ABBA ½

PREDICTION OF HEART DISEASE ON TWO DATASETS USING SVM, NAIVES BAYES, LOGISTIC REGRESSION, DECISION TREE, RANDOM FOREST, KNN, KMEAN. ARTIFICIAL NEURAL NETWORK WITH OPTIMIZERS - SGD, ADAM, FTRL, ADAGARD & EXPLAINABLE AI (BLACKBOX)

●●●

- Binnu Ponugoti
- Avivartta Krishna

Statistical Analysis of Datasets

	Age	Sex	Chest pain type	BP	Cholesterol	FBS over 120	EKG results	Max HR	Exercise angina	ST depression	Slope of ST	Number of vessels fluro	Thallium	Target
0	70	1	4	130	322	0	2	109	0	2.4	2	3	3	Presence
1	67	0	3	115	564	0	2	160	0	1.6	2	0	7	Absence
2	57	1	2	124	261	0	0	141	0	0.3	1	0	7	Presence
3	64	1	4	128	263	0	0	105	1	0.2	2	1	7	Absence
4	74	0	2	120	269	0	2	121	1	0.2	1	1	3	Absence

Dataset

	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	target
0	52	1	0	125	212	0	1	168	0	1.0	2	2	3	0
1	53	1	0	140	203	1	0	155	1	3.1	0	0	3	0
2	70	1	0	145	174	0	1	125	1	2.6	0	0	3	0
3	61	1	0	148	203	0	1	161	0	0.0	2	1	3	0
4	62	0	0	138	294	1	1	106	0	1.9	1	3	2	0

Index of Both Datasets

Small dataset

Index(['Age', 'Sex', 'Chest pain type', 'BP', 'Cholesterol', 'FBS over 120',
'EKG results', 'Max HR', 'Exercise angina', 'ST depression',
'Slope of ST', 'Number of vessels fluro', 'Thallium', 'Target'],
dtype='object')

Large dataset

Index(['age', 'sex', 'cp', 'trestbps', 'chol', 'fbs', 'restecg', 'thalach',
'exang', 'oldpeak', 'slope', 'ca', 'thal', 'target'],
dtype='object')

DATASET INFORMATION

```
RangeIndex: 270 entries, 0 to 269
Data columns (total 14 columns):
 #   Column                  Non-Null Count  Dtype
---  ------                  --------------  -----
 0   Age                     270 non-null    int64
 1   Sex                     270 non-null    int64
 2   Chest pain type         270 non-null    int64
 3   BP                      270 non-null    int64
 4   Cholesterol             270 non-null    int64
 5   FBS over 120            270 non-null    int64
 6   EKG results             270 non-null    int64
 7   Max HR                  270 non-null    int64
 8   Exercise angina         270 non-null    int64
 9   ST depression           270 non-null    float64
 10  Slope of ST             270 non-null    int64
 11  Number of vessels fluro 270 non-null    int64
 12  Thallium                270 non-null    int64
 13  Target                  270 non-null    object
dtypes: float64(1), int64(12), object(1)
memory usage: 29.7+ KB
```

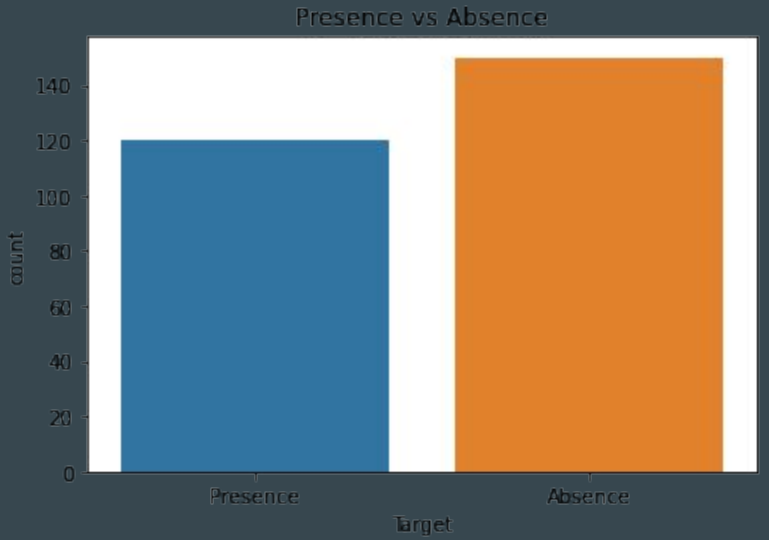
```
RangeIndex: 1025 entries, 0 to 1024
Data columns (total 14 columns):
 #   Column    Non-Null Count  Dtype
---  ------    --------------  -----
 0   age       1025 non-null   int64
 1   sex       1025 non-null   int64
 2   cp        1025 non-null   int64
 3   trestbps  1025 non-null   int64
 4   chol      1025 non-null   int64
 5   fbs       1025 non-null   int64
 6   restecg   1025 non-null   int64
 7   thalach   1025 non-null   int64
 8   exang     1025 non-null   int64
 9   oldpeak   1025 non-null   float64
 10  slope     1025 non-null   int64
 11  ca        1025 non-null   int64
 12  thal      1025 non-null   int64
 13  target    1025 non-null   int64
dtypes: float64(1), int64(13)
memory usage: 112.2 KB
```

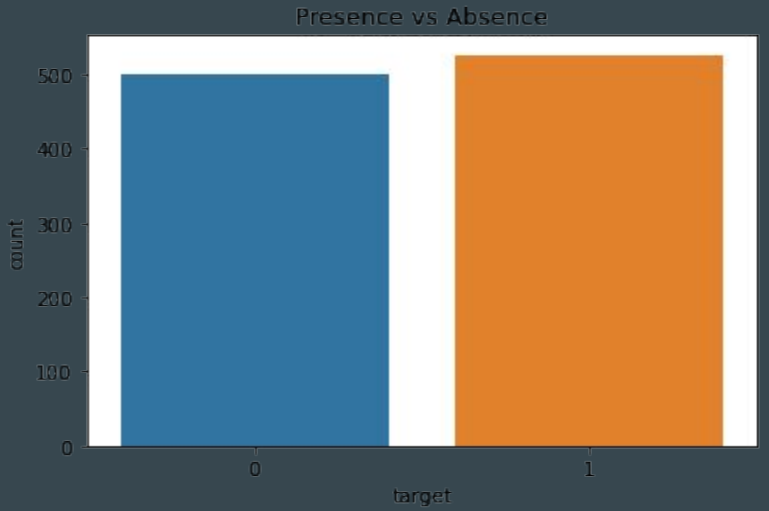
	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	target
count	1025.000000	1025.000000	1025.000000	1025.000000	1025.00000	1025.000000	1025.000000	1025.000000	1025.000000	1025.000000	1025.000000	1025.000000	1025.000000	1025.000000
mean	54.434146	0.695610	0.942439	131.611707	246.00000	0.149268	0.529756	149.114146	0.336585	1.071512	1.385366	0.754146	2.323902	0.513171
std	9.072290	0.460373	1.029641	17.516718	51.59251	0.356527	0.527878	23.005724	0.472772	1.175053	0.617755	1.030798	0.620660	0.500070
min	29.000000	0.000000	0.000000	94.000000	126.00000	0.000000	0.000000	71.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
25%	48.000000	0.000000	0.000000	120.000000	211.00000	0.000000	0.000000	132.000000	0.000000	0.000000	1.000000	0.000000	2.000000	0.000000
50%	56.000000	1.000000	1.000000	130.000000	240.00000	0.000000	1.000000	152.000000	0.000000	0.800000	1.000000	0.000000	2.000000	1.000000
75%	61.000000	1.000000	2.000000	140.000000	275.00000	0.000000	1.000000	166.000000	1.000000	1.800000	2.000000	1.000000	3.000000	1.000000
max	77.000000	1.000000	3.000000	200.000000	564.00000	1.000000	2.000000	202.000000	1.000000	6.200000	2.000000	4.000000	3.000000	1.000000

