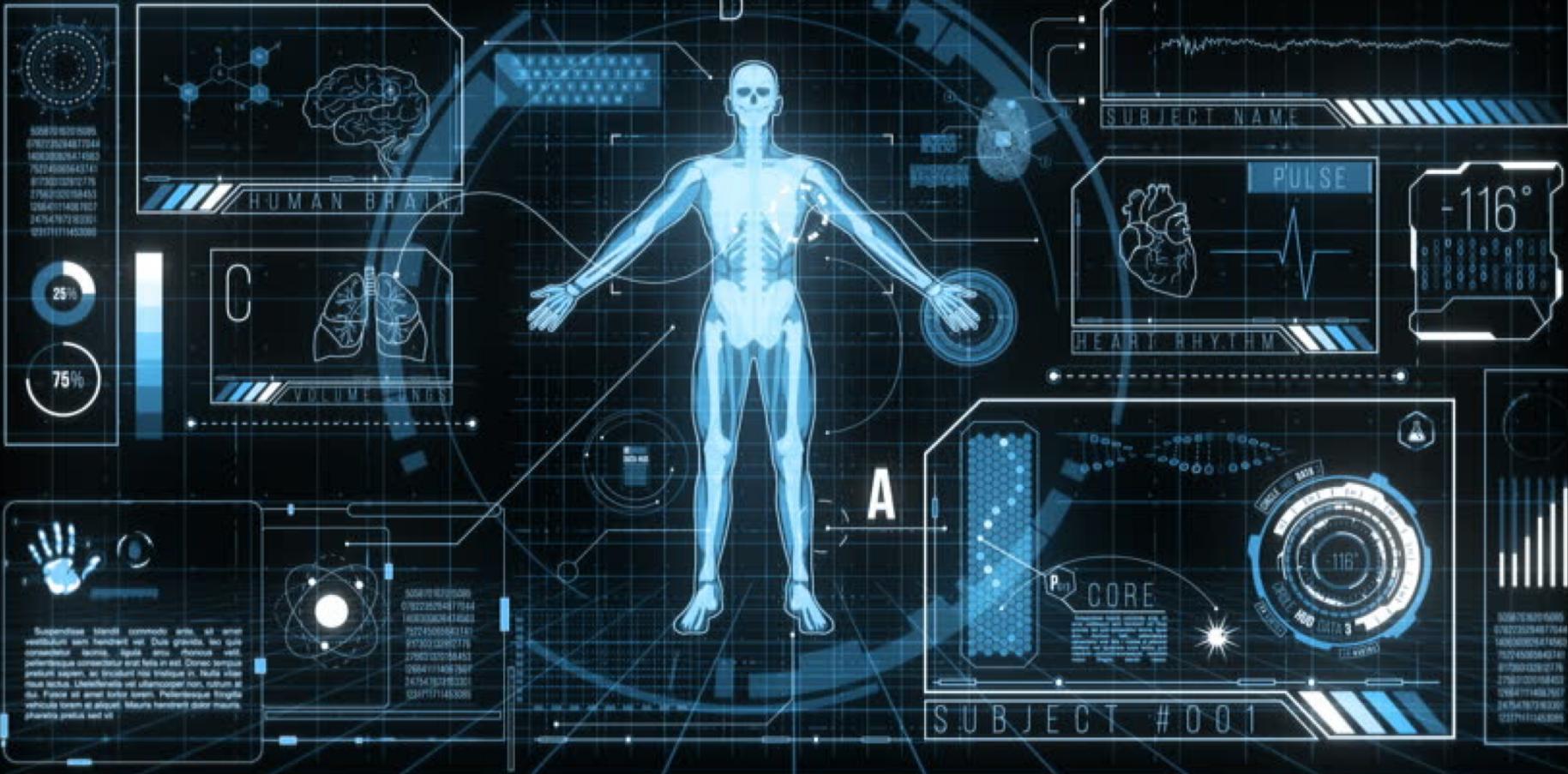
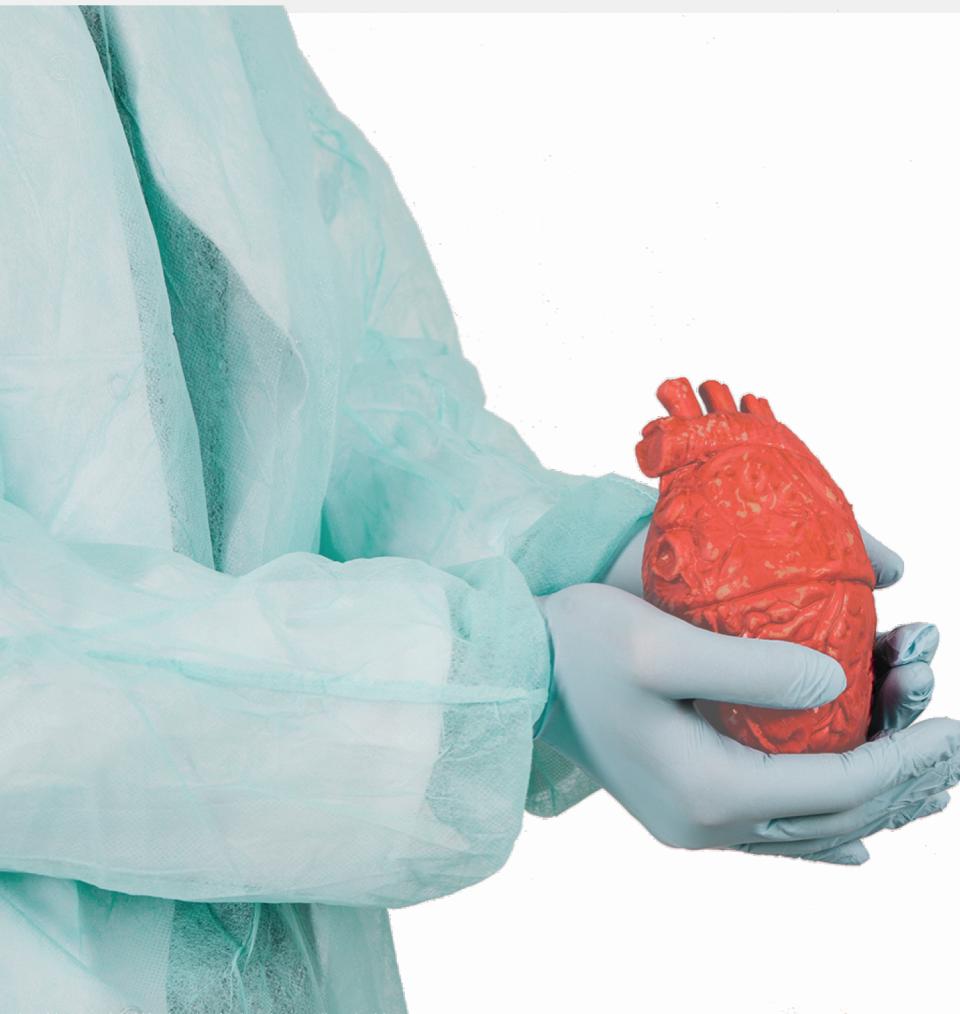


