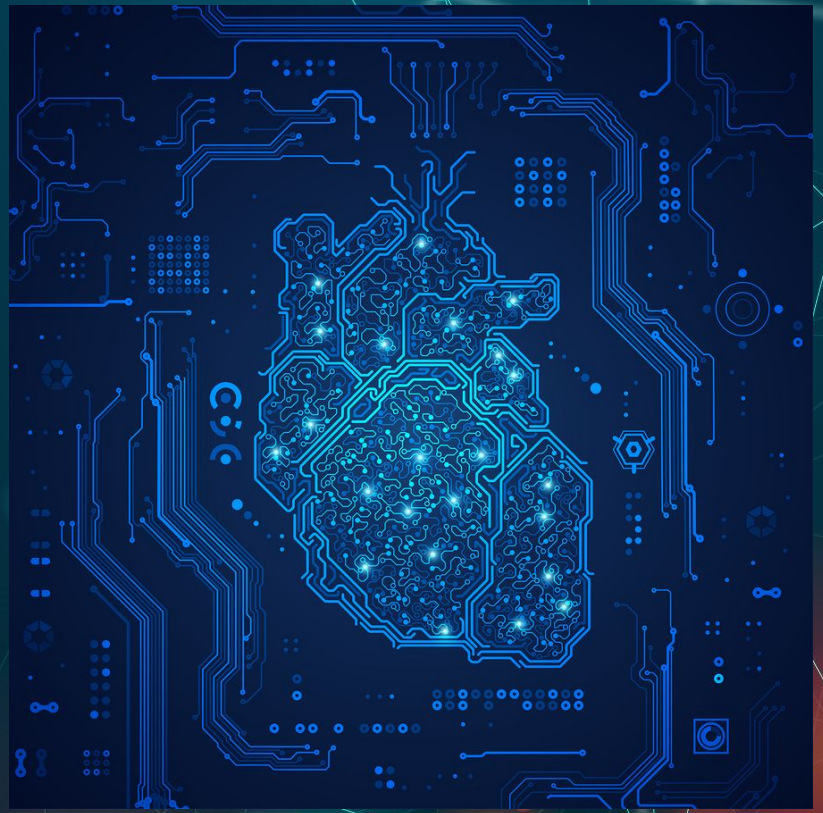
Heart Disease Prediction



Group number : 6

Abhinandhu A : AM.EN.U4AIE20002
Hariprasad S : AM.EN.U4AIE20035
M Mahadev : AM.EN.U4AIE20045
Marasani Jayasurya : AM.EN.U4AIE20048

INTRODUCTION

- Heart disease describes a range of conditions that affect your heart. Heart diseases include: Blood vessel disease, such as coronary artery disease. Heart rhythm problems (arrhythmias) Heart defects you're born with (congenital heart defects)

- With early direction of this with the available symptoms can save the life of the patient from critical conditions.

- High blood pressure, High blood cholesterol levels, Diabetes are some of the reasons for heart disease, of this is identified early and the person is made aware of his situation, it could be a life saver

PROBLEM

- Traditional cardiovascular risk factors such as blood pressure, cholesterol levels, diabetes, and smoking status, are valuable in predicting who will develop coronary heart disease (CHD).
- **Machine Learning** can play an essential role in predicting presence/absence of Heart diseases, Locomotor disorders,and more. Given all these attributes , a model is built in order to predict the whether he/she has Heart disease or not.
- Such information, if predicted well in advance, can provide important insights to doctors who can then adapt their diagnosis and treatment per patient basis.

IMPLEMENTATION

As the solution to this problem considering 4 models :