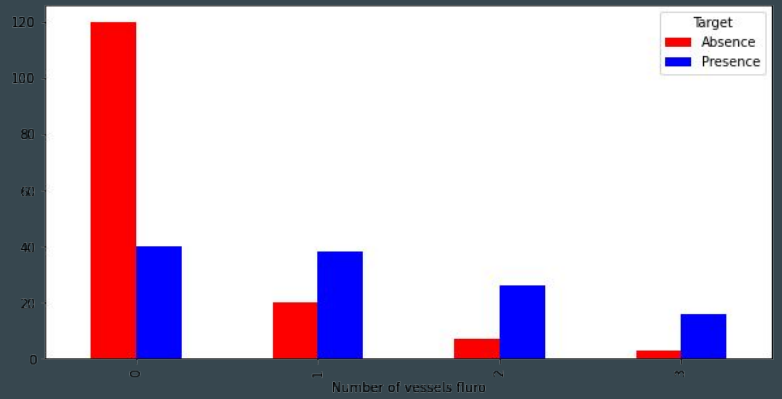
DATASET DESCRIPTION

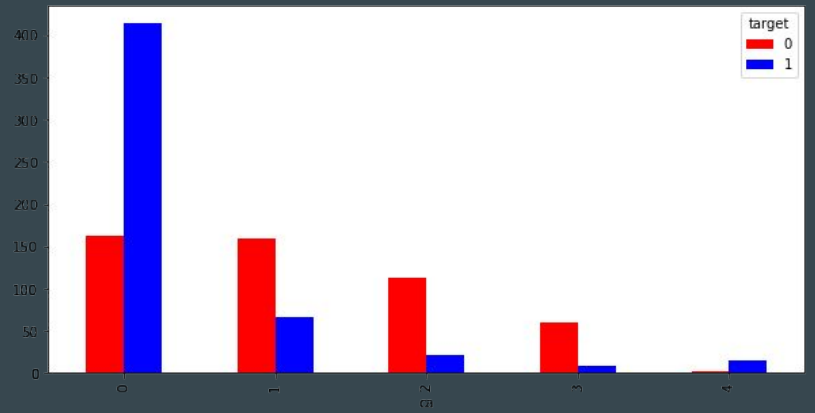
	Age	Sex	Chest pain type	BP	Cholesterol	FBS over 120	EKG results	Max HR	Exercise angina	ST depression	Slope of ST	Number of vessels fluro	Thallium
count	270.000000	270.000000	270.000000	270.000000	270.000000	270.000000	270.000000	270.000000	270.000000	270.00000	270.000000	270.000000	270.000000
mean	54.433333	0.677778	3.174074	131.344444	249.659259	0.148148	1.022222	149.677778	0.329630	1.05000	1.585185	0.670370	4.696296
std	9.109067	0.468195	0.950090	17.861608	51.686237	0.355906	0.997891	23.165717	0.470952	1.14521	0.614390	0.943896	1.940659
min	29.000000	0.000000	1.000000	94.000000	126.000000	0.000000	0.000000	71.000000	0.000000	0.00000	1.000000	0.000000	3.000000
25%	48.000000	0.000000	3.000000	120.000000	213.000000	0.000000	0.000000	133.000000	0.000000	0.00000	1.000000	0.000000	3.000000
50%	55.000000	1.000000	3.000000	130.000000	245.000000	0.000000	2.000000	153.500000	0.000000	0.80000	2.000000	0.000000	3.000000
75%	61.000000	1.000000	4.000000	140.000000	280.000000	0.000000	2.000000	166.000000	1.000000	1.60000	2.000000	1.000000	7.000000
max	77.000000	1.000000	4.000000	200.000000	564.000000	1.000000	2.000000	202.000000	1.000000	6.20000	3.000000	3.000000	7.000000

