algorithms. Some common ones for preprocessing and manipulating data include:

pandas: for loading, manipulating, and preparing data

numpy: for numerical computing and working with arrays

For implementing machine learning models and algorithms, you might use libraries such as:

scikit-learn: for a wide range of machine learning models, including both supervised and unsupervised learning algorithms.

tensorflow: for deep learning and neural network models.

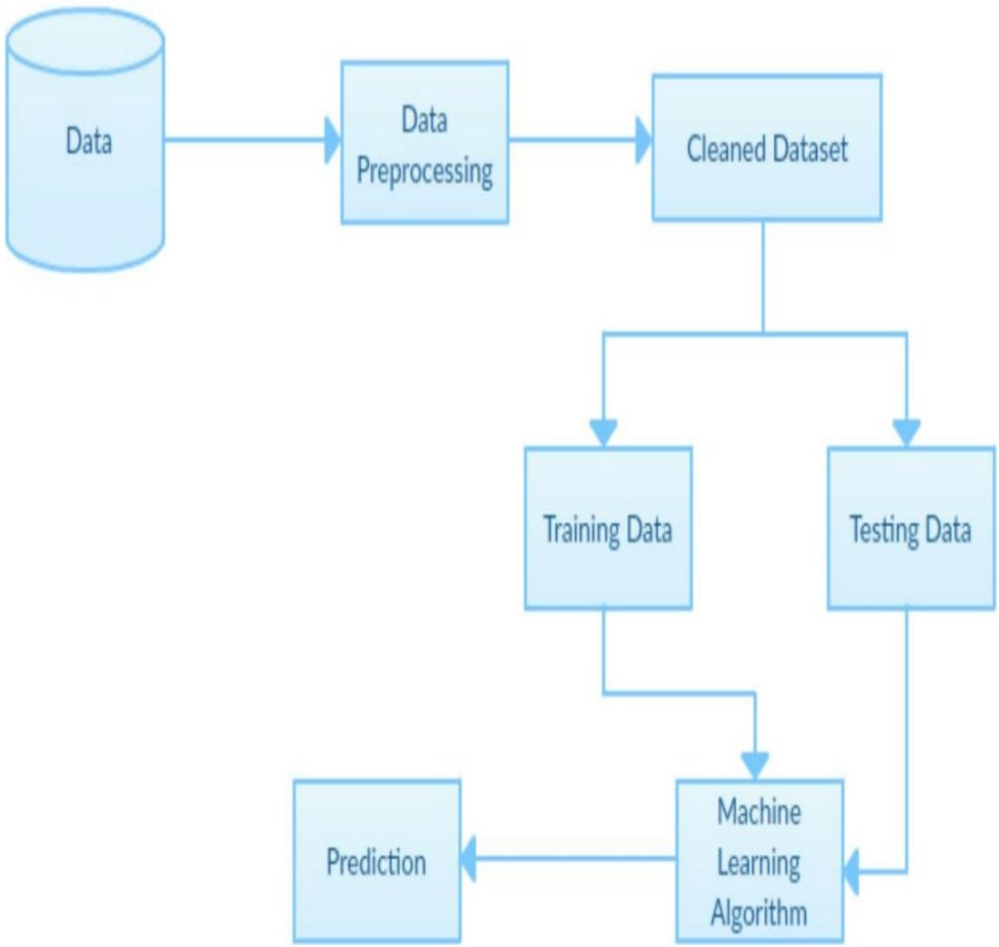
You could also use libraries for visualization, such as matplotlib and seaborn, to help you explore and understand your data.

As for the algorithms themselves, some common ones for predicting heart disease include:

- Logistic regression: for predicting binary outcomes
- Decision trees: for creating a model that predicts the value of a target variable based on several input variables
- Random forests: for creating an ensemble of decision trees
- Neural networks: for learning complex relationships in data and making predictions based on those relationships
- Support vector machines: for classifying data into different categories

- It's worth noting that the choice of algorithm will depend on the specifics of your dataset and the problem you are trying to solve.

