Sri Lanka Institute of Information and Technology

Heart Failure Prediction

Statement of Work

Group 38

Fundamental of Data Mining - IT3051

Submitted by:

1. IT20216528 – Wanaguru D. R. S
2. IT20121792 – Kolonne R. U
3. IT20227340 – Dolawatta M. A
4. IT20012410 – Rajapaksha D. S. D
5. IT20457952 – Samarakoon S. M. D. H

1. Background

Heart Disease Predicting

This dataset includes data of patients who were tested to have heart failure or not by analyzing another 13 facts. These data dates from 1988 and consists of four databases: Cleveland, Hungary, Switzerland, and Long Beach V. Most of the people diagnose heart failure due to some other diseases and personal factors such as age, sex. This project will Predict the presence of heart disease in the patient by using 1025 rows of real historical data. As heart failure has become a reason for massive number of deaths in the present, our team hope to provide some help to predict heart failure status of a patient in advance.

This data set includes information such as

1. age
2. sex
3. chest pain type (4 values) (cp)
4. resting blood (trestbps)
5. serum cholesterol in mg/dl (chol)
6. fasting blood sugar > 120mg/dl (fbs)
7. resting electrocardiographic results (value 0,1,2) (restecg)
8. maximum heart rate achieved (thalach)
9. exercise induced angina (exang)
10. oldpeak = ST depression induced by exercise relative to rest (oldpeak)
11. the slope of the peak exercise ST segment (slope)
12. number of major vessels (0-3) colored by floursopy (ca)
13. thalassemia: 0 = normal, 1 = fixed defect, 2 = reversable defect (thal)
14. Target

Dataset: https://www.kaggle.com/datasets/johnsmith88/heart-disease-dataset

2. Scope of Work

This dataset contains 13 heart disease predicted attributes. The dataset is planned to use for building a predictive data model using python language. The model will use classification and regression mining tasks and predict whether the patient will have a heart disease or not. Our aim is to analyze the attributes and come up with an early decision about heart diseases. If there are some unwanted data in the dataset, as the first step we hope to remove them and clean the dataset in our preprocessing part.

We must divide our dataset into two groups as training set and testing set. Training set will be used to build the model and testing set will be used to determine the accuracy of the data set.

2.1 Business Problem

To predict whether the patient has a probability to diagnose heart disease or not by analyzing age, sex, chest pain type, resting blood, serum cholesterol, fasting blood sugar, resting electrocardiographic results, maximum heart rate achieved, exercise induced angina, ST depression induced by exercise relative to rest, the slope of the peak exercise ST segment, number of major vessels, and thalassemia

3. Activities

Step 1 –

- Finding a dataset to build the model.
- Observing dataset to understand the scenario
- Getting approval to the data set
- Submitting the SOW to get approval before the implementation

Step 2 –

- Reading the dataset and checking missing values
- Cleaning Data after checking dirtiness
- Perform exploratory analysis
- Feature selection
- Handling outliers
- Feature importance
- Splitting dataset and model building
- Play with multiple ml models (classification and regression) and get the best accuracy model

Step 3 –

- Deploying the model
- Testing the model in the user interface

Step 04 –

- Evaluating the final project report
- Submitting the Final report.

4. Approach

Implementation of the ML model

In this project we are going to follow some common approaches used in data mining such as data cleansing, feature selection and data conversion. In the data cleansing process, we have checked whether the missing values are available or not. We hope to replace numerical missing values with mean values of the relevant column and categorical/nominal missing values with most common value of the column. We hope to use classification and regression model to predict whether the patient is going to have a heart failure or not. After testing the accuracy of the model, we plan to improve the model further. After that we can use the model to predict patient's status.

Finally, we deploy the model using flask framework and help with HTML, CSS, JS technologies. Project will fully function after testing user interfaces.

5. Deliverables

Language

- Python

Libraries

- Numpy
- Pandas
- Matlplotlib
- Seaborn
- Ploty
- Sklearn
- XGBoost
- Pickle

Framework

- Flask (python base)

Other technologies

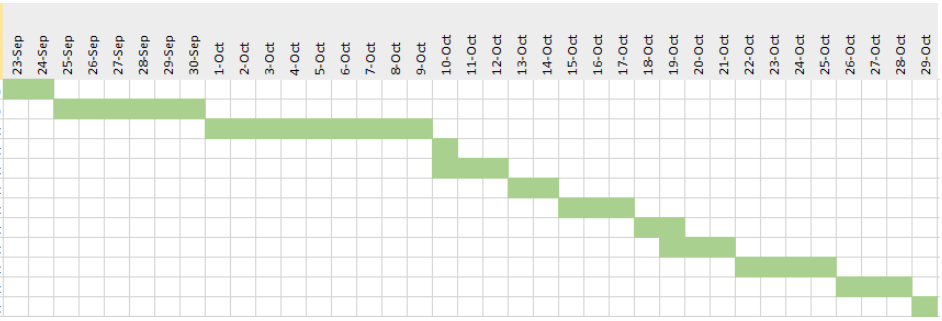
- HTML
- CSS
- JS

IDE

- Jupyter Notebook
- VS code

6. Project Plan & Timeline

Task Name	Duration	Start Date	End Date
Finding A dataset	2	23-Sep	24-Sep
Understanding the Real world Problem	6	25-Sep	30-Sep
Prepaire SOW	9	1-Oct	9-Oct
Approvel for the project	1	10-Oct	10-Oct
Exploring the dataset	3	10-Oct	12-Oct
Designing the model	2	13-Oct	14-Oct
Buiding the model	3	15-Oct	17-Oct
Designing the userinterface	2	18-Oct	19-Oct
Developing the userinterface	3	19-Oct	21-Oct
Get conclusion	4	22-Oct	25-Oct
Create the Final report	3	26-Oct	28-Oct
Presenting yhe Final report	1	29-Oct	29-Oct



7. Assumptions

- Records in the dataset are real
- Records in the dataset are accurate
- Dataset is static
- Data fields and the predictive column are related to each other.

8. Project team, Roles, and Responsibilities

IT number	Name	Responsibilities
IT20216528	Wanaguru D R S	• Data Preprocessing • Documentation • Frontend Application Development • Backend Development • Analyze the result
IT20121792	Kolonne R U	• Data Mining Model Building • Documentation • Frontend Application Development • Backend Development • Analyze the result
IT20227340	Dolawatta M A	• Data Mining Model Building • Documentation • Frontend Application Development • Backend Development • Select the best model
IT20012410	Rajapaksha D S D	• Data Mining Model Building • Documentation • Frontend Application Development • Backend Development • Select the best model
IT20457952	Samarakoon S M D H	• Data Preprocessing • Documentation • Frontend Application Development • Backend Development • Select the best model