PREDICTING CARDIO VASCULAR DISEASE

Dhvani Patel

|| MEDICINE FILES ||





31%

DEATHS EACH YEAR

50%

NON COMMUNICABLE
DISEASE

\$219 Billion

IN US ALONE



“The problem with heart disease is that
the first symptom is often fatal....”

—MICHAEL PHELPS



LEADING RISK FACTORS



PREDICT
HEART DISEASE
FROM
EARLY
WARNINGS SIGNS...

METHOD



01
**DATA
COLLECTION**
70,000 INSTANCES
11 FEATURES

02
**DATA
CLEANING**
OUTLIERS
TRANSFORMATIONS

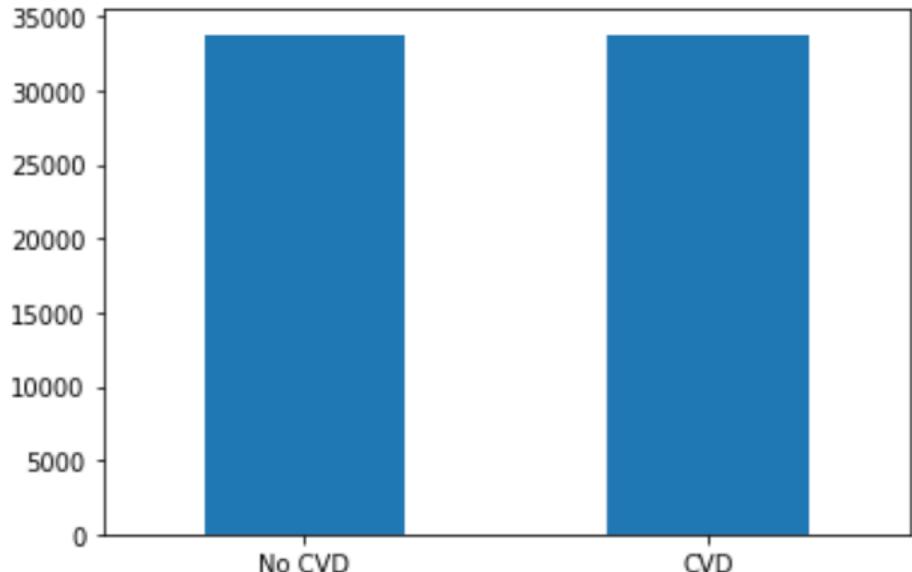
03
**FEATURE
ENGINEERING**
FEATURE SCALING
ADDING FEATURES
FEATURE REDUCTION

04
MODELING
RANGE OF
CLASSIFICATION
MODELS

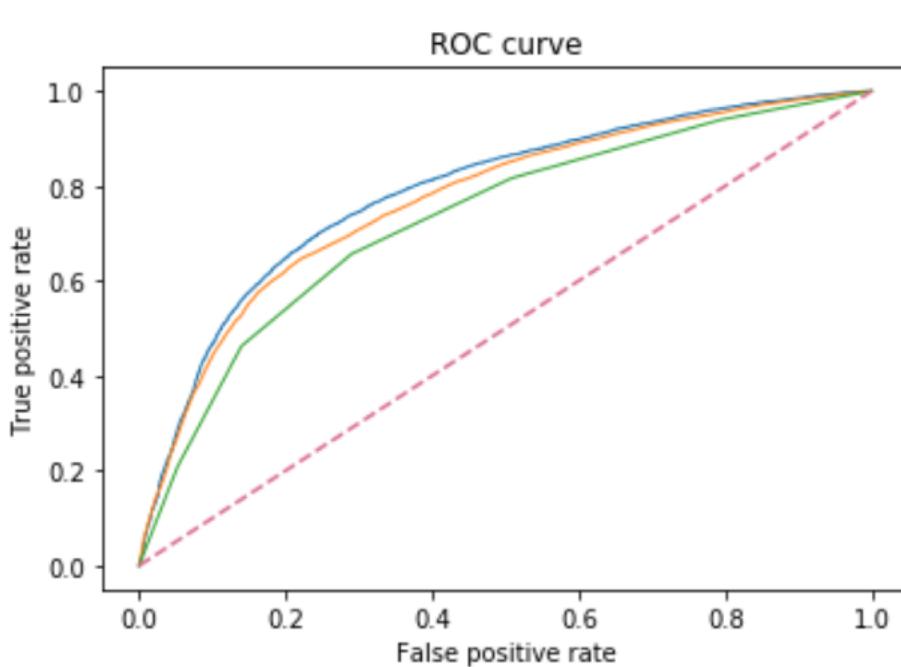
