

Final Project on “Predicting heart disease”

Problem Statement –

You are the Data Scientist at a medical research facility. The facility wants you to build a machine learning model to classify if given the data of a patient should tell if the patient is at risk of heart attack

Heart Disease Dataset:

UCI Heart Disease Dataset

(<https://archive.ics.uci.edu/ml/datasets/Heart+Disease?spm=5176.100239.blogcont54260.8.TRNGoO>)

Lab Environment: Jupiter Notebooks

Domain – Healthcare

Tasks to be done:

A) Data Analysis:

- a. Import the dataset
- b. Get information about dataset (mean, max, min, quartiles etc.)
- c. Find the correlation between all fields.

B) *Data Visualization:*

- a. Visualize the number of patients having a heart disease and not having a heart disease.
- b. Visualize the age and weather patient has disease or not
- c. Visualize correlation between all features using a heat map

C) Logistic Regression:

- a. Build a simple logistic regression model
 - i. Divide the dataset in 70:30 ratio
 - ii. Build the model on train set and predict the values on test set
 - iii. Build the confusion matrix and get the accuracy score

D) Decision Tree:

- a. Build a decision tree model
 - i. Divide the dataset in 70:30 ratio

Project Work

- ii. Build the model on train set and predict the values on test set
- iii. Build the confusion matrix and calculate the accuracy
- iv. Visualize the decision tree using the graphviz package

E) Random Forest:

- a. Build a Random Forest model
 - i. Divide the dataset in 70:30 ratio
 - ii. Build the model on train set and predict the values on test set
 - iii. Build the confusion matrix and calculate the accuracy
 - iv. Visualize the model using the graphviz package

F) Select the best model

- a. Print the confusion matrix of all classifiers
- b. Print the classification report of all classifiers
- c. Calculate Recall Precision and F1 score of all the models
- d. Visualize confusion matrix using heatmaps
- e. Select the best model based on the best accuracies