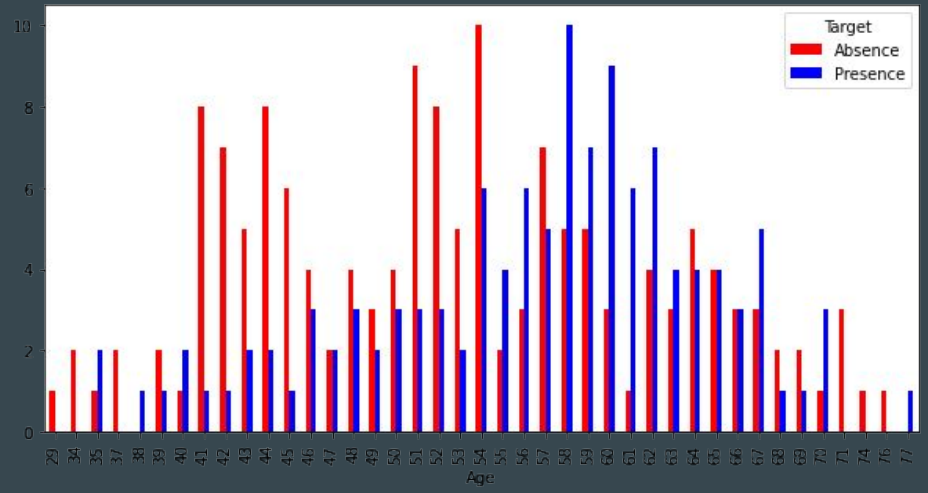
PRESENCE VS ABSENCE - HEART DISEASE

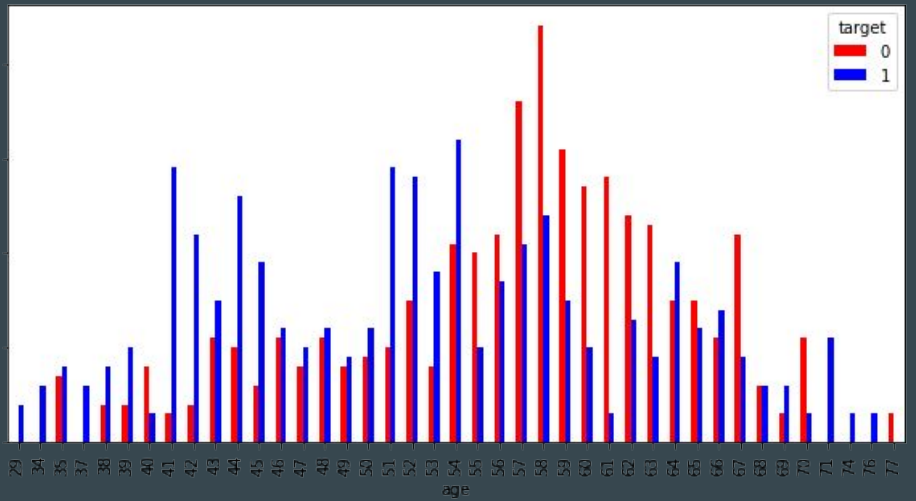


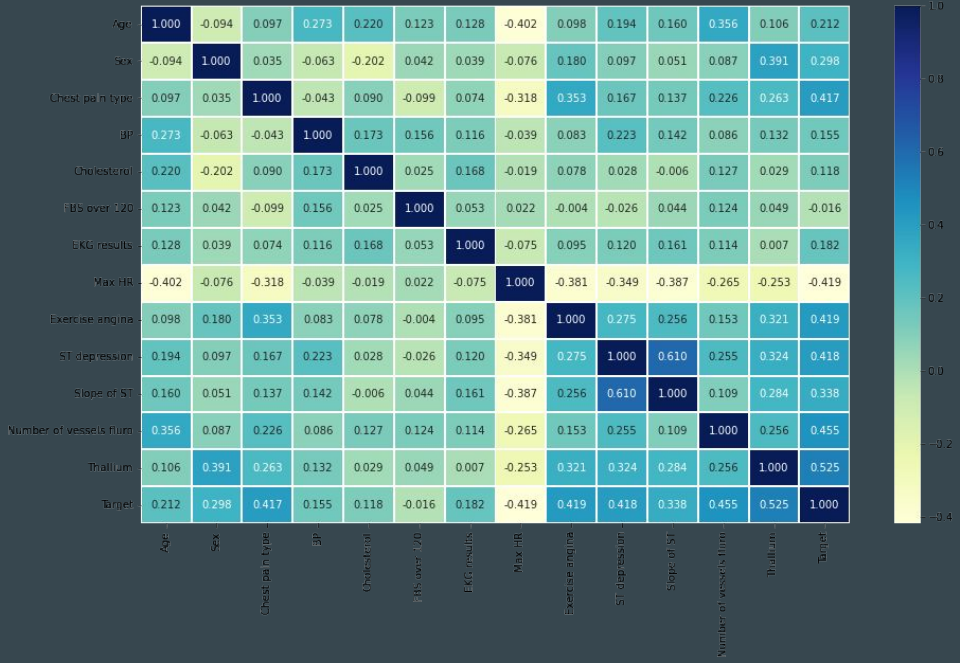


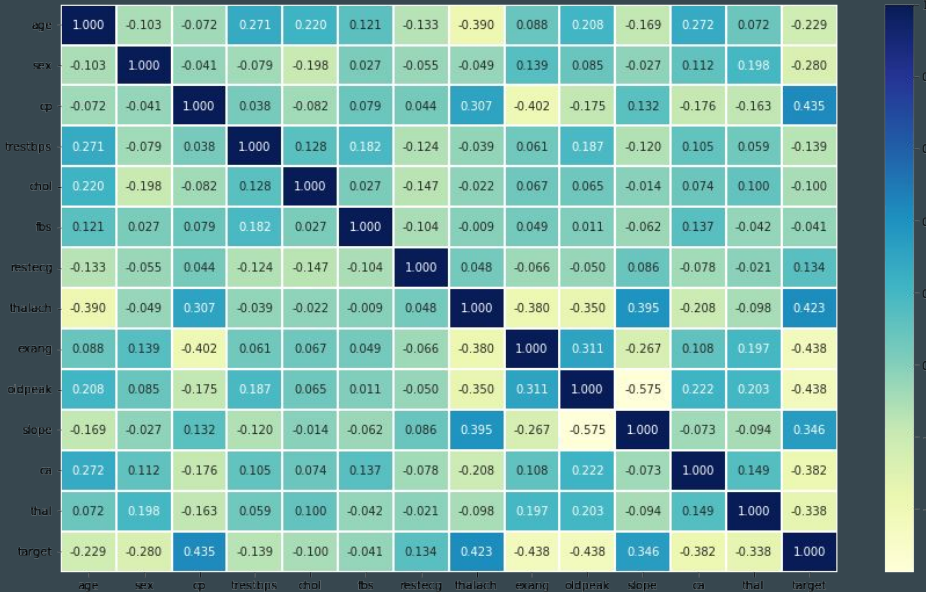





Target
Absence
Presence
Age


target
0
1
age


Age 1.000 -0.094 0.097 0.273 0.220 0.123 0.128 -0.402 0.098 0.194 0.160 0.356 0.106 0.212
Sex -0.094 1.000 0.035 -0.063 -0.202 0.042 0.039 -0.076 0.180 0.097 0.051 0.087 0.391 0.298
Chest pain type 0.097 0.035 1.000 -0.043 0.090 -0.099 0.074 -0.318 0.353 0.167 0.137 0.226 0.263 0.417
BP 0.273 -0.063 -0.043 1.000 0.173 0.156 0.116 -0.039 0.083 0.223 0.142 0.086 0.132 0.155
Cholesterol 0.220 -0.202 0.090 0.173 1.000 0.025 0.168 -0.019 0.078 0.028 -0.006 0.127 0.029 0.118
FBS over 120 0.123 0.042 -0.099 0.156 0.025 1.000 0.053 0.022 -0.004 -0.026 0.044 0.124 0.049 -0.016
EKG results 0.128 0.039 0.074 0.116 0.168 0.053 1.000 -0.075 0.095 0.120 0.161 0.114 0.007 0.182
Max HR -0.402 -0.076 -0.318 -0.039 -0.019 0.022 -0.075 1.000 -0.381 -0.349 -0.387 -0.265 -0.253 -0.419
Exercise angina 0.098 0.180 0.353 0.083 0.078 -0.004 0.095 -0.381 1.000 0.275 0.256 0.153 0.321 0.419
ST depression 0.194 0.097 0.167 0.223 0.028 -0.026 0.120 -0.349 0.275 1.000 0.610 0.255 0.324 0.418
Slope of ST 0.160 0.051 0.137 0.142 -0.006 0.044 0.161 -0.387 0.256 0.610 1.000 0.109 0.284 0.338
Number of vessels fluro 0.356 0.087 0.226 0.086 0.127 0.124 0.114 -0.265 0.153 0.255 0.109 1.000 0.256 0.455
Thallium 0.106 0.391 0.263 0.132 0.029 0.049 0.007 -0.253 0.321 0.324 0.284 0.256 1.000 0.525
Target 0.212 0.298 0.417 0.155 0.118 -0.016 0.182 -0.419 0.419 0.418 0.338 0.455 0.525 1.000
Age
Sex
Chest pain type
BP
Cholesterol
FBS over 120
EKG results
Max HR
Exercise angina
ST depression
Slope of ST
Number of vessels fluro
Thallium
Target
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4


age 1.000 -0.103 -0.072 0.271 0.220 0.121 -0.133 -0.390 0.088 0.208 -0.169 0.272 0.072 -0.229
sex -0.103 1.000 -0.041 -0.079 -0.198 0.027 -0.055 -0.049 0.139 0.085 -0.027 0.112 0.198 -0.280
cp -0.072 -0.041 1.000 0.038 -0.082 0.079 0.044 0.307 -0.402 -0.175 0.132 -0.176 -0.163 0.435
trestbps 0.271 -0.079 0.038 1.000 0.128 0.182 -0.124 -0.039 0.061 0.187 -0.120 0.105 0.059 -0.139
chol 0.220 -0.198 -0.082 0.128 1.000 0.027 -0.147 -0.022 0.067 0.065 -0.014 0.074 0.100 -0.100
fbs 0.121 0.027 0.079 0.182 0.027 1.000 -0.104 -0.009 0.049 0.011 -0.062 0.137 -0.042 -0.041
restecg -0.133 -0.055 0.044 -0.124 -0.147 -0.104 1.000 0.048 -0.066 -0.050 0.086 -0.078 -0.021 0.134
thalach -0.390 -0.049 0.307 -0.039 -0.022 -0.009 0.048 1.000 -0.380 -0.350 0.395 -0.208 -0.098 0.423
exang 0.088 0.139 -0.402 0.061 0.067 0.049 -0.066 -0.380 1.000 0.311 -0.267 0.108 0.197 -0.438
oldpeak 0.208 0.085 -0.175 0.187 0.065 0.011 -0.050 -0.350 0.311 1.000 -0.575 0.222 0.203 -0.438
slope -0.169 -0.027 0.132 -0.120 -0.014 -0.062 0.086 0.395 -0.267 -0.575 1.000 -0.073 -0.094 0.346
ca 0.272 0.112 -0.176 0.105 0.074 0.137 -0.078 -0.208 0.108 0.222 -0.073 1.000 0.149 -0.382
thal 0.072 0.198 -0.163 0.059 0.100 -0.042 -0.021 -0.098 0.197 0.203 -0.094 0.149 1.000 -0.338
target -0.229 -0.280 0.435 -0.139 -0.100 -0.041 0.134 0.423 -0.438 -0.438 0.346 -0.382 -0.338 1.000
age
sex
cp
trestbps
chol
fbs
restecg
thalach
exang
oldpeak
slope
ca
thal
target
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4

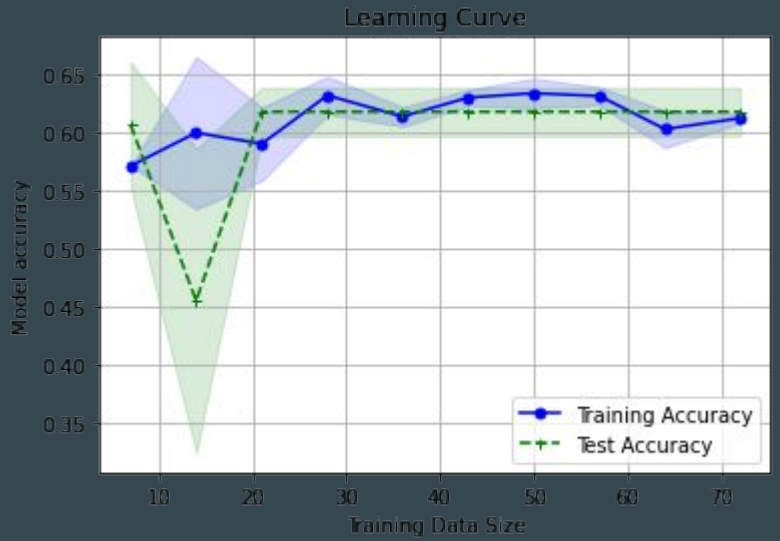
SMALL DATASET - RESULTS

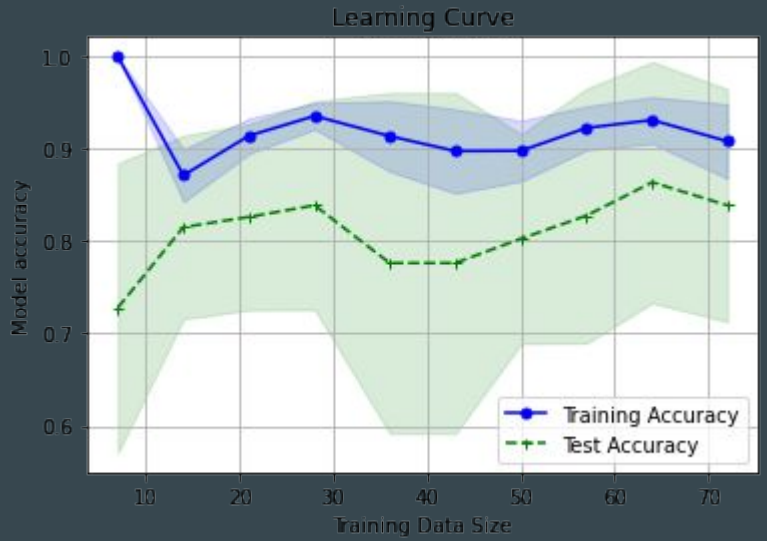
SVM

	precision	recall	f1-score	support
0	0.82	0.69	0.75	59
1	0.42	0.59	0.49	22
accuracy			0.67	81
macro avg	0.62	0.64	0.62	81
weighted avg	0.71	0.67	0.68	81