PERFORMANCE METRIC

- ACCURACY SCORE
- AUROC SCORE

DATA BALANCE



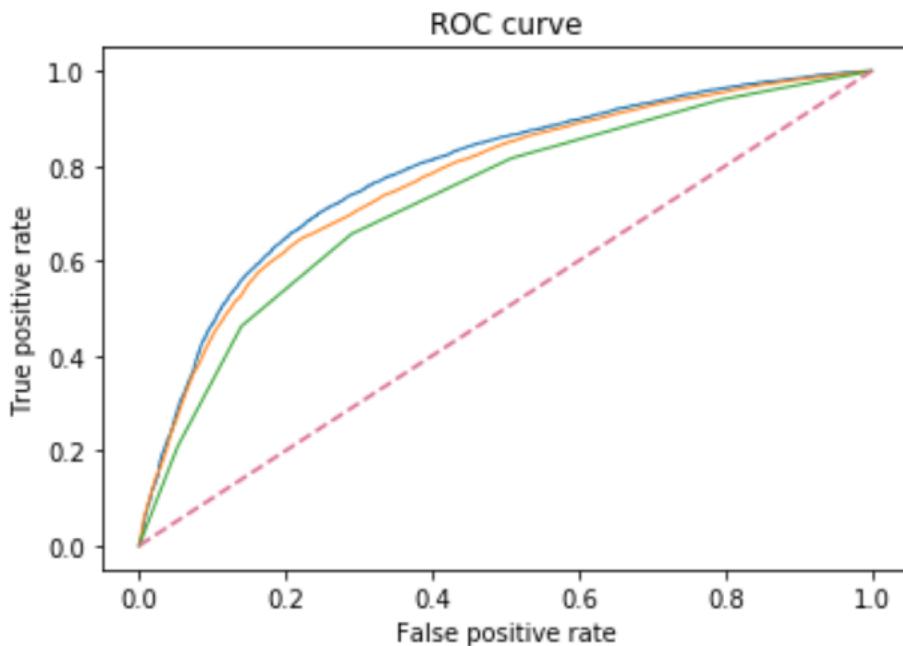
MODEL COMPARISON



AUROC Score

- LOGISTIC REGRESSION : 0.78
- RANDOM FOREST CLASSIFIER : 0.77
- K- NEAREST NEIGHBOURS : 0.73

MODEL COMPARISON



AUROC Score

- LOGISTIC REGRESSION : 0.78
- RANDOM FOREST CLASSIFIER : 0.77
- K- NEAREST NEIGHBOURS : 0.73

Accuracy Score

- LOGISTIC REGRESSION : 0.72
- RANDOM FOREST CLASSIFIER : 0.71
- K- NEAREST NEIGHBOURS : 0.68

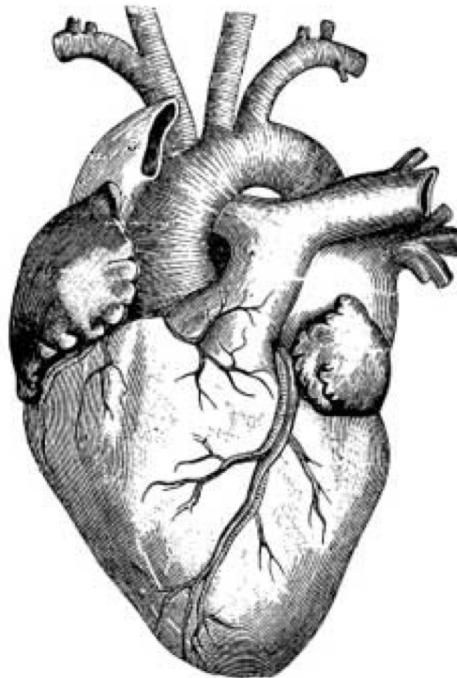
IMPORTANCE FEATURES

Systolic Blood
Pressure

Age

Cholesterol

Body Mass
Index



IMPORTANCE FEATURES