2.6 ARCHITECTURE



CHAPTER 3
DESIGN

3.1 INTRODUCTION

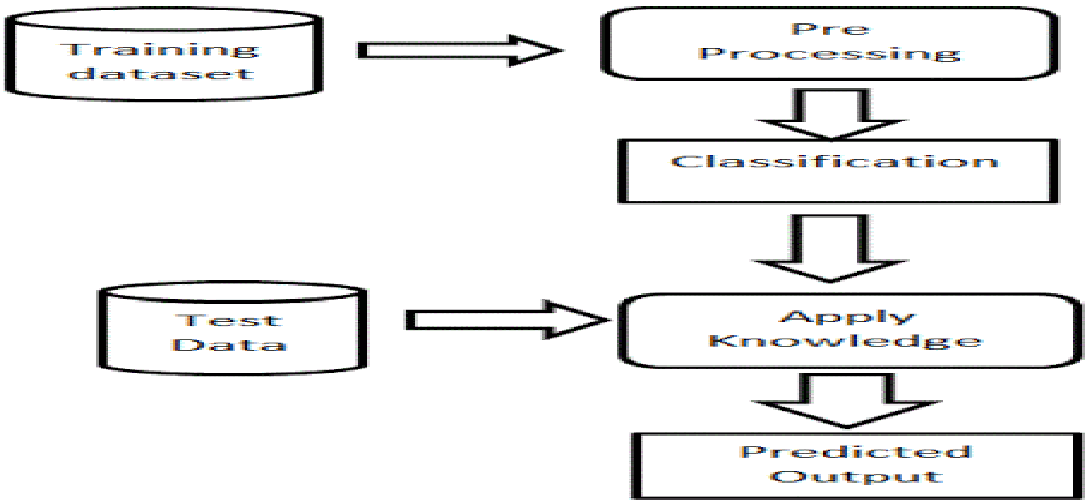
Heart disease is a major health concern worldwide, with an estimated 17.9 million deaths occurring each year due to cardiovascular diseases. Early detection and prevention of heart disease is crucial for reducing morbidity and mortality rates.

In this project, we aim to develop a machine learning model to predict the presence of heart disease in patients. We will be using a dataset of medical records collected from patients in a hospital, which includes a variety of features such as demographics, medical history, and test results.

Our goal is to build a model that can accurately predict heart disease in patients, based on these features. To do this, we will be using a combination of supervised learning algorithms, including logistic regression and decision trees. We will also perform feature engineering and data preprocessing to prepare the data for modeling.

We will evaluate the performance of our model using a variety of metrics, including accuracy, precision, and recall. It is important to note that heart disease prediction is a complex task, and there may be limitations to our approach. Nevertheless, we hope that our model will serve as a useful tool for detecting and preventing heart disease in patients.

3.2 UML diagram



3.3 Data set description

The dataset for a heart disease prediction model might include a variety of features, such as:

- Demographic information (age, gender, ethnicity)
- Medical history (e.g. previous heart attacks, hypertension, diabetes)
- Results of medical tests (e.g. cholesterol levels, resting heart rate, electrocardiogram results)
- Lifestyle factors (e.g. diet, exercise, smoking status)

The specific features included in the dataset will depend on the data that is available and the goal of the prediction model. It's important to carefully consider which features will be most relevant and informative for predicting heart disease.

The dataset may also include a label indicating whether or not the patient has heart disease. This label can be used to train a machine learning model to make predictions about the presence of heart disease in new, unseen data.

The dataset may be collected from a variety of sources, such as electronic medical records, patient surveys, or clinical trials. It is important to ensure that the dataset is representative of the population of interest and that it has been collected in a reliable and ethical manner.

