

Final Project Report

Introduction to Data Analytics

Project Title:
Heart Attack Prediction

Prepared by:
Kartik Nain(N01516607)

ITE 5201 – Summer 2022
Humber College

1. Problem Statement

Prediction of Heart Attack Occurrence among people of different ages.

2. Dataset Description

	Age	Sex	CPTYPE	RestBP	Chol	FastingBS	RestECG	Max.HR	EIAng	OldPeak	CA	ThalRate	output
0	63	1	3	145	233	1	0	150	0	2.3	0	1	1
1	37	1	2	130	250	0	1	187	0	3.5	0	2	1
2	41	0	1	130	204	0	0	172	0	1.4	0	2	1
3	56	1	1	120	236	0	1	178	0	0.8	0	2	1
4	57	0	0	120	354	0	1	163	1	0.6	0	2	1

The dataset has 13 columns providing information about various factors such as age, sex, Resting Blood Pressure, ECG, Cholesterol Level and many more.

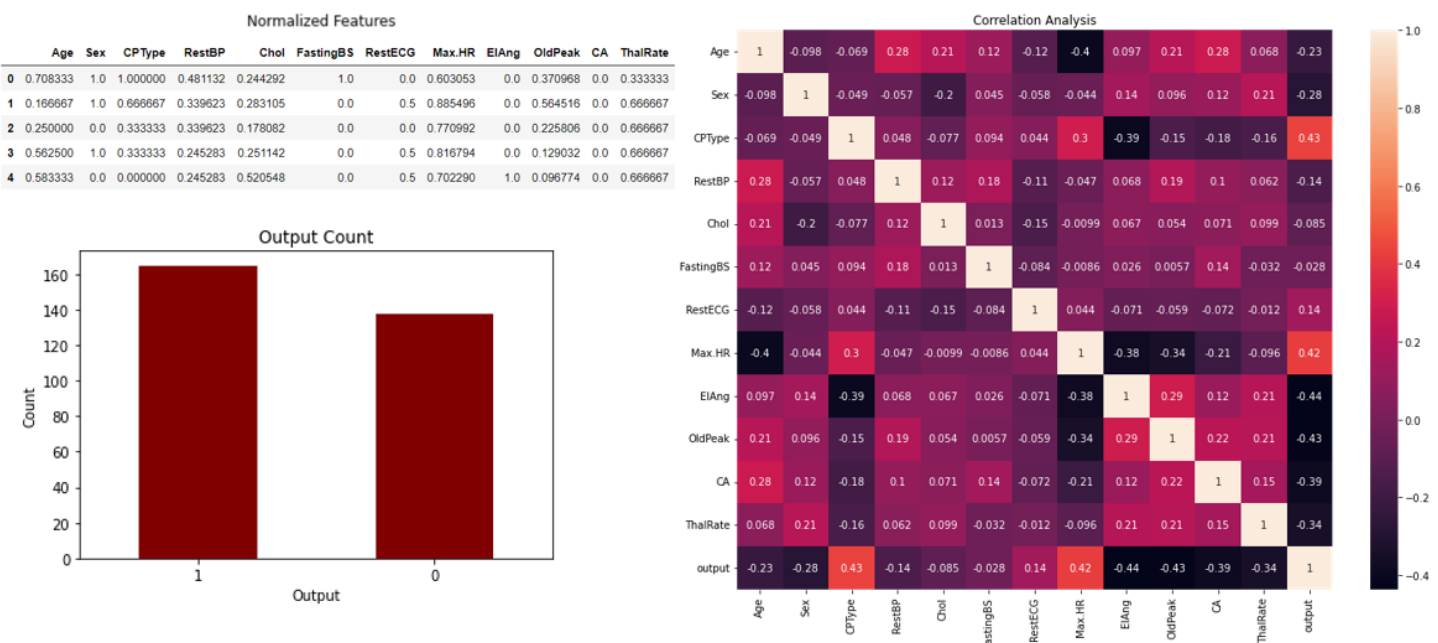
“Age”, “RestBP”, “Chol”, “Max.HR”, “OldPeak” columns consist of numerical data while the rest consist of categorical data.