Smoking

Systolic Blood Pressure

Height

Age

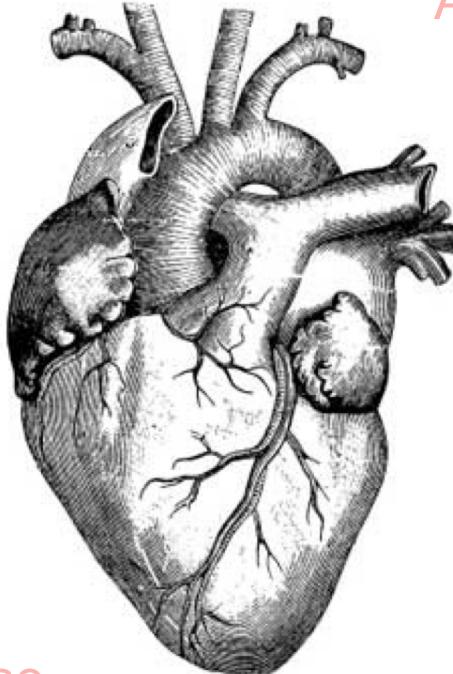
Glucose

Activity

Cholesterol

Alcohol

Body Mass Index



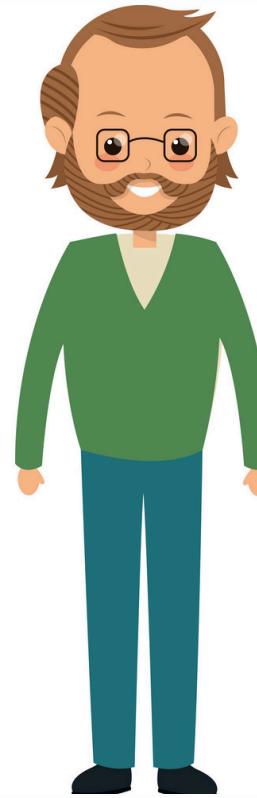


MODEL PERFORMANCE BY AGE GROUP

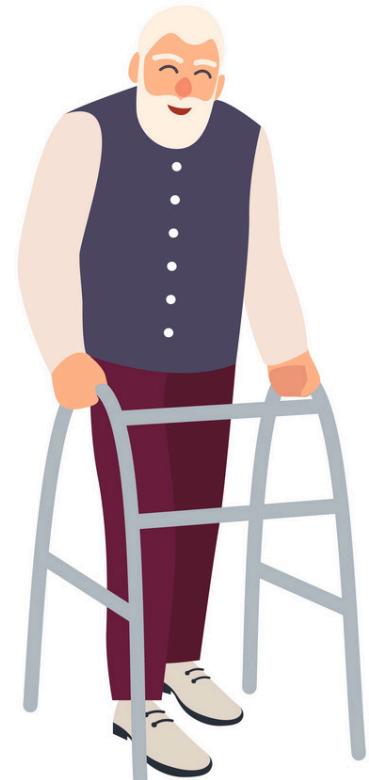
SEGMENTING BY AGE



AGE 40-50



AGE 50-60



AGE 60-Above

MODEL PERFORMANCE BY AGE GROUP



AGE 40-50

ACCURACY SCORE : 0.78



AGE 50-60

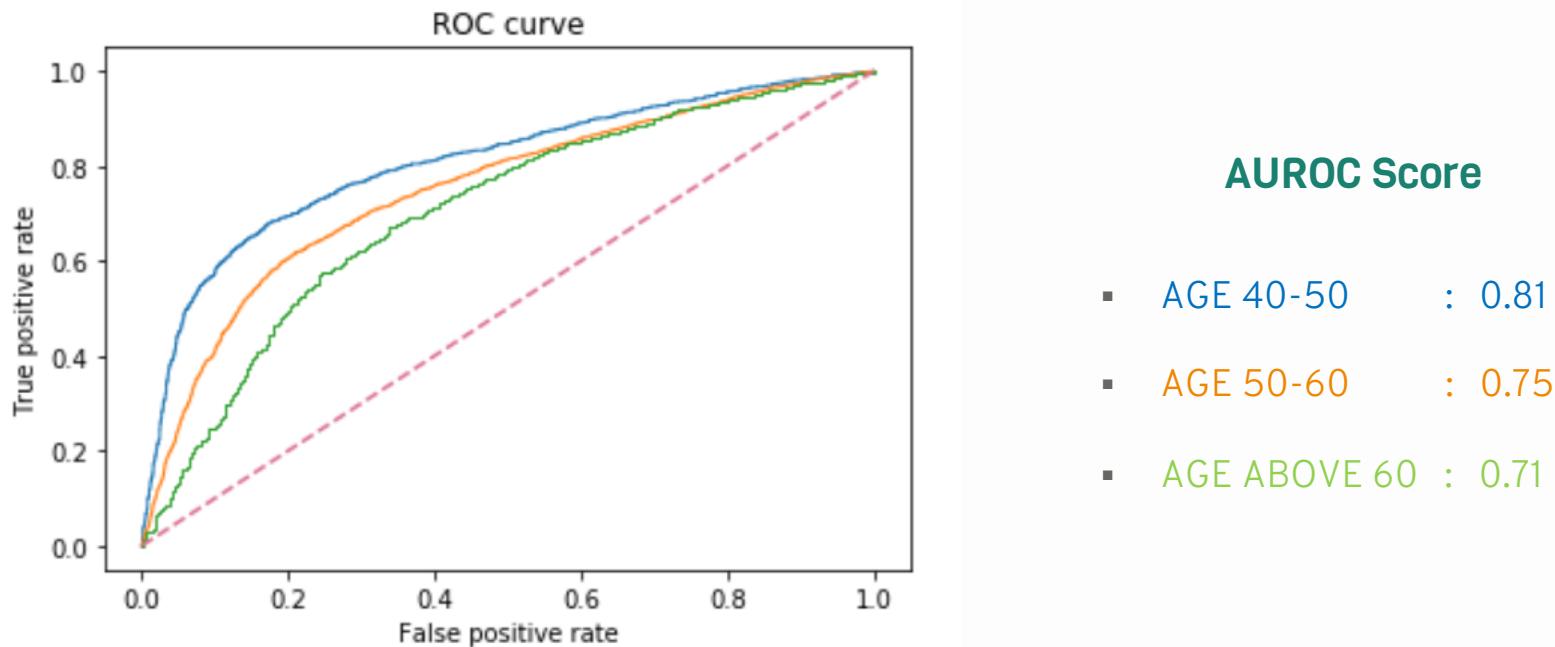
ACCURACY SCORE : 0.70



AGE 60-Above

ACCURACY SCORE : 0.70

MODEL PERFORMANCE



INFERENCE

MORE RECOVERY TIME

BETTER DIET

BETTER LIFESTYLE

PULSE TRACKING
BIOSENSOR