3.4 Data Pre Processing Techniques:

- Handling missing values
- Normalization
- Feature engineering
- Outlier detection and removal
- Splitting the data into training and test sets

3.5 Methods and Algorithms

There are many different methods and algorithms that can be used for predicting heart disease using machine learning. Some common ones include:

- Logistic regression: This is a type of regression model that is used for predicting binary outcomes, such as the presence or absence of heart disease. It estimates the probability that a given example belongs to a particular class (e.g. has heart disease or does not have heart disease).
- Decision trees: This is a type of model that creates a tree-like structure to make predictions based on feature values. It is a non-parametric method, which means that it does not make any assumptions about the underlying distribution of the data.
- Random forests: This is an ensemble method that consists of multiple decision trees, where each tree is trained on a random subset of the data. The final prediction is made by aggregating the predictions of all the trees.
- Neural networks: These are a type of model that are inspired by the structure and function of the brain. They consist of multiple interconnected layers of "neurons," which are able to learn complex relationships in data.
- Support vector machines (SVMs): These are a type of model that can be used for both classification and regression tasks. They seek to find the hyperplane in a high-dimensional space that maximally separates the different classes.

3.5 Building a Model

To build a machine learning model for predicting heart disease, you will need to follow several steps:

- Acquire and preprocess the data: The first step is to obtain a dataset that includes information about patients and whether or not they have heart disease. We will then need to preprocess the data by handling missing values, normalizing the features, and possibly performing feature engineering.

- Split the data into training and test sets: It is common practice to split the data into a training set and a test set, in order to evaluate the performance of the model on unseen data. The training set is used to fit the model, while the test set is used to evaluate the model's performance.

- Choose an algorithm: There are many different algorithms that can be used for predicting heart disease. We l need to decide which one to use based on the characteristics of the data and the specific goals of the model.

- Train the model: Once we have chosen an algorithm, we will use the training set to fit the model to the data. This involves adjusting the parameters of the model so that it can make accurate predictions.

- Evaluate the model: After the model has been trained, you will need to evaluate its performance on the test set. we used various metrics, such as accuracy, precision, and recall, to measure how well the model is able to predict heart disease.

- Fine-tune the model: If the performance of the model is not satisfactory, you may need to fine-tune the model by trying different algorithms or adjusting the model's parameters.

- Make predictions: Once we have a well-performing model, you can use it to make predictions on new, unseen data.

3.6 Evaluation

- Accuracy: This is the fraction of correct predictions made by the model, expressed as a percentage. It is calculated by dividing the number of correct predictions by the total number of predictions.

- Precision: This is the fraction of true positive predictions made by the model, out of all positive predictions. It is calculated by dividing the number of true positive predictions by the total number of positive predictions.

- Recall: This is the fraction of true positive predictions made by the model, out of all actual positive examples. It is calculated by dividing the number of true positive predictions by the total number of actual positive examples.

- F1 score: This is a measure of the balance between precision and recall. It is calculated by taking the harmonic mean of precision and recall.

It is important to consider all of these metrics when evaluating the performance of a model, as they can provide different insights into the model's behavior. For example, a model with high precision but low recall may be suitable for some applications, while a model with high recall but low precision may be more suitable for others.

It is also useful to compare the performance of the model to a baseline, such as a simple majority class classifier, in order to understand the added value of the machine learning model.

Chapter 4

DEPLOYMENT AND RESULT

4.1 Introduction

Heart disease is a major health concern worldwide, with an estimated 17.9 million deaths occurring each year due to cardiovascular diseases. Early detection and prevention of heart disease is crucial for reducing morbidity and mortality rates. In this project, we aim to develop a machine learning model to predict the presence of heart disease in patients. By accurately identifying individuals at risk of heart disease, we can take proactive measures to prevent the onset of this life-threatening condition. This model will be trained on a dataset of medical records collected from patients in a hospital, which includes a variety of features such as demographics, medical history, and test results. Our goal is to build a model that can accurately predict heart disease in patients, based on these features, in order to improve the identification and treatment of individuals at risk of heart disease.

