

Alena Sanchez

April 4, 2022

Professor Brahma

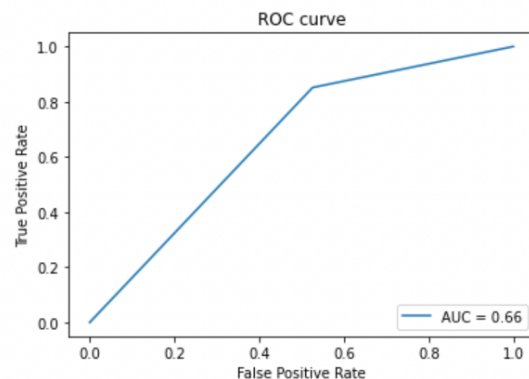
CA-05

Part 3: Evaluating Model Performance

With this assignment, we have developed a program to help predict whether a patient is at risk of cardiovascular disease (CVD) based on various features. With this, we must understand the importance of evaluating model performance as it is critical to have higher and better performing models when working in the healthcare domain. From the analysis we were able to conclude that race, hlthlm25, and parrptdiab have the highest influence on CVD risk. hlthlm25 is the “Frequency that physical health or emotional problems interfered with normal social activities” and parrptdiab is the “History of Diabetes”.

	feature	coefficients
1	race	0.835409
15	hlthlm25	0.574986
11	parrptdiab	0.487852
2	educat	0.300273
3	mstat	0.140960
14	tired25	0.138130
12	bend25	0.118151
10	srhype	0.098597
6	waist	0.073862
4	hip	0.046018
13	happy25	0.038402
9	tea15	0.034349
5	neck20	0.033492
7	av_weight_kg	0.018081
0	age_s1	0.004720
8	cgpkylr	0.000061

TRUE NEGATIVE:
217
=====
FALSE POSITIVE:
241
=====
FALSE NEGATIVE:
91
=====
TRUE POSITIVE:
521



Overall, we can conclude that the model's performance is not the strongest for this given domain as the AUC is below .7. Additionally, we can see that there is high rate of False Positives, which is not a good indicator of a good model for one analyzing patient data. Therefore, I would suggest that a new model be developed in order to accurately predict CVD risk.