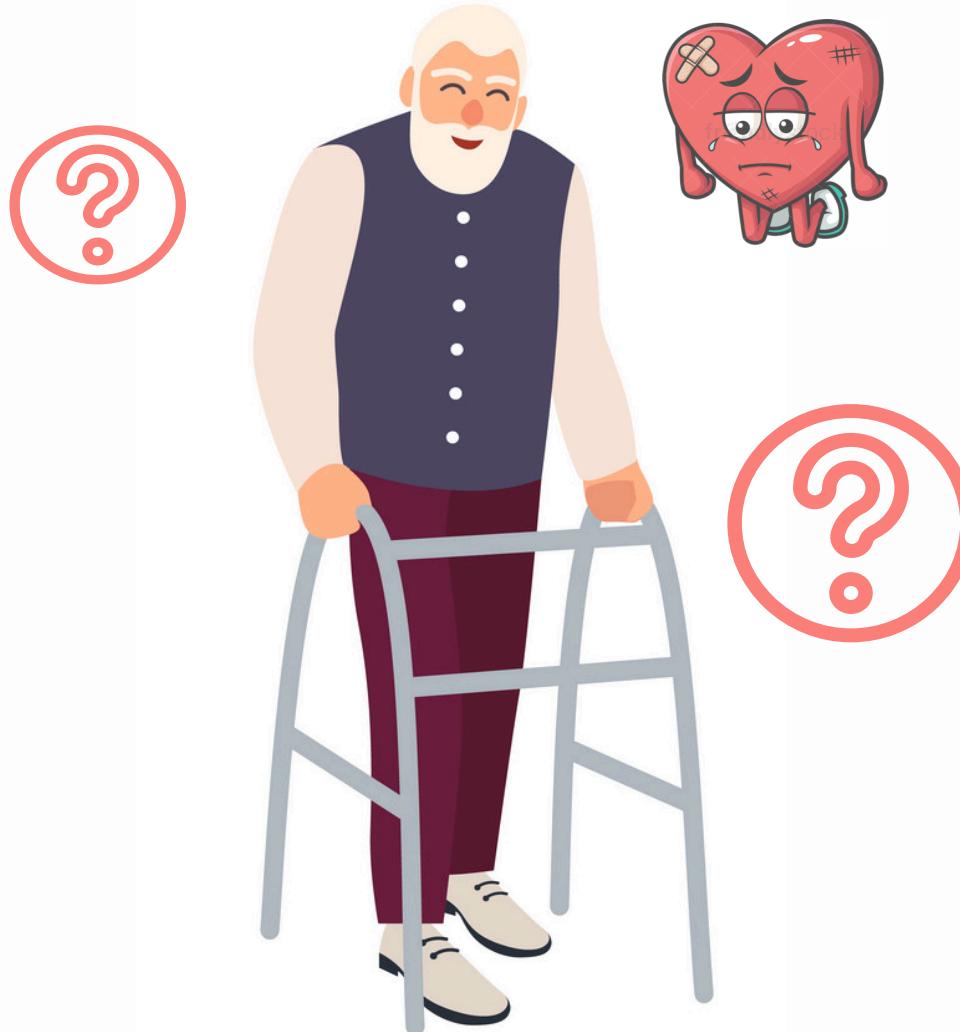
INFERENCE

INVESTIGATE OTHER
FEATURES



INFERENCE

INVESTIGATE OTHER
FEATURES



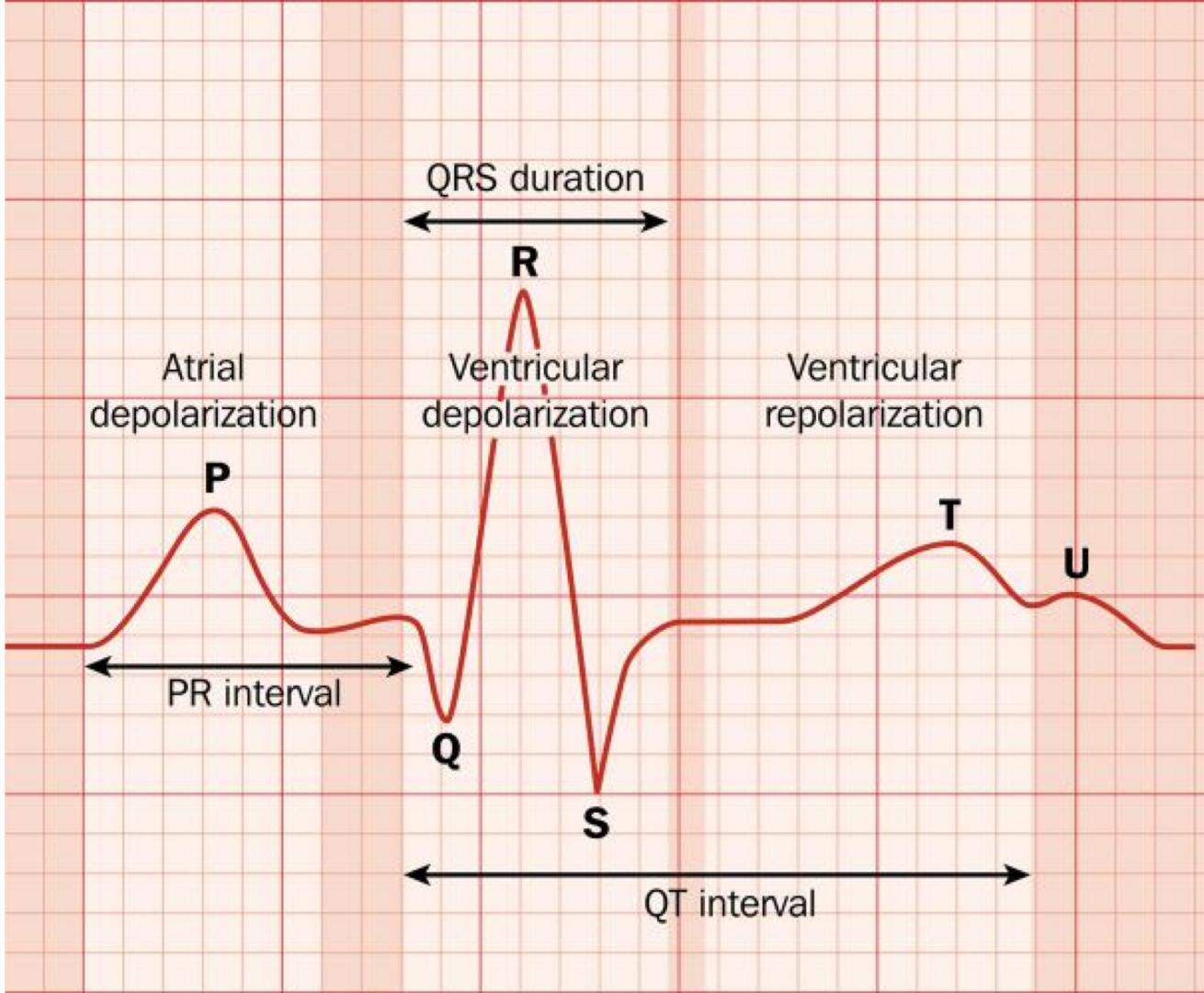
Future Works

Analyzing Biosensor
Data to alert
Before Onset of
Major CVD
Condition like
Heart Attack



Future Works

- ECG Data
- Pulse Rate
- Precise History



THANK YOU

Questions ??

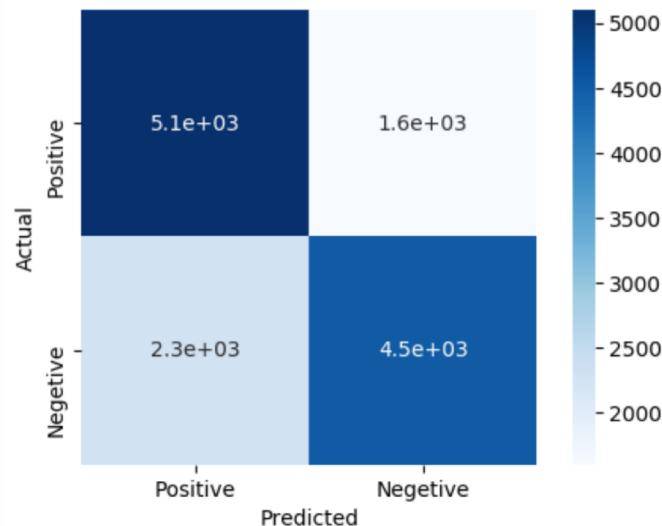


APPENDIX



BASELINE MODELING LOGISTIC REGRESSION

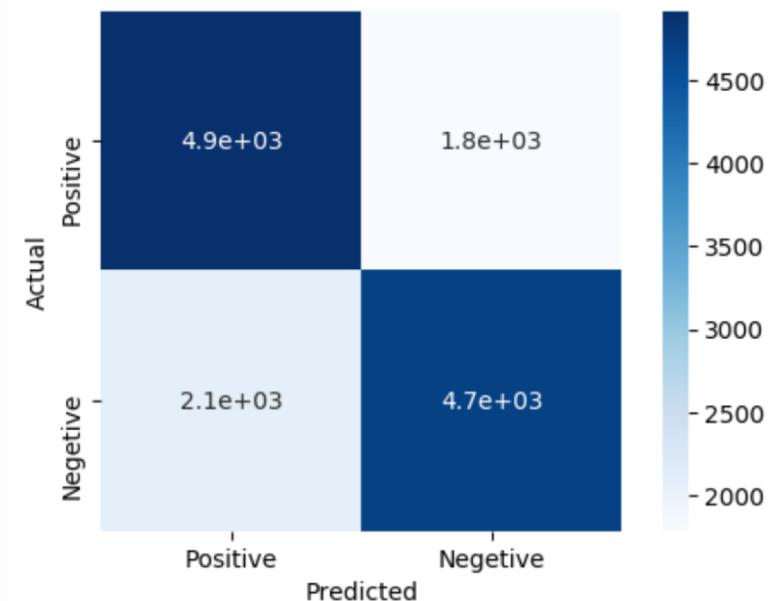
Logistic Regression: 0.7127738306690349
Precision Score: 0.7388607801534193
Recall Score: 0.6649529964747356
F1 Score: 0.6999613451874759



