**Project Report**

**On**

**Heart Disease Prediction**

**Under**

**Vocational training**

**AI with Python**

|  |  |
| --- | --- |
|  |  |
| **Submitted by –**  **Dhruvika Sharma**  **Aniket Kalta** | **Project Mentor Name** |
|  |  |

**Table of Contents**

|  |  |
| --- | --- |
|  | **Page No** |
| 1. **Introduction** | **3** |
| 1. **Problem statement** | **4** |
| 1. **Objectives** | **5** |
| 1. **Design Methodology (Block Diagram or Work flow)** | **6** |
| **5 Dataset Explanation** | **10** |
| **6 Results – Front end clips, back end design clips, data base etc** | **12** |
| **7 Conclusion** | **13** |
| **8 References** | **14** |

**INTRODUCTION**

According to the World Health Organization, every year 12 million deaths occur worldwide due to Heart Disease. The load of cardiovascular disease is rapidly increasing all over the world from the past few years. Many researches have been conducted in attempt to pinpoint the most influential factors of heart disease as well as accurately predict the overall risk. Heart Disease is even highlighted as a silent killer which leads to the death of the person without obvious symptoms. The early diagnosis of heart disease plays a vital role in making decisions on lifestyle changes in high-risk patients and in turn reduce the complications. This project aims to predict future Heart Disease by analyzing data of patients which classifies whether they have heart disease or not using machine-learning algorithms.

This is important to medical fields. If such a prediction is accurate enough, we can not only avoid wrong diagnosis but also save human resources. With accurate predictions, we can solve the unnecessary trouble. Besides, if we can apply our machine learning tool into medical prediction, we will save human resource because we do not need complicated diagnosis process in hospitals.

**PROBLEM STATEMENT**

Heart disease can be managed effectively with a combination of lifestyle changes, medicine and, in some cases, surgery. With the right treatment, the symptoms of heart disease can be reduced and the functioning of the heart improved. The predicted results can be used to prevent and thus reduce cost for surgical treatment and other expensive. The overall objective of my work will be to predict accurately with few tests and attributes the presence of heart disease. Attributes considered form the primary basis for tests and give accurate results more or less. Many more input attributes can be taken but our goal is to predict with few attributes and faster efficiency the risk of having heart disease.

Decisions are often made based on doctors’ intuition and experience rather than on the knowledge rich data hidden in the data set and databases. This practice leads to unwanted biases, errors and excessive medical costs which affects the quality of service provided to patients. Data mining holds great potential for the healthcare industry to enable health systems to systematically use data and analytics to identify inefficiencies and best practices that improve care and reduce costs. The successful application of data mining in highly visible fields like e-business, marketing and retail has led to its application in other industries and sectors. Among these sectors just discovering is healthcare. There is a wealth of data available within the healthcare systems. However, there is a lack of effective analysis tools to discover hidden relationships and trends in the data for African genres.

**OBJECTIVES**

**Main Objectives.**

The main objective of this research is to develop a heart prediction system. The system can discover and extract hidden knowledge associated with diseases from a historical heart data set Heart disease prediction system aims to exploit data mining techniques on medical data set to assist in the prediction of the heart diseases.

**Specific Objectives.**

• Provides new approach to concealed patterns in the data.

• Helps avoid human biasness.

• To implement Naïve Bayes Classifier that classifies the disease as per the input of the user.

• Reduce the cost of medical tests.