Naive Bayes

	precision	recall	f1-score	support
0	0.88	0.85	0.86	52
1	0.74	0.79	0.77	29
accuracy			0.83	81
macro avg	0.81	0.82	0.81	81
weighted avg	0.83	0.83	0.83	81



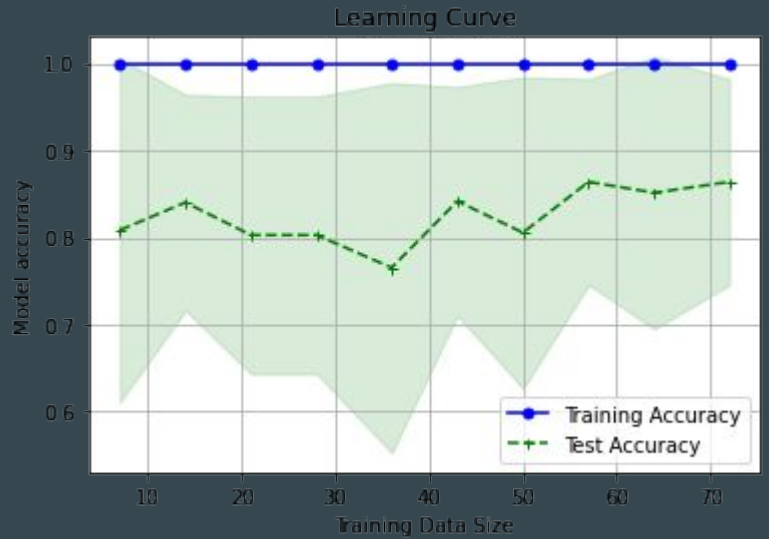


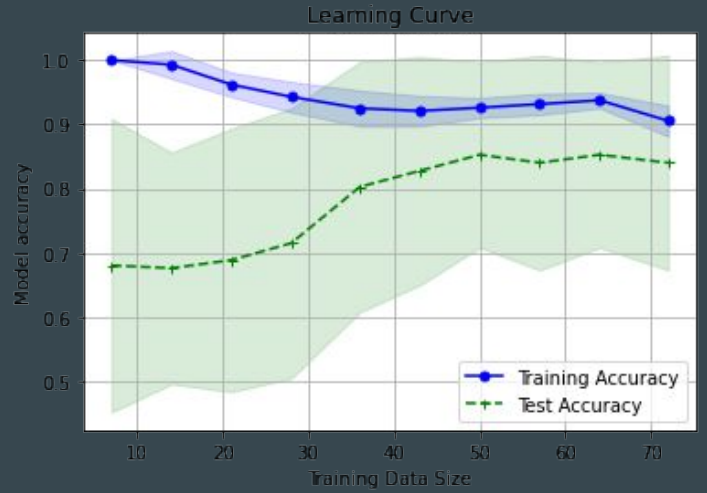
Decision Tree

	precision	recall	f1-score	support
0	0.96	0.96	0.96	145
1	0.96	0.96	0.96	163
accuracy			0.96	308
macro avg	0.96	0.96	0.96	308
weighted avg	0.96	0.96	0.96	308

Logistic Regression

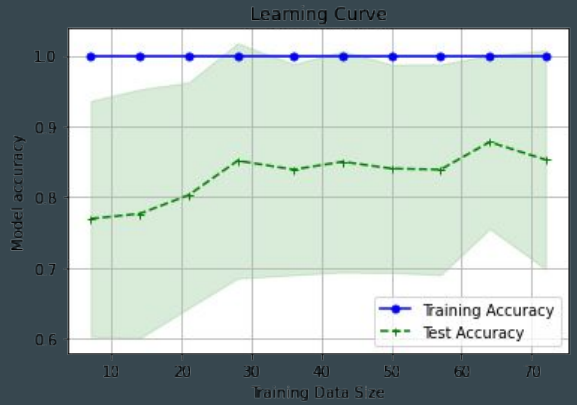
	precision	recall	f1-score	support
0	0.88	0.85	0.86	52
1	0.74	0.79	0.77	29
accuracy			0.83	81
macro avg	0.81	0.82	0.81	81
weighted avg	0.83	0.83	0.83	81





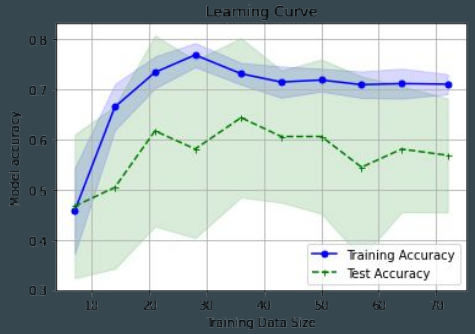
Random Forest

	precision	recall	f1-score	support
0	0.92	0.82	0.87	56
1	0.68	0.84	0.75	25
accuracy			0.83	81
macro avg	0.80	0.83	0.81	81
weighted avg	0.85	0.83	0.83	81



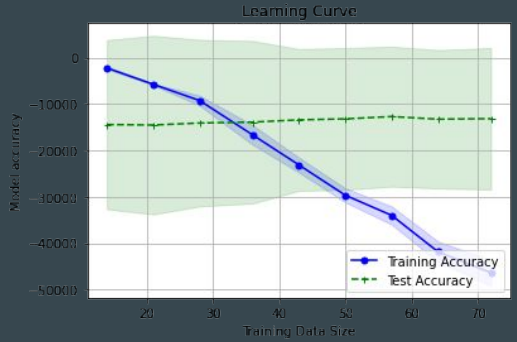
KNN

	precision	recall	f1-score	support
0	0.82	0.73	0.77	56
1	0.52	0.64	0.57	25
accuracy			0.70	81
macro avg	0.67	0.69	0.67	81
weighted avg	0.73	0.70	0.71	81



KMeans

	precision	recall	f1-score	support
0	0.18	0.75	0.29	12
1	0.13	0.36	0.19	11
2	0.00	0.00	0.00	20
3	0.00	0.00	0.00	16
4	0.00	0.00	0.00	2
5	0.00	0.00	0.00	3
6	0.00	0.00	0.00	11
7	0.00	0.00	0.00	6
accuracy			0.16	81
macro avg	0.04	0.14	0.06	81
weighted avg	0.04	0.16	0.07	81



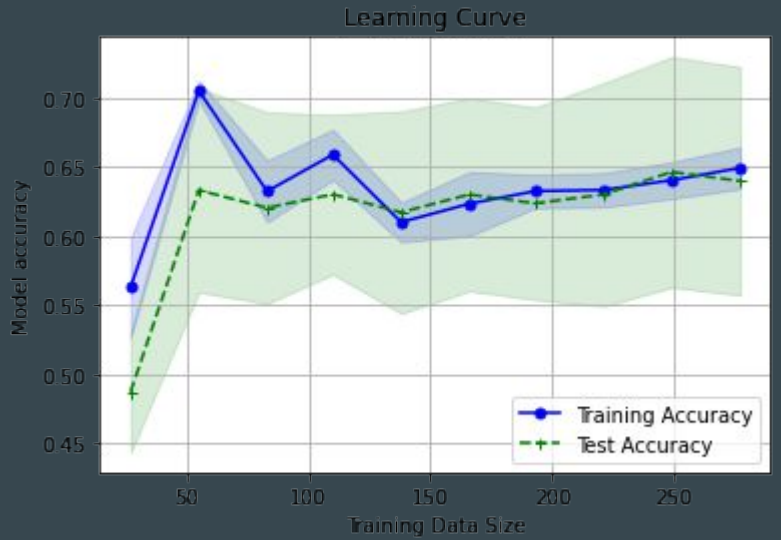
BIG DATASET - RESULTS

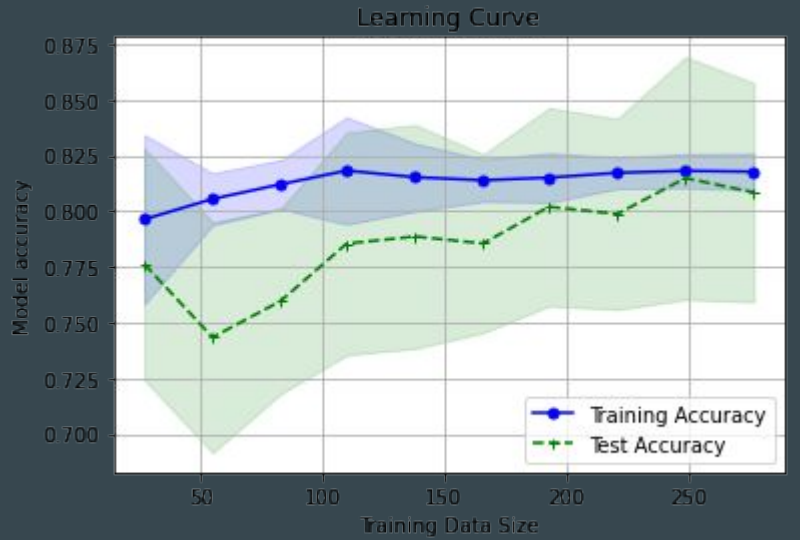
SVM

	precision	recall	f1-score	support
0	0.66	0.65	0.65	146
1	0.69	0.69	0.69	162
accuracy			0.67	308
macro avg	0.67	0.67	0.67	308
weighted avg	0.67	0.67	0.67	308