	Feature	Co-efficient
5	cholesterol	0.449485
0	gender	0.260762
3	ap_hi	0.052983
10	Age in Years	0.030161
2	weight	0.015898
4	ap_lo	-0.001627
1	height	-0.056444
6	gluc	-0.138515
7	smoke	-0.203580
8	alco	-0.245378
9	active	-0.387316

Cross Validation Scores: [0.71008326 0.71720629 0.71211841 0.71320196 0.70940883]

BASELINE MODELING RANDOM FOREST



Random Forest Modelling: 0.7131438721136767

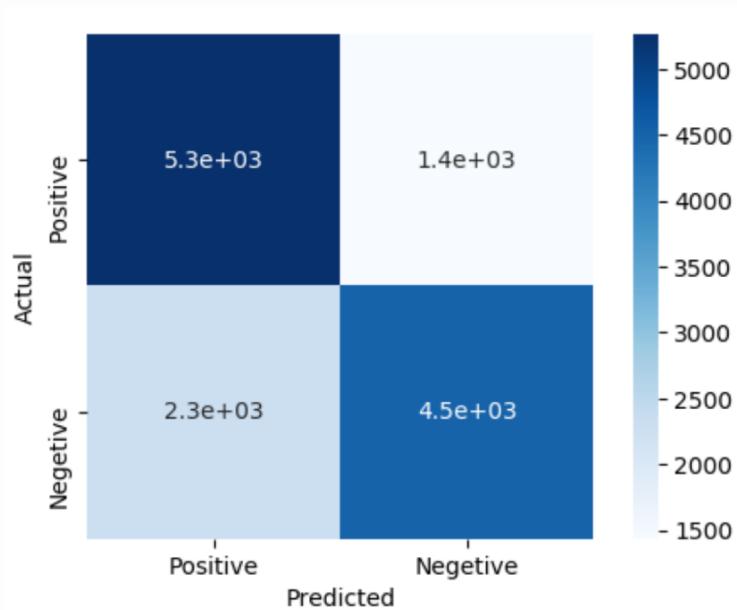
Precision Score: 0.7252612169637369

Recall Score: 0.6933019976498237

F1 Score: 0.7089215980775008

Cross Validation Scores: [0.71045328 0.71526364 0.7147086 0.70691091 0.71088907]

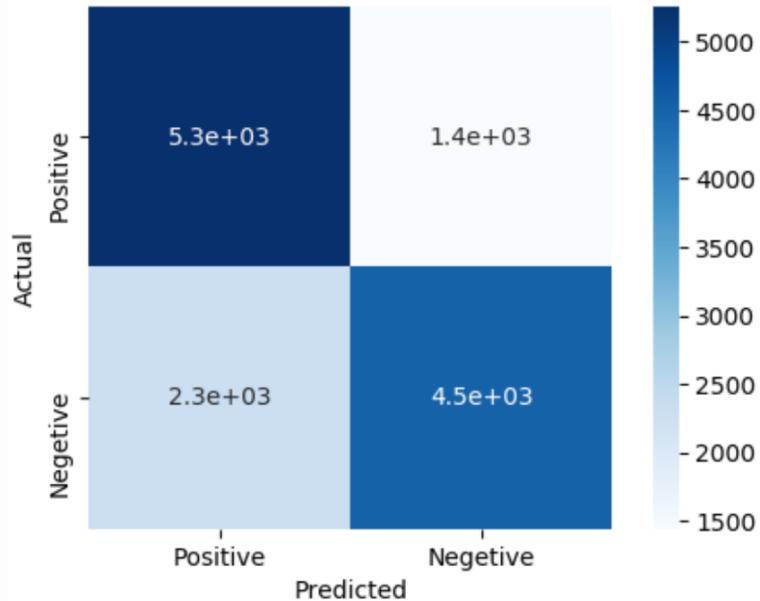
BASELINE MODELING LOGISTIC REGRESION – STANDERDIZED SCALING



Logistic Regression: 0.7244671403197158
Precision Score: 0.7591130522425668
Recall Score: 0.6637779083431258
F1 Score: 0.7082517044118799
Cross Validation Scores: [0.72460685 0.73145236 0.72673451 0.72254603 0.72411879]

	Feature	Co-efficient
3	ap_hi	0.922370
5	cholesterol	0.493795
10	Age in Years	0.347699
2	weight	0.155334
4	ap_lo	0.108685
0	gender	-0.020711
1	height	-0.024291
6	gluc	-0.123400
7	smoke	-0.144039
8	alco	-0.211939
9	active	-0.217840

FEATURE ENGINEERING – BMI LOGISTIC REGRESSION



Logistic Regression: 0.7244671403197158

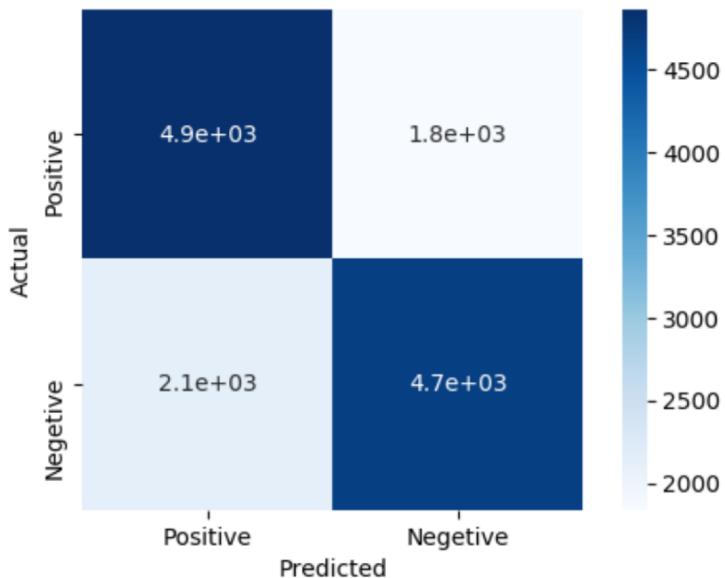
Precision Score: 0.7583319376988779

Recall Score: 0.6650998824911868

F1 Score: 0.7086626496595978

Cross Validation Scores: [0.72183164 0.72895467 0.72368178 0.72115829 0.72365621]