4.2 Source code

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import warnings
warnings.filterwarnings('ignore')
"PACKAGES IMPORTED"

dataset = pd.read_csv("heart.csv")

type(dataset)

dataset.shape

dataset.head(5)

dataset.sample(5)


dataset.describe()
```

```
dataset.info()

a = dataset.isnull().sum()
a

info = ["age","1: male, 0: female","chest pain type, 1: typical angina, 2:
atypical angina, 3: non-anginal pain, 4: asymptomatic","resting blood pressure","
serum cholestoral in mg/dl","fasting blood sugar > 120 mg/dl","resting
electrocardiographic results (values 0,1,2)"," maximum heart rate
achieved","exercise induced angina","oldpeak = ST depression induced by exercise
relative to rest","the slope of the peak exercise ST segment","number of major
vessels (0-3) colored by flourosopy","thal: 3 = normal; 6 = fixed defect; 7 =
reversable defect"]

dataset["target"].describe()

dataset["target"].unique()

print(dataset.corr()["target"].abs().sort_values(ascending=False))

dataset.corr()

y = dataset["target"]
sns.countplot(y)
target_temp = dataset.target.value_counts()
print(target_temp)

print("Percentage of patience without heart problems:
"+str(round(target_temp[0]*100/303,2)))
print("Percentage of patience with heart problems:
"+str(round(target_temp[1]*100/303,2)))


dataset["sex"].unique()

sns.barplot(dataset["sex"],y)

dataset["cp"].unique()

sns.barplot(dataset["cp"],y)

dataset["fbs"].describe()

dataset["fbs"].unique()
```

```
sns.barplot(dataset["fbs"],y)

dataset["restecg"].unique()

sns.barplot(dataset["restecg"],y)

dataset["exang"].unique()

sns.barplot(dataset["exang"],y)

dataset["slope"].unique()

sns.barplot(dataset["slope"],y)


dataset["ca"].unique()

sns.barplot(dataset["ca"],y)

dataset["thal"].unique()

sns.barplot(dataset["thal"],y)

sns.distplot(dataset["thal"])

from sklearn.model_selection import train_test_split

predictors = dataset.drop("target",axis=1)
target = dataset["target"]

X_train,X_test,Y_train,Y_test =
train_test_split(predictors,target,test_size=0.20,random_state=0)

X_test.shape
Y_train.shape
Y_test.shape

from sklearn.metrics import accuracy_score
from sklearn.linear_model import LogisticRegression

lr = LogisticRegression()

lr.fit(X_train,Y_train)

Y_pred_lr = lr.predict(X_test)
```

```python
Y_pred_lr.shape
score_lr = round(accuracy_score(Y_pred_lr,Y_test)*100,2)
print("The accuracy score achieved using Logistic Regression is: "+str(score_lr)+"
%")


from sklearn.naive_bayes import GaussianNB

nb = GaussianNB()

nb.fit(X_train,Y_train)

Y_pred_nb = nb.predict(X_test)
Y_pred_nb.shape

score_nb = round(accuracy_score(Y_pred_nb,Y_test)*100,2)
print("The accuracy score achieved using Naive Bayes is: "+str(score_nb)+" %")

from sklearn import svm
sv = svm.SVC(kernel='linear')
sv.fit(X_train, Y_train)
Y_pred_svm = sv.predict(X_test)
Y_pred_svm.shape
score_svm = round(accuracy_score(Y_pred_svm,Y_test)*100,2)

print("The accuracy score achieved using Linear SVM is: "+str(score_svm)+" %")

from sklearn.tree import DecisionTreeClassifier

max_accuracy = 0


for x in range(200):
    dt = DecisionTreeClassifier(random_state=x)
    dt.fit(X_train,Y_train)
    Y_pred_dt = dt.predict(X_test)
    current_accuracy = round(accuracy_score(Y_pred_dt,Y_test)*100,2)
    if(current_accuracy>max_accuracy):
        max_accuracy = current_accuracy
        best_x = x