Naive Bayes

	precision	recall	f1-score	support
0	0.76	0.78	0.77	141
1	0.81	0.79	0.80	167
accuracy			0.79	308
macro avg	0.78	0.79	0.78	308
weighted avg	0.79	0.79	0.79	308



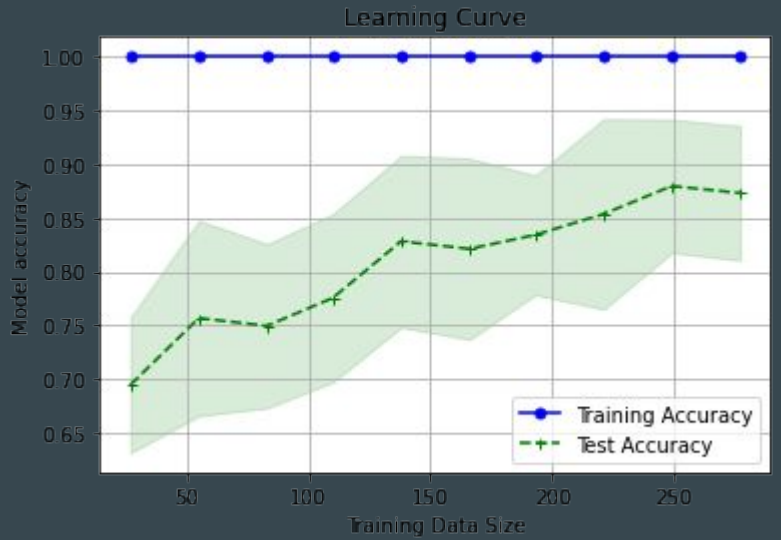


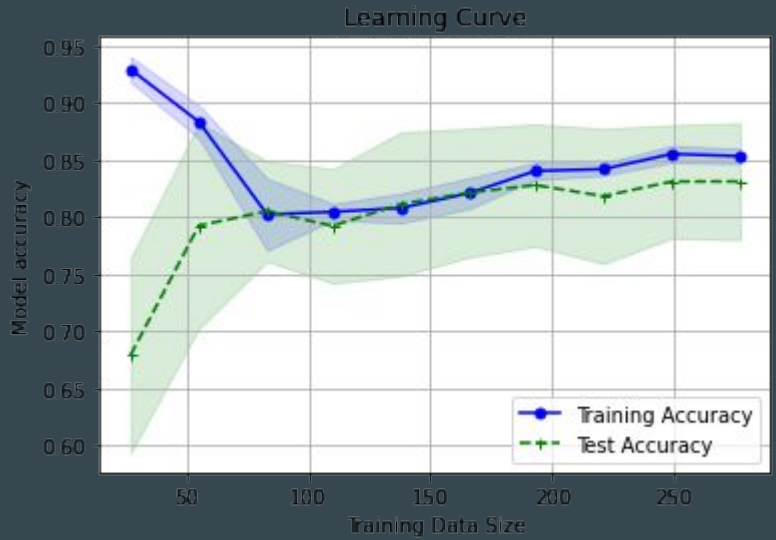
Decision Tree

	precision	recall	f1-score	support
0	0.96	0.96	0.96	145
1	0.96	0.96	0.96	163
accuracy			0.96	308
macro avg	0.96	0.96	0.96	308
weighted avg	0.96	0.96	0.96	308

Logistic Regression

	precision	recall	f1-score	support
0	0.75	0.84	0.80	129
1	0.88	0.80	0.84	179
accuracy			0.82	308
macro avg	0.81	0.82	0.82	308
weighted avg	0.82	0.82	0.82	308





Random Forest

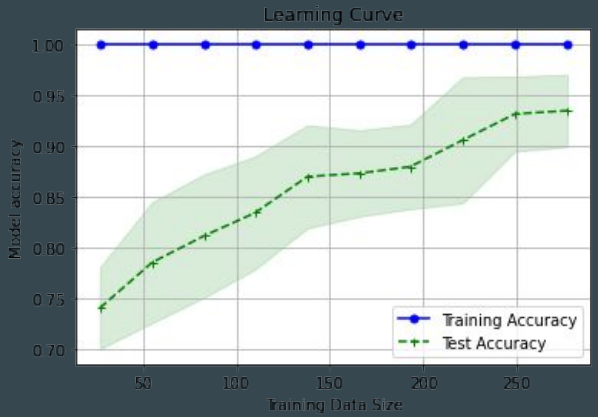
	precision	recall	f1-score	support
0	0.98	0.98	0.98	145
1	0.98	0.98	0.98	163
accuracy			0.98	308
macro avg	0.98	0.98	0.98	308
weighted avg	0.98	0.98	0.98	308

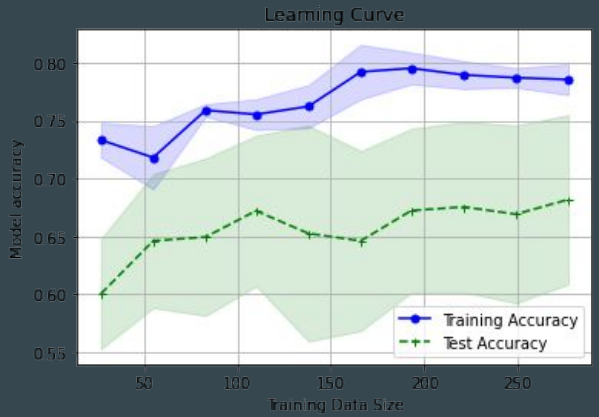
KNN

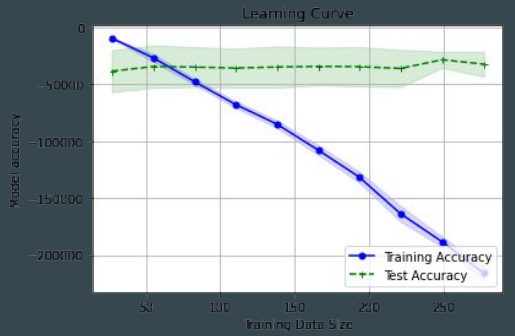
	precision	recall	f1-score	support
0	0.73	0.68	0.71	155
1	0.70	0.75	0.72	153
accuracy			0.71	308
macro avg	0.72	0.71	0.71	308
weighted avg	0.72	0.71	0.71	308

KMeans

	precision	recall	f1-score	support
0	0.14	0.51	0.23	41
1	0.06	0.28	0.09	32
2	0.00	0.00	0.00	89
3	0.00	0.00	0.00	18
4	0.00	0.00	0.00	49
5	0.00	0.00	0.00	10
6	0.00	0.00	0.00	48
7	0.00	0.00	0.00	21
accuracy			0.10	308
macro avg	0.03	0.10	0.04	308
weighted avg	0.03	0.10	0.04	308