	Feature	Co-efficient
5	cholesterol	0.496090
2	weight	0.070827
3	ap_hi	0.056713
10	Age in Years	0.046796
0	gender	0.037637
4	ap_lo	0.008676
1	height	-0.070875
7	smoke	-0.081191
8	alco	-0.083471
6	gluc	-0.150642
11	BMI	-0.155744
9	active	-0.225888



Random Forest Modelling: 0.7068531675547661

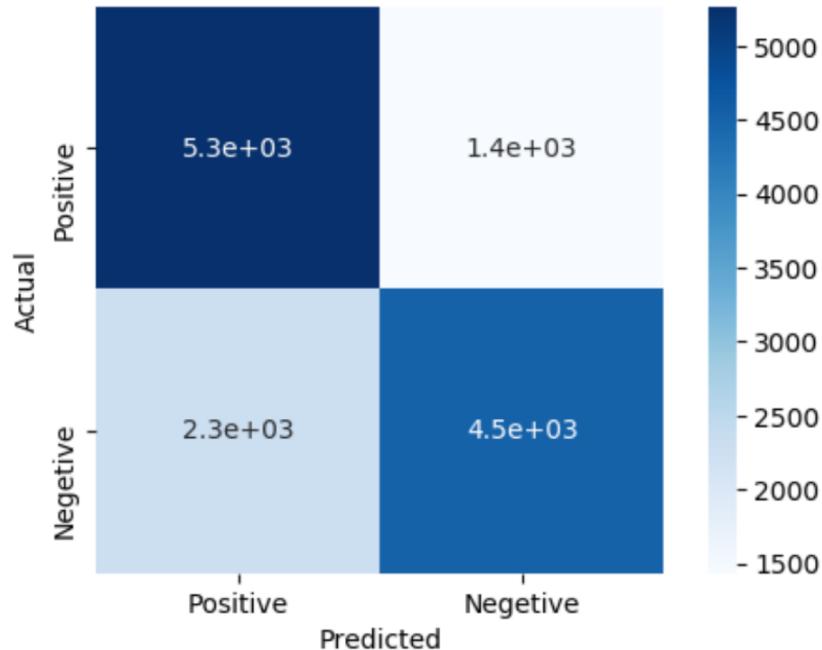
Precision Score: 0.7182278092902039

Recall Score: 0.6881609870740305

F1 Score: 0.7028730027754858

Cross Validation Scores: [0.70601295 0.70962072 0.71295097 0.70432047 0.71486724]

FEATURE ENGINEERING – BMI LOGISTIC REGRESSION – STANDERDIZED SCALING



Logistic Regression: 0.7244671403197158

Precision Score: 0.7589390632868894

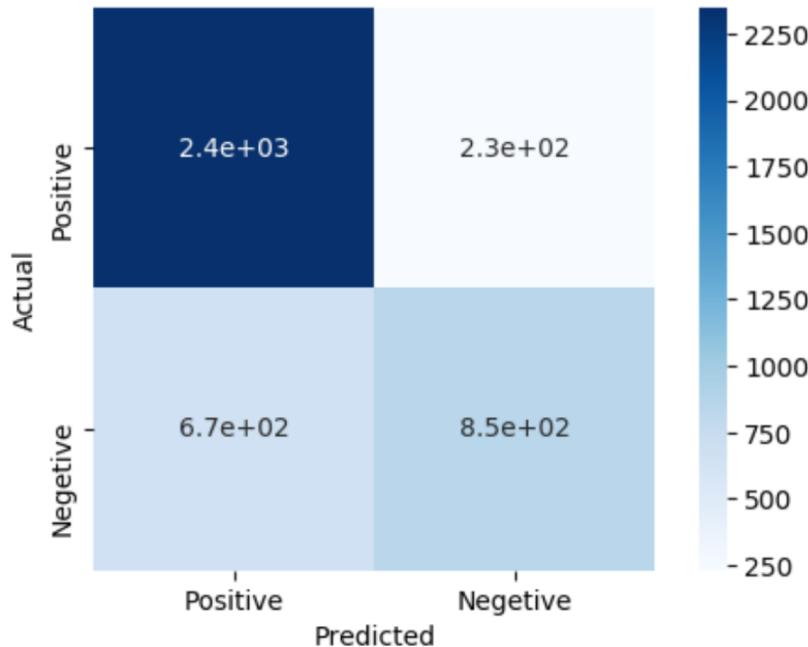
Recall Score: 0.6640716803760282

F1 Score: 0.70834312573443

Cross Validation Scores: [0.72460685 0.73117484 0.72636448 0.72254603 0.72411879]

Feature	Co-efficient	
3	ap_hi	0.922225
5	cholesterol	0.493519
10	Age in Years	0.347548
2	weight	0.173139
4	ap_lo	0.108607
0	gender	-0.018674
11	BMI	-0.019183
1	height	-0.035592
6	gluc	-0.123453
7	smoke	-0.143646
8	alco	-0.212372
9	active	-0.217891

AGE 40-50
LOGISTIC REGRESSION



Feature	Co-efficient
5 cholesterol	0.672712
0 gender	0.132172
2 weight	0.092348
3 ap_hi	0.079229
10 Age in Years	0.045983
4 ap_lo	0.012376
6 gluc	-0.069198
1 height	-0.090427
12 Sys/Dia	-0.134983
7 smoke	-0.174498
9 active	-0.191237
11 BMI	-0.211661
8 alco	-0.305551