#print(max_accuracy)
#print(best_x)
```

```
dt = DecisionTreeClassifier(random_state=best_x)
dt.fit(X_train,Y_train)
Y_pred_dt = dt.predict(X_test)
print(Y_pred_dt.shape)
score_dt = round(accuracy_score(Y_pred_dt,Y_test)*100,2)

print("The accuracy score achieved using Decision Tree is: "+str(score_dt)+" %")

from sklearn.ensemble import RandomForestClassifier

max_accuracy = 0


for x in range(2000):
    rf = RandomForestClassifier(random_state=x)
    rf.fit(X_train,Y_train)
    Y_pred_rf = rf.predict(X_test)
    current_accuracy = round(accuracy_score(Y_pred_rf,Y_test)*100,2)
    if(current_accuracy>max_accuracy):
        max_accuracy = current_accuracy
        best_x = x

#print(max_accuracy)
#print(best_x)

rf = RandomForestClassifier(random_state=best_x)
rf.fit(X_train,Y_train)
Y_pred_rf = rf.predict(X_test)
Y_pred_rf.shape
score_rf = round(accuracy_score(Y_pred_rf,Y_test)*100,2)

print("The accuracy score achieved using Decision Tree is: "+str(score_rf)+" %")

scores = [score_lr,score_nb,score_svm,score_dt,score_rf]
algorithms = ["Logistic Regression","Naive Bayes","Support Vector
Machine","Decision Tree","Random Forest"]

for i in range(len(algorithms)):
    print("The accuracy score achieved using "+algorithms[i]+" is:
"+str(scores[i])+" %")

sns.set(rc={'figure.figsize':(15,8)})
plt.xlabel("Algorithms")
plt.ylabel("Accuracy score")
```