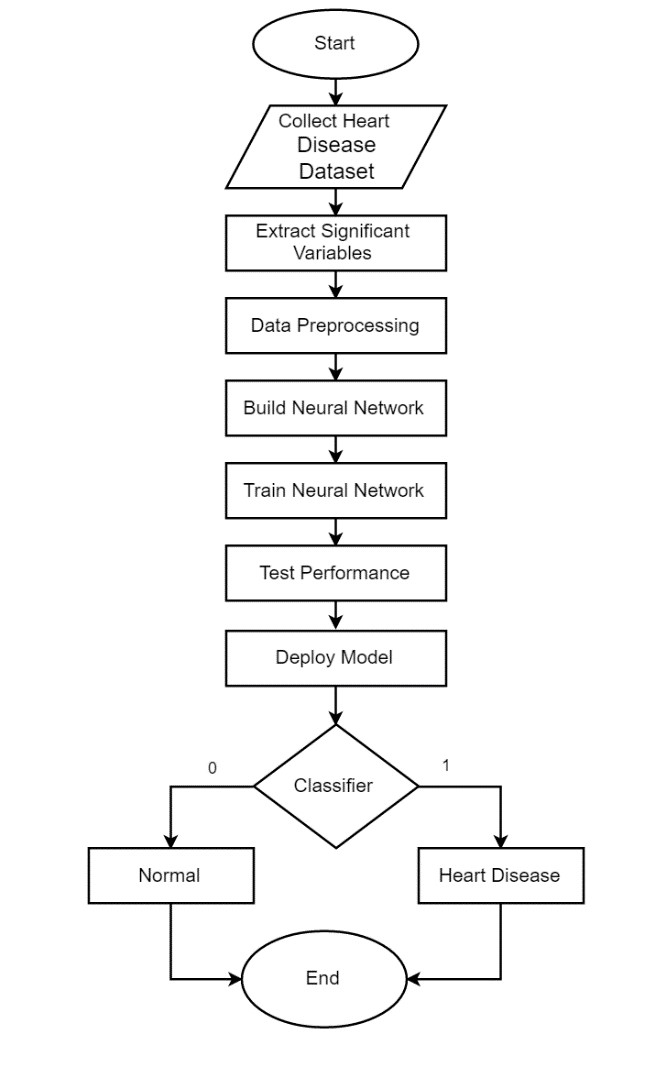
**DESIGN METHODOLOGY**

In this system we are implementing effective heart attack prediction system using Naïve Bayes algorithm. We can give the input as in CSV file or manual entry to the system. After taking input the algorithms apply on that input that is Naïve Bayes. After accessing data set the operation is performed and effective heart attack level is produced. The proposed system will add some more parameters significant to heart attack with their weight, age and the priority levels are by consulting expertise doctors and the medical experts. The heart attack prediction system designed to help the identify different risk levels of heart attack like normal, low or high and also giving the prescription details with related to the predicted result.

|  |  |
| --- | --- |
| **Algorithm Used** | **Accuracy** |
| Logistic Regression | 79% |
| SVC | 80% |
| K Neighbors Classifier | 81% |
| Random Forest Classifier | 85% |

**Project Flow Chart**

This will be the proposed flow chart that the system will look like



# Data Flow Diagram

**Dataset**

**Preprocessing**

**Accuracy Calculation**

**Pattern Matching**

**Prediction**

**Rule Generation**

**Result**

**Proposed Model**

**UCI Respository**

**Heart Attack**

**Dataset**

**Preprocessing**

**Decentralization**

**Conclusion**

**and**

**Suggestion**

**Feature Selection**

**Evaluation and**

**Comparison of**

**Results**

**Classifications**

**Model**

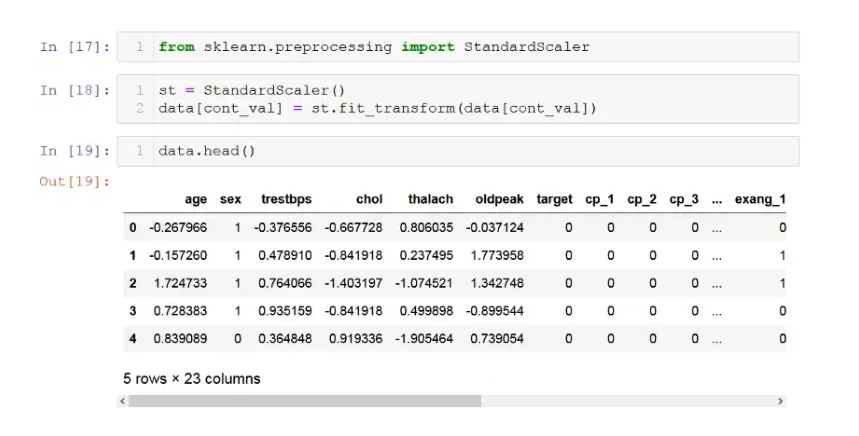
**DATASET AND FEATURES**

Our dataset is based on UCI heart Disease Data Set [6] and we have 303 instances. According to UCI, “This database contains 76 attributes, but all published experiments refer to using a subset of 13 of them.”We guess too many features will bring too much noise so people has done feature extraction and reduce 76 features to 13 features. To better understand the meaning of the features, we have the responsibility to explain some of the attributes of original dataset from UCI as follows:

## Dataset Features:

- age  
- sex  
- chest pain type (4 values)  
— Value 0: typical angina  
— Value 1: atypical angina  
— Value 2: non-anginal pain  
— Value 3: asymptomatic  
- trestbps: resting blood pressure (in mm Hg on admission to the hospital)  
- chol: serum cholestoral in mg/dl  
- fbs: (fasting blood sugar > 120 mg/dl) (1 = true; 0 = false)  
- restecg: resting electrocardiographic results  
— Value 0: normal  
— Value 1: having ST-T wave abnormality (T wave inversions and/or ST elevation or depression of > 0.05 mV)  
— Value 2: showing probable or definite left ventricular hypertrophy by Estes’ criteria  
- thalach: maximum heart rate achieved  
- exang: exercise induced angina (1 = yes; 0 = no)  
- oldpeak = ST depression induced by exercise relative to rest  
- slope: the slope of the peak exercise ST segment  
— Value 1: up-sloping  
— Value 2: flat  
— Value 3: down-sloping  
- ca: number of major vessels (0–3) colored by fluoroscope  
- thal: 3 = normal; 6 = fixed defect; 7 = reversible defect  
- target : 0=low risk of heart attack, 1=high risk of heart attack