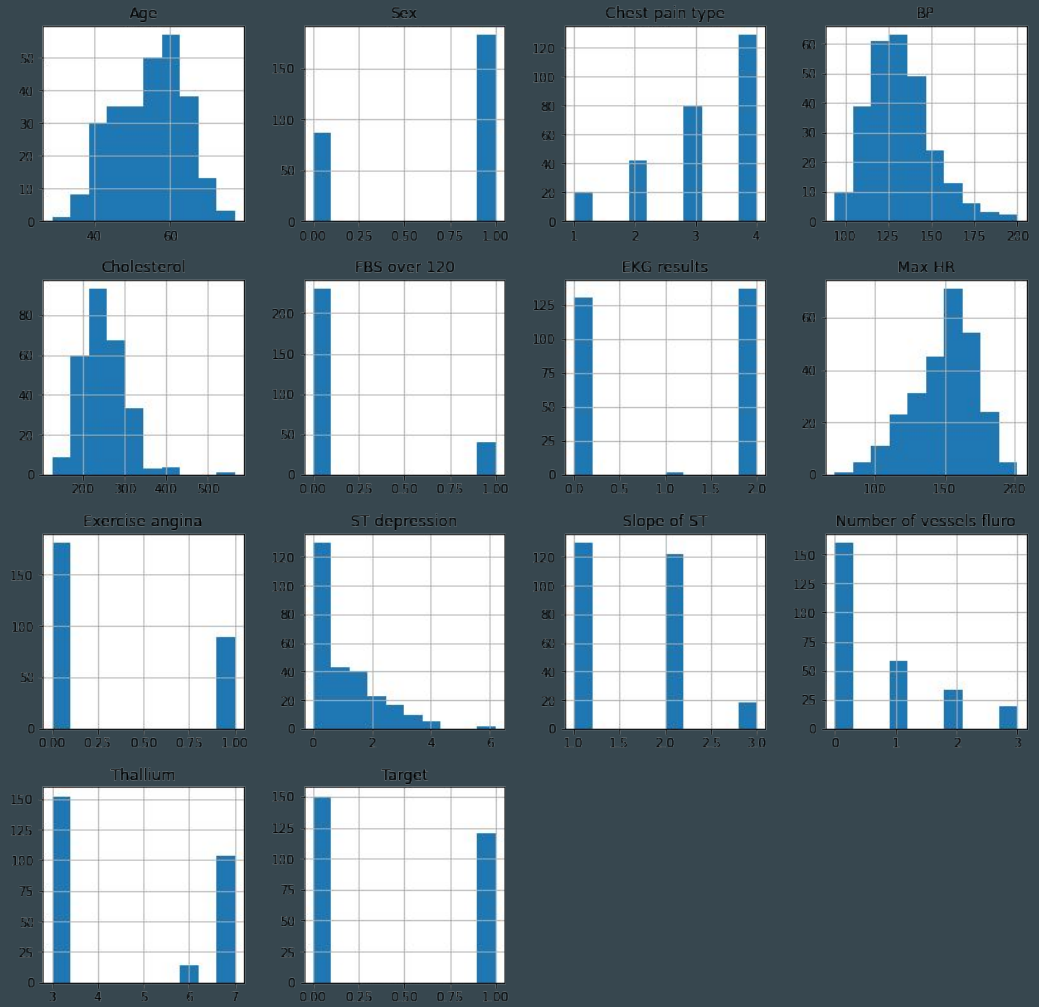


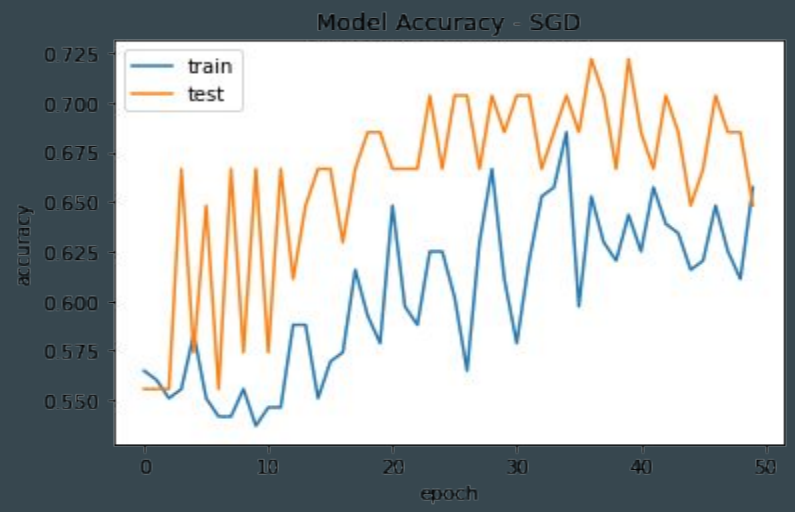


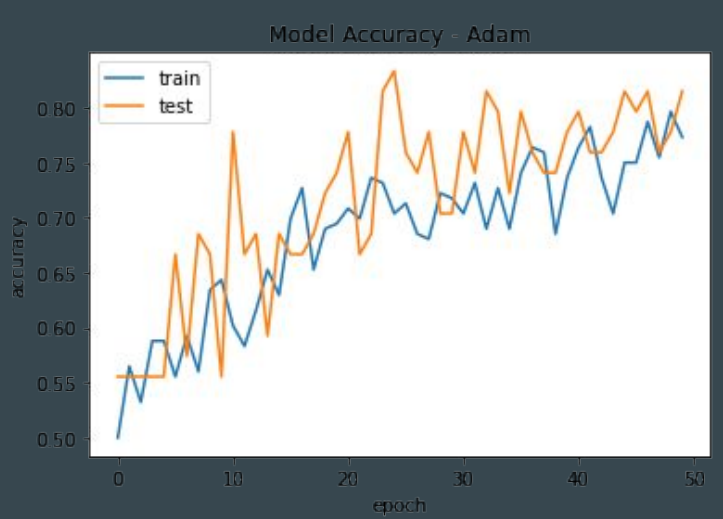


NEURAL NETWORK - RESULTS

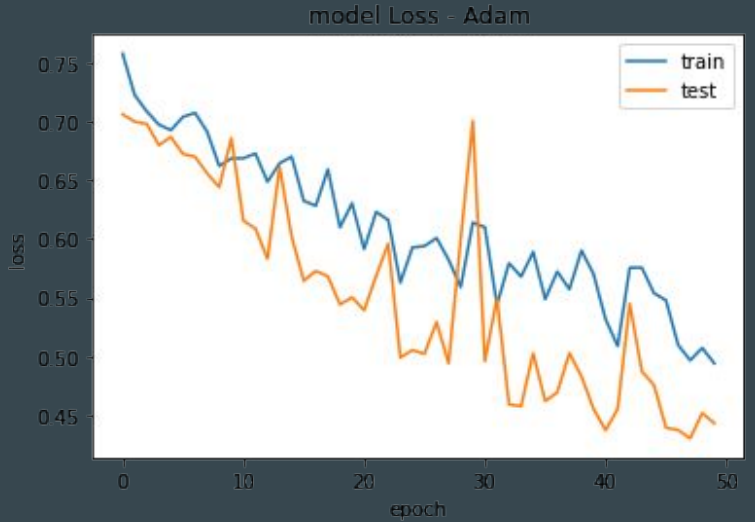
SMALL DATASET

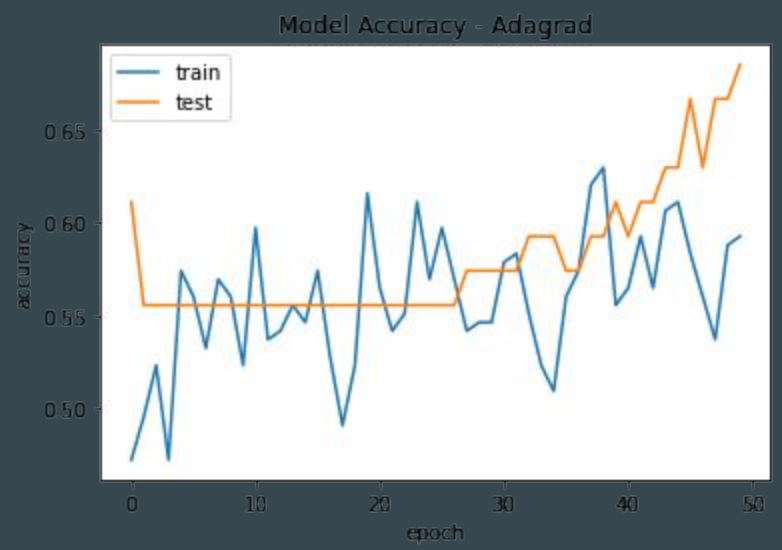

Age
Sex
Chest pain type
BP
Cholesterol
FBS over 120
EKG results
Max HR
Exercise angina
ST depression
Slope of ST
Number of vessels fluro
Thallium
Target