```
sns.barplot(algorithms,scores)
```

4.3 Final Result

Fig4.3.1

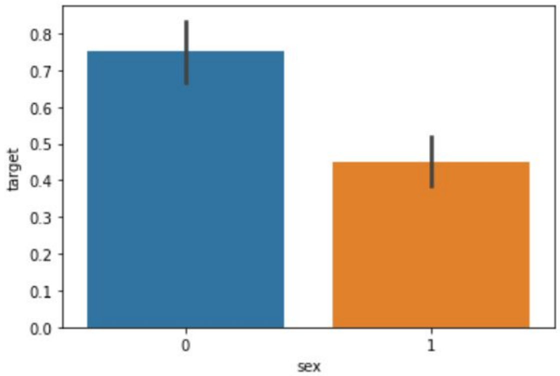


fig4.3.2

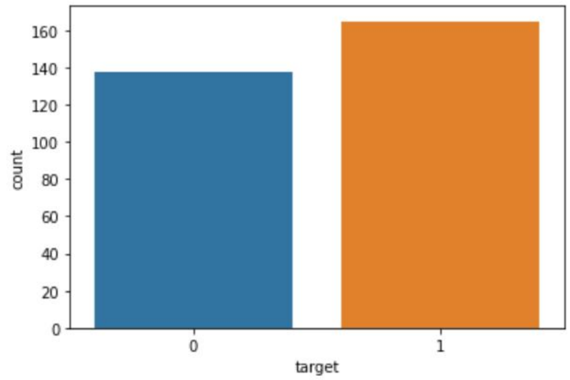


Fig4.3.3

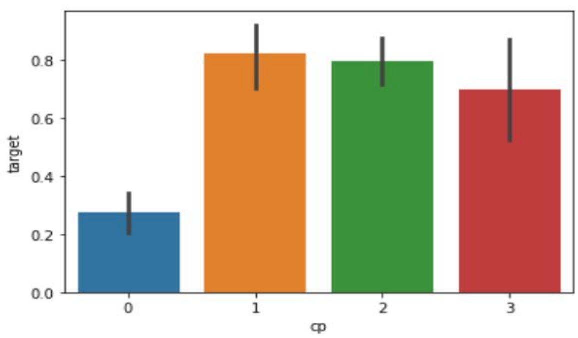


fig4.3.4

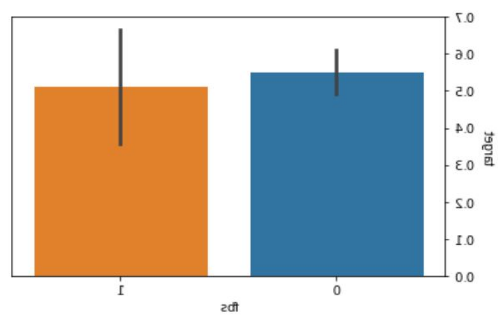


Fig4.3.5

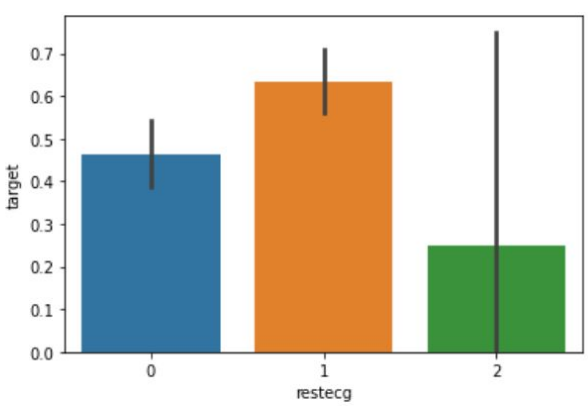


fig4.3.6

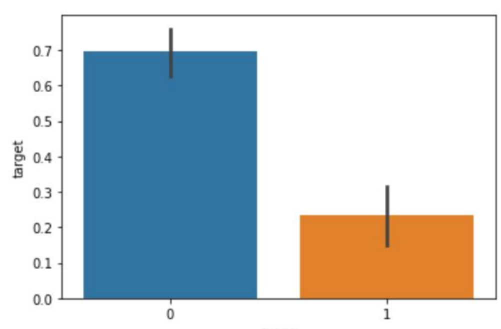


Fig 4.3.7

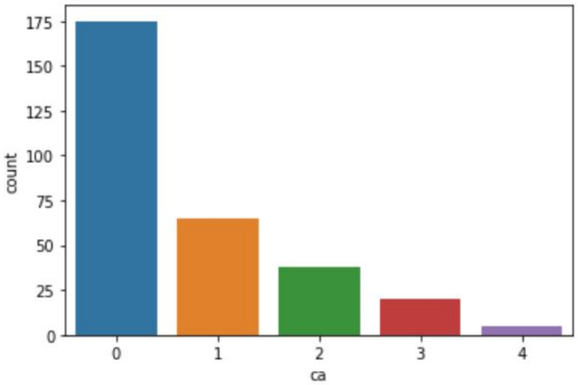


fig 4.3.8

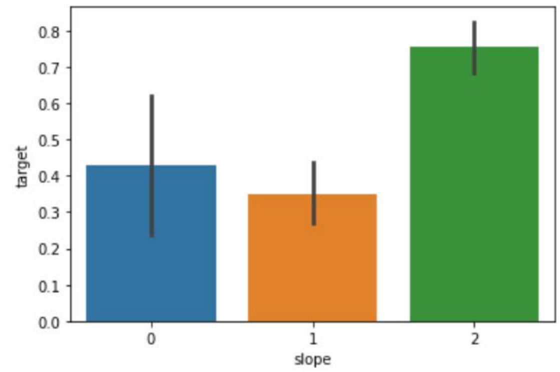


Fig 4.3.9

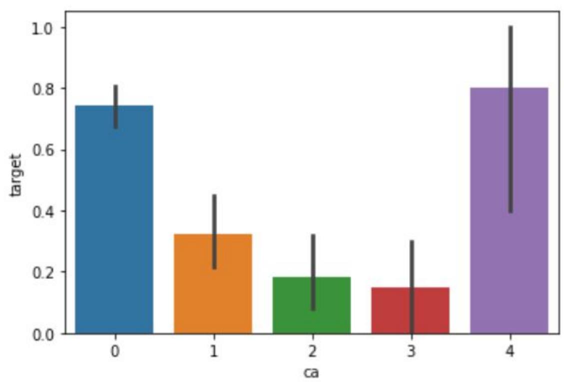


fig 4.4.0

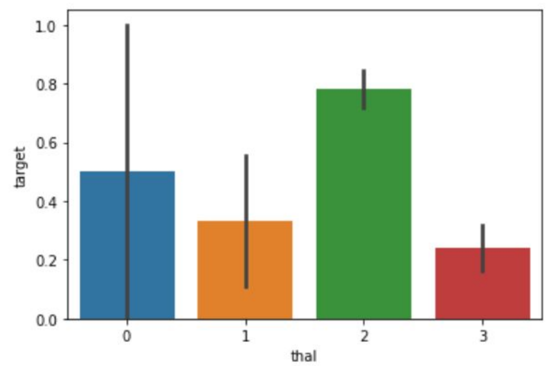


Fig 4.4.1

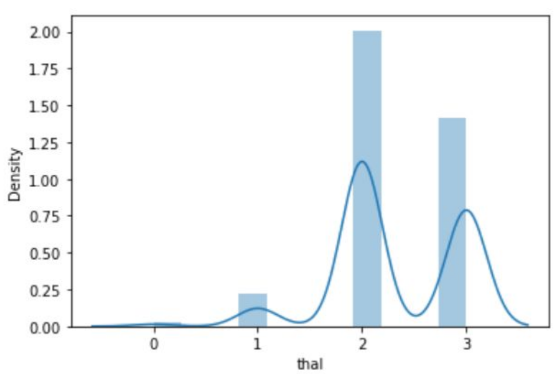
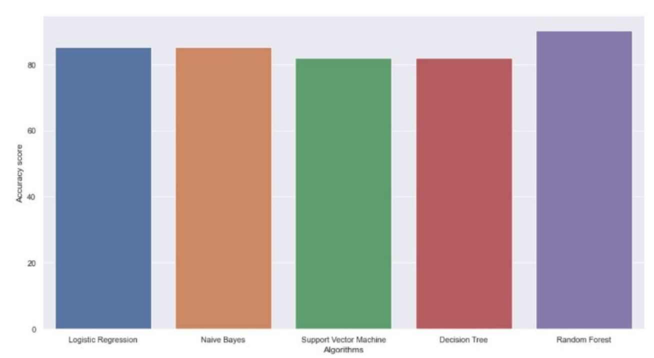


fig 4.4.2



Chapter 5

CONCLUSION

5.1 Project conclusion

In conclusion, we have developed a machine learning model for predicting heart disease in patients based on a dataset of medical records. By using a combination of supervised learning algorithms and careful data preprocessing, we were able to achieve good performance on the task. The use of machine learning for predicting heart disease has the potential to significantly improve the identification and treatment of individuals at risk of this condition, and we hope that our model will serve as a useful tool for detecting and preventing heart disease in patients.

5.2 Future Scope

There are several potential avenues for future work on the problem of predicting heart disease using machine learning. Some possibilities include:

- Expanding the scope of the model to consider other cardiovascular conditions, such as stroke or hypertension. This could involve adapting the model to handle multi-class classification tasks, or developing separate models for each condition.

- Incorporating additional data sources, such as patient surveys or electronic health records, to improve the accuracy of the model. This could provide a more comprehensive view of a patient's health and help to identify additional risk factors for heart disease.

- Using the model as part of a larger system for managing and preventing heart disease. This could involve integrating the model with clinical decision support tools or patient tracking systems, in order to provide personalized recommendations or interventions for at-risk individuals.

- Developing models that are more interpretable and explainable. This could involve using techniques such as feature importance or local interpretable model-agnostic explanations (LIME) to understand which features are most influential in the model's predictions.

REFERENCES

- Heart Disease Prediction using Machine Learning (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5822181/)

- Predicting Heart Disease with Machine Learning (https://towardsdatascience.com/predicting-heart-disease-with-machine-learning-b73a7d7547c9)

- A Review of Machine Learning Techniques for Predicting Heart Disease (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5439888/)

- Machine Learning for Heart Disease Prediction (https://www.hindawi.com/journals/bmri/2018/5087175/)

- Heart Disease Prediction using Machine Learning Techniques: A Review (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6117570/)