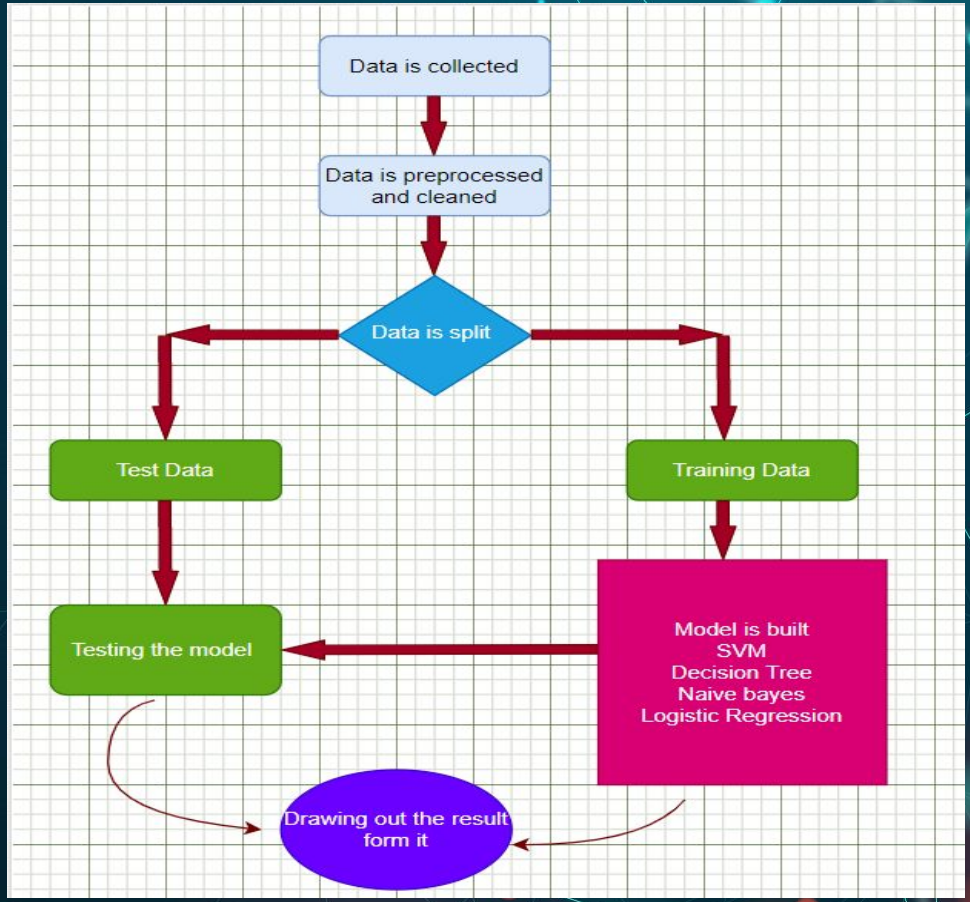
- SVM
- Decision tree
- Naive bayes
- Logistic regression

Taking these four, predictions are made by each model thereby identifying whether a person with the given attributes have heart disease or not.