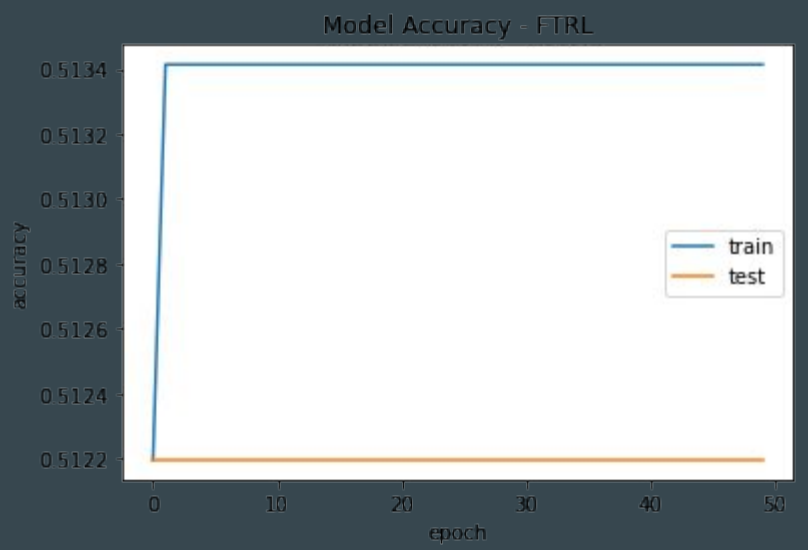

Model Accuracy - SGD
train
test
accuracy
epoch

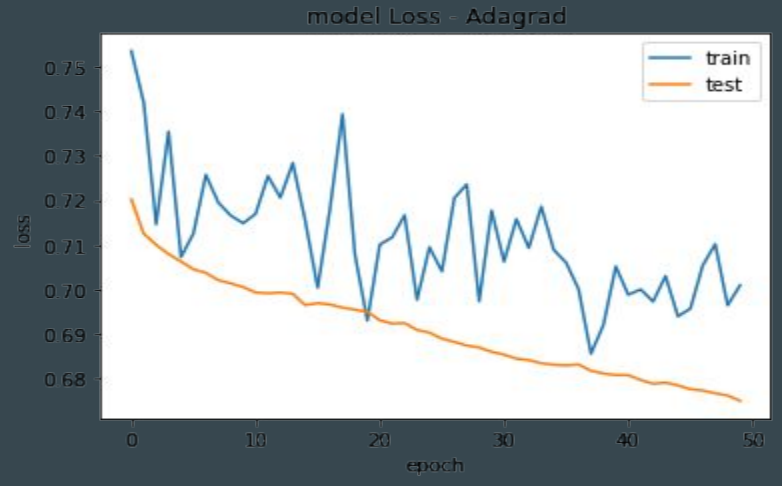

Model Accuracy - Adam
train
test
accuracy
epoch

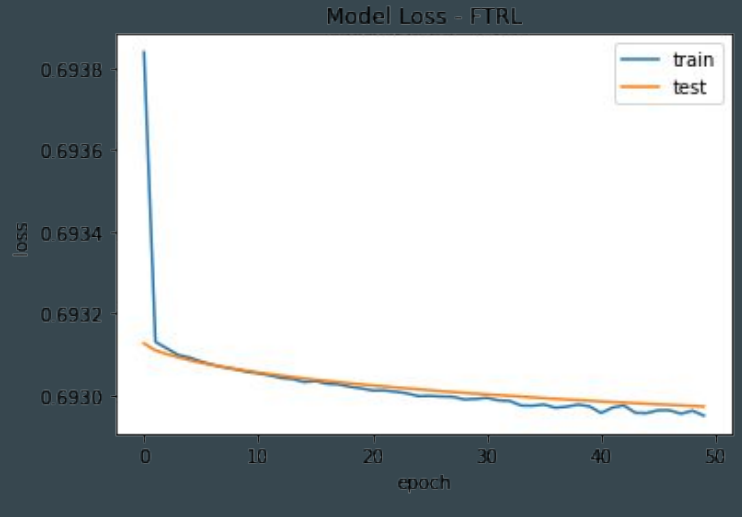

model Loss - SGD
train
test
loss
epoch


model Loss - Adam
train
test
loss
epoch

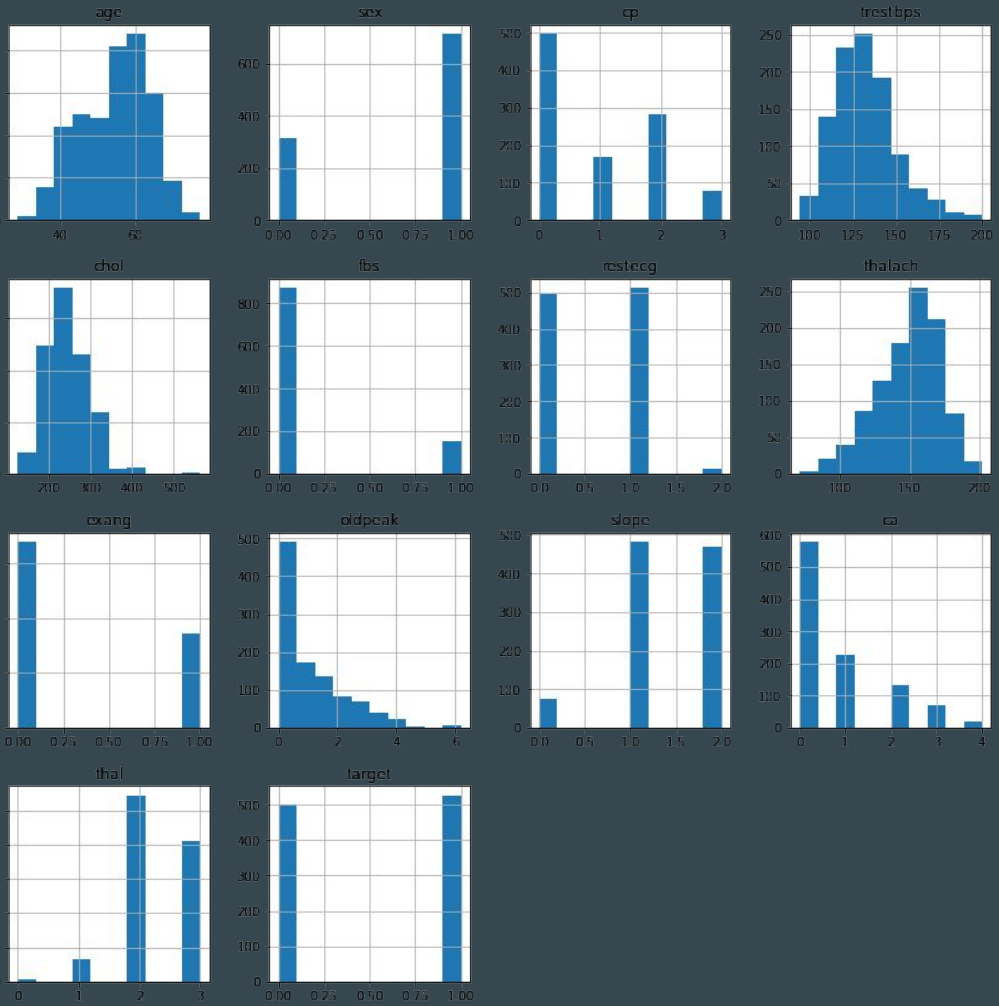

Model Accuracy - Adagrad
train
test
accuracy
epoch

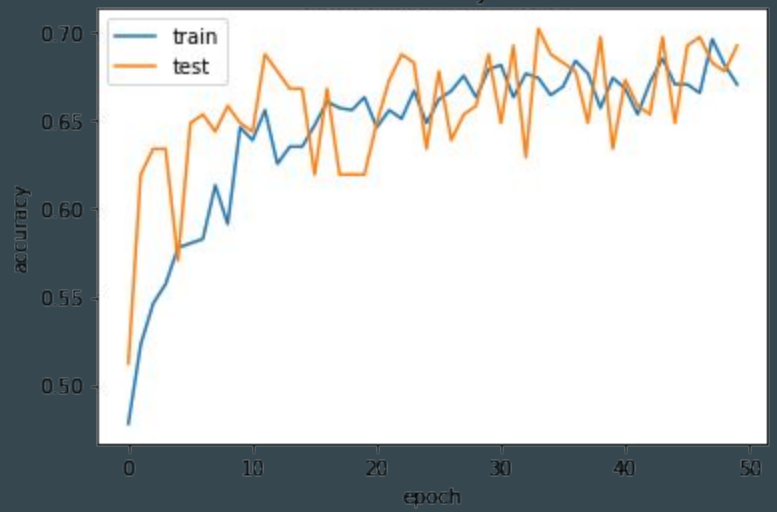

Model Accuracy - FTRL
train
test
accuracy
epoch

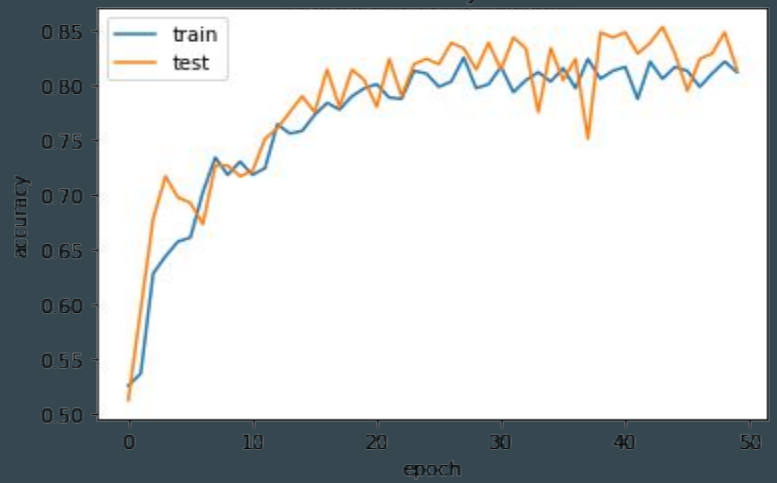

model Loss - Adagrad
train
test
loss
epoch