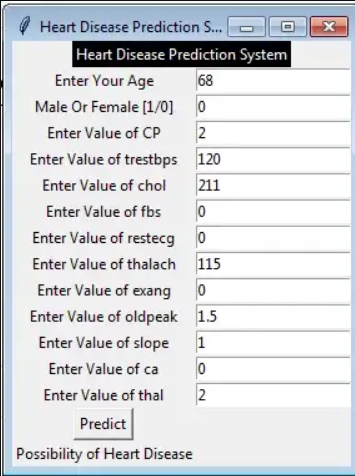
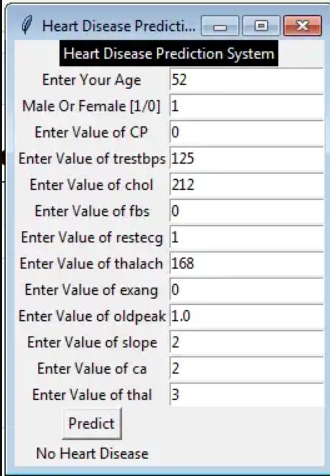
As you can see from the output above, This dataset contains 13 features and 1 target variable.

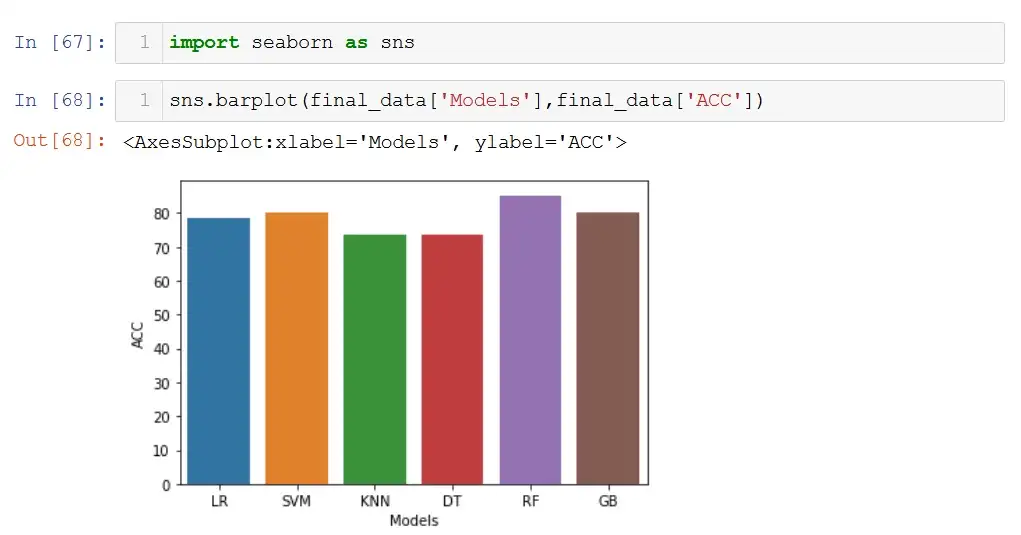
****

**RESULTS**

As we can see from the plot, a Random forest classifier is the best algorithm for this dataset.

We have trained our model on X\_train and y\_train (means on 80% data only). Before model deployment, we have to train our selected mode on 100% data.



**Conclusion**

The proposed system is GUI-based, user-friendly, scalable, reliable and an expandable system. The proposed working model can also help in reducing treatment costs by providing Initial diagnostics in time. The model can also serve the purpose of training tool for medical students and will be a soft diagnostic tool available for physician and cardiologist. General physicians can utilize this tool for initial diagnosis of cardio-patients. There are many possible improvements that could be explored to improve the scalability and accuracy of this prediction system. As we have developed a generalized system, in future we can use this system for the analysis of different data sets. The performance of the health’s diagnosis can be improved significantly by handling numerous class labels in the prediction process, and it can be another positive direction of research. In DM warehouse, generally, the dimensionality of the heart database is high, so identification and selection of significant attributes for better diagnosis of heart disease are very challenging tasks for future research.

**REFERENCES**

Content - Wikipedia - https://www.wikipedia.org/

Google - <https://www.google.com>

Dataset - <https://www.kaggle.com/rashikrahmanpritom/heart-attack-analysis-prediction-dataset>

Anaconda Software - <https://www.anaconda.com/products/distribution#Downloads>

Project Mentor - Mrs. Himani