Github link: https://github.com/hari4321/heart_disease_pred

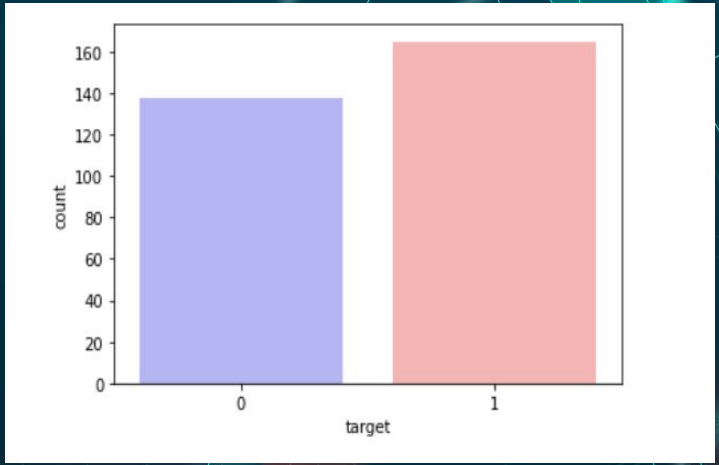
DESIGN

Implementation Flow chart

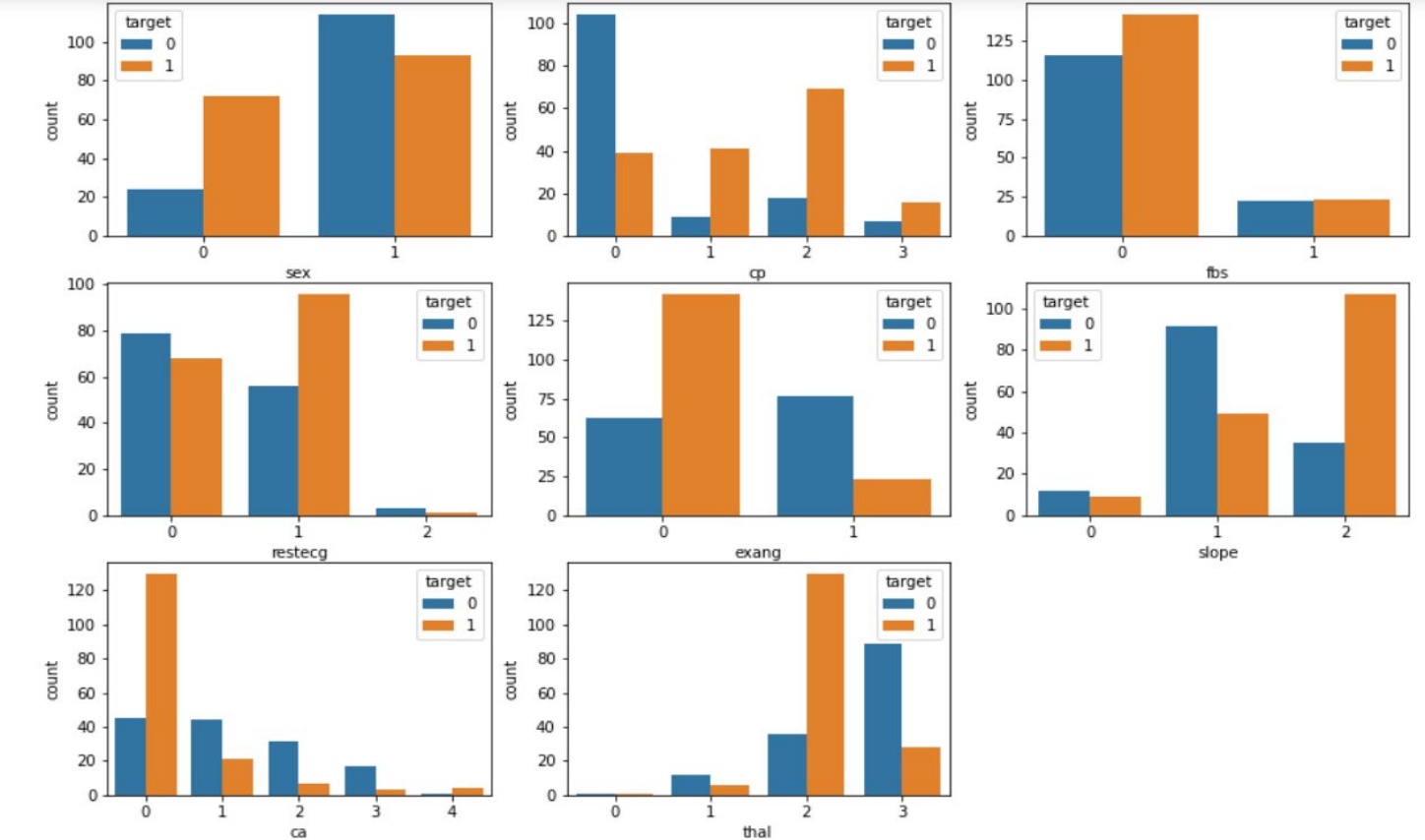


DATASET

- The dataset is taken from kaggle. There are 303 rows and 14 columns/attributes.
- There are no null value that is the dataset is already cleaned.
- The target attribute is the one which determines whether the person has disease or not.
  in the dataset it is of the ratio 165:138.



SAMPLE VISUALIZATION OF THE DATASET



OBSERVATION

- One of the major observation which we could draw was about the parameter turning in SVM:
  The value of kernel = 'linear', C=4, gamma=1

- The model which has the highest test data accuracy is Naive bayes

- The model which has the highest train data accuracy is SVM (with robust scaler)

RESULT

The accuracy scores for train and test data obtained for different models are :

Model	Scaling	Train data	Test data
Logistic regression	No	0.8553719008264463	0.8032786885245902
Naive bayes	No	0.8471074380165289	0.8471074380165289
Decision tree	No	1	0.7704918032786885
SVM	Min max scaler	0.913223140495867	0.8032786885245902
	Standard scaling	0.913223140495867	0.8032786885245902
	Robust Scaling	0.9173553719008262	0.7868852459016393
	Normalizer	0.6528925619834711	0.6557377049180328

REFERENCES

- https://seaborn.pydata.org/introduction.html
- https://youtu.be/eG663q4KjVw
- https://youtu.be/yZTBMMdP0ww
- https://youtu.be/-o3AxdVcUtQ
- https://youtu.be/MoM6migh0JM
- https://link.springer.com/article/10.1007/s42979-020-00365-y

REFERENCES

- https://youtu.be/TLdXM0A7SR8
- https://www.kaggle.com/akshukla283/heart-disease-prediction
- https://www.youtube.com/watch?v=xE9c1cJf48
- https://scikit-learn.org/stable/modules/naive_bayes.html
- https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LogisticRegression.html
- https://scikit-learn.org/stable/modules/tree.html

REFERENCES

- https://scikit-learn.org/stable/modules/svm.html
- https://www.youtube.com/watch?v=xE9c1c1f48A
- https://machinelearningmastery.com/standardscaler-and-minmaxscaler-transforms-in-python/
- https://iopscience.iop.org/article/10.1088/1757-899X/1022/1/012072/meta