Out of these 13 columns, 12 are used as features to predict the output which is the “output” column. ‘1’ in the “output” column indicates that the person will have a heart attack while ‘0’ indicates otherwise.

3. Dataset Analysis and Observations

The dataset doesn’t require any pre-processing like replacing or deleting null or duplicate values respectively as there are none. Also, the categorical data was already in binary form so one-hot encoding was not required either.

Normalization of features is carried out and exploratory data analysis is performed. According to the bar chart, 165 people in the dataset have suffered from a heart attack while 138 have not after factoring in all the above mentioned features.

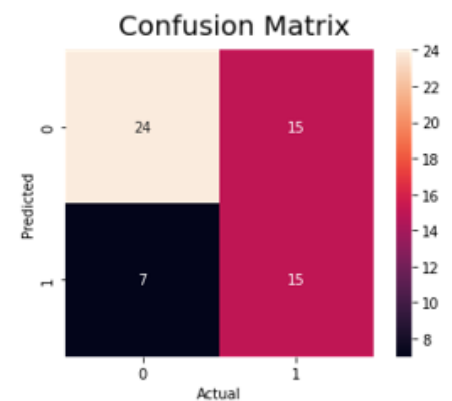


4. Proposed Prediction Models

Two prediction models are used:-

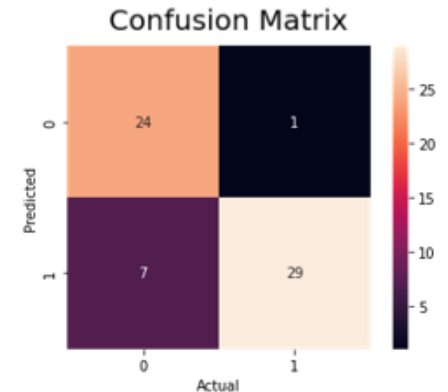
Perceptron Model:

Classification Report				
	precision	recall	f1-score	support
0	0.62	0.77	0.69	31
1	0.68	0.50	0.58	30
accuracy			0.64	61
macro avg	0.65	0.64	0.63	61
weighted avg	0.65	0.64	0.63	61



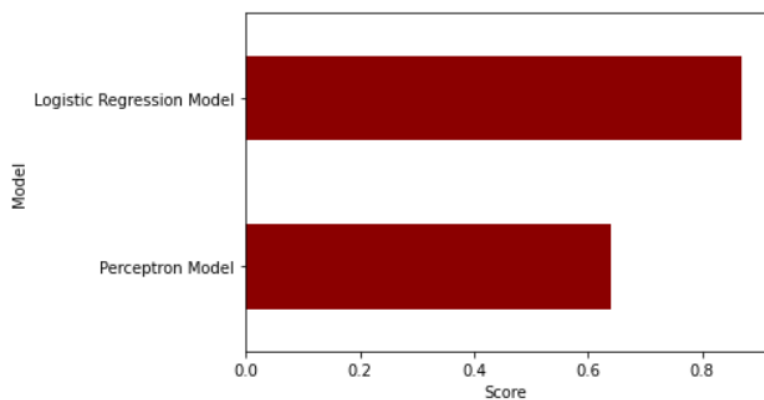
Logistic Regression Model:

Classification Report				
	precision	recall	f1-score	support
0	0.96	0.77	0.86	31
1	0.81	0.97	0.88	30
accuracy			0.87	61
macro avg	0.88	0.87	0.87	61
weighted avg	0.88	0.87	0.87	61



5. Results and Discussions:

After comparing both the models, it is clear that an accuracy jump to 87% is observed when implementing Logistic Regression Model in place of Perceptron Model which has 64% accuracy.



Result: With the help of the LR Model, we can predict the occurrence of heart attack among people using certain features with an accuracy of 87%.