Logistic Regression: 0.781005859375

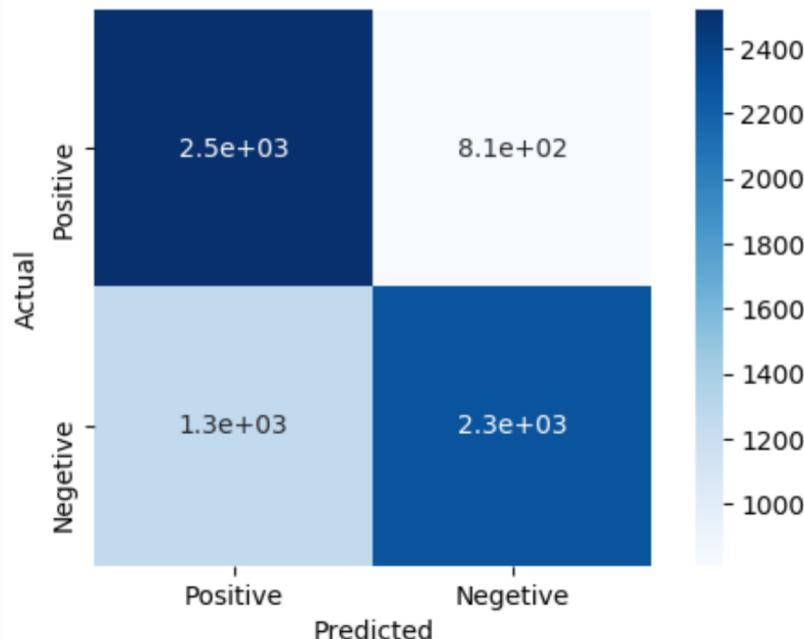
Precision Score: 0.7861111111111111

Recall Score: 0.5603960396039604

F1 Score: 0.654335260115607

Cross Validation Scores: [0.79829112 0.77815075 0.78730546 0.79859628 0.77899878]

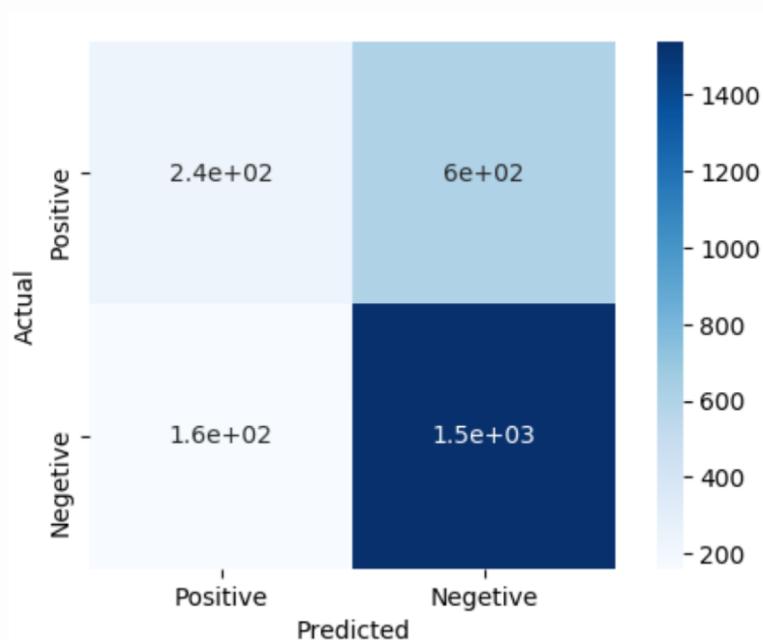
AGE 50-60
LOGISTIC REGRESSION



	Feature	Co-efficient
5	cholesterol	0.477015
0	gender	0.069863
2	weight	0.063709
3	ap_hi	0.056764
10	Age in Years	0.037763
4	ap_lo	0.004934
1	height	-0.065284
12	Sys/Dia	-0.075405
8	alco	-0.121282
6	gluc	-0.123228
11	BMI	-0.137403
7	smoke	-0.166989
9	active	-0.308717

Logistic Regression: 0.6980802792321117
Precision Score: 0.7374717103136114
Recall Score: 0.6434414668547249
F1 Score: 0.687255197348599

Cross Validation Scores: [0.7053263 0.71277949 0.70332667 0.71314306 0.69818182]



Feature	Co-efficient	
5	cholesterol	0.464560
10	Age in Years	0.074898
2	weight	0.065419
3	ap_hi	0.039444
4	ap_lo	0.004047
12	Sys/Dia	-0.019284
8	alco	-0.039475
1	height	-0.060612
7	smoke	-0.077373
11	BMI	-0.155693
6	gluc	-0.170918
0	gender	-0.175580
9	active	-0.198909

Logistic Regression: 0.7003937007874016

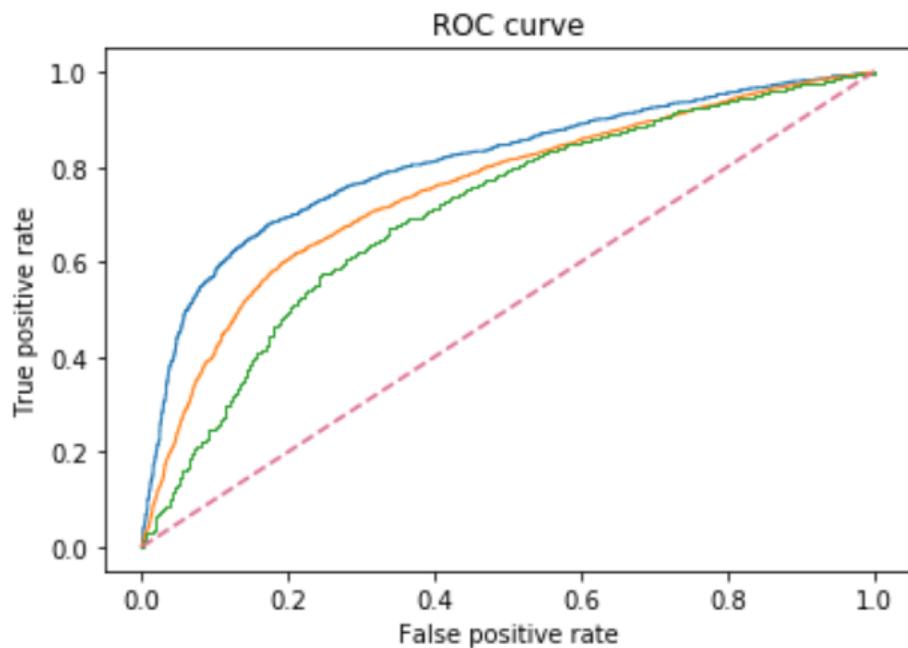
Precision Score: 0.719626168224299

Recall Score: 0.9053497942386831

F1 Score: 0.8018745118458734

Cross Validation Scores: [0.68011811 0.69045276 0.71850394 0.68011811 0.70029528]

COMPARISON BY AGE



AUROC Score

- AGE 40-50 : 0.81
- AGE 50-60 : 0.75
- AGE ABOVE 60 : 0.71

Accuracy Score

- AGE 40-50 : 0.78
- AGE 50-60 : 0.70
- AGE ABOVE 60 : 0.70