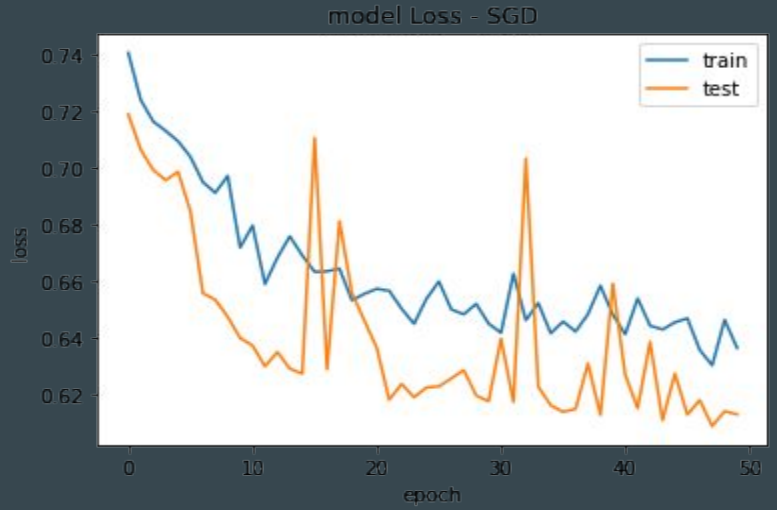

Model Loss - FTRL
train
test
loss
epoch

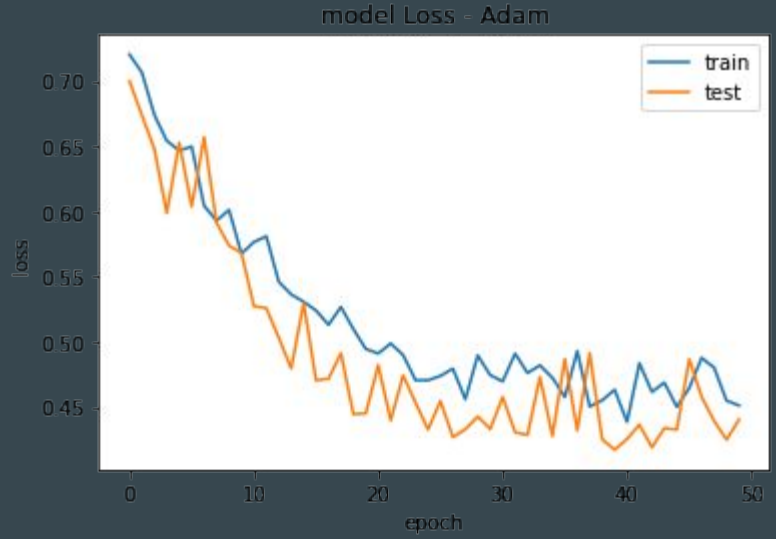
BIG DATASET

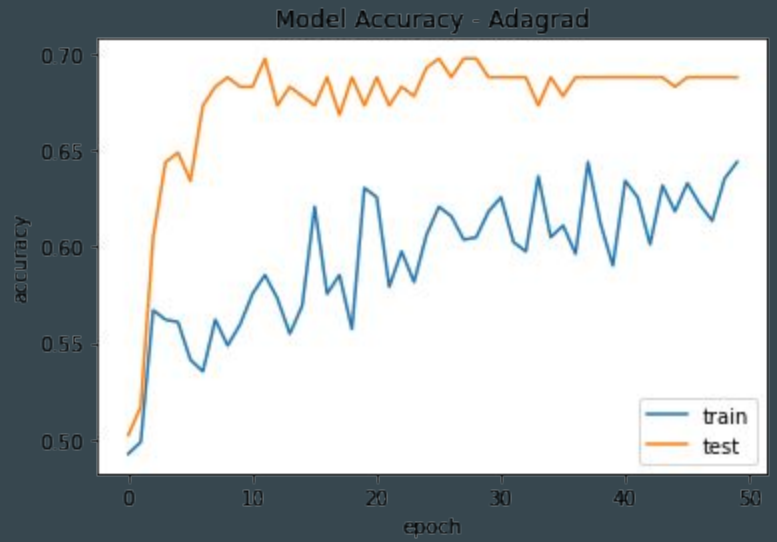

age
sex
cp
trestbps
chol
fbs
restecg
thalach
exang
oldpeak
slope
ca
thal
target

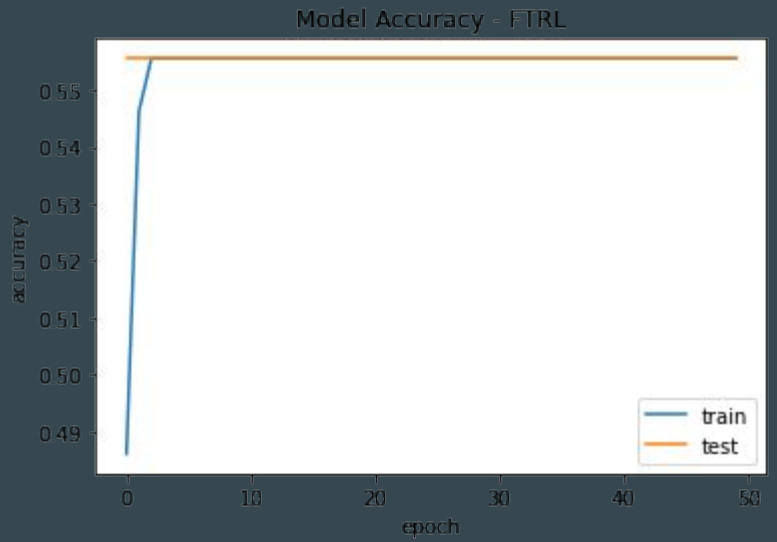

Model Accuracy - SGD
train
test
accuracy
epoch


Model Accuracy - Adam
train
test
accuracy
epoch

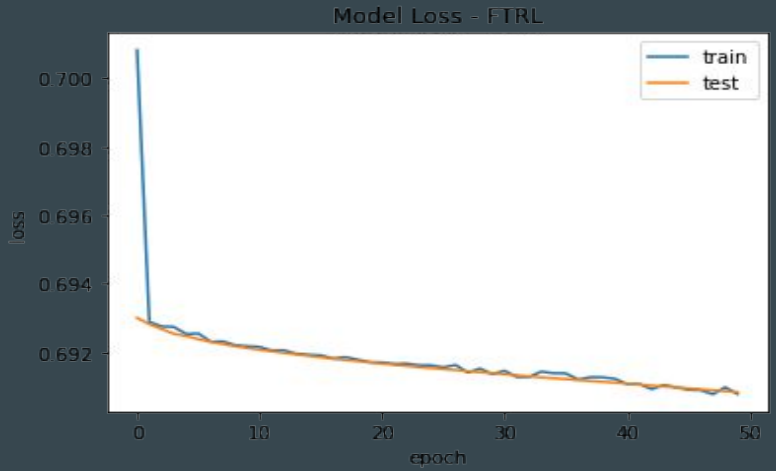

model Loss - SGD
train
test
loss
epoch


model Loss - Adam
train
test
loss
epoch

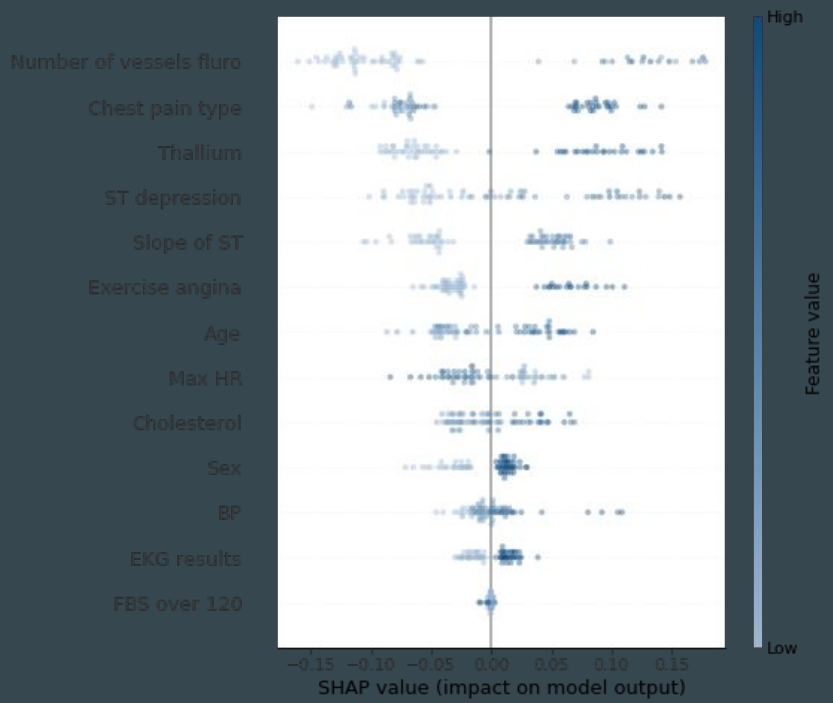

Model Accuracy - Adagrad
accuracy
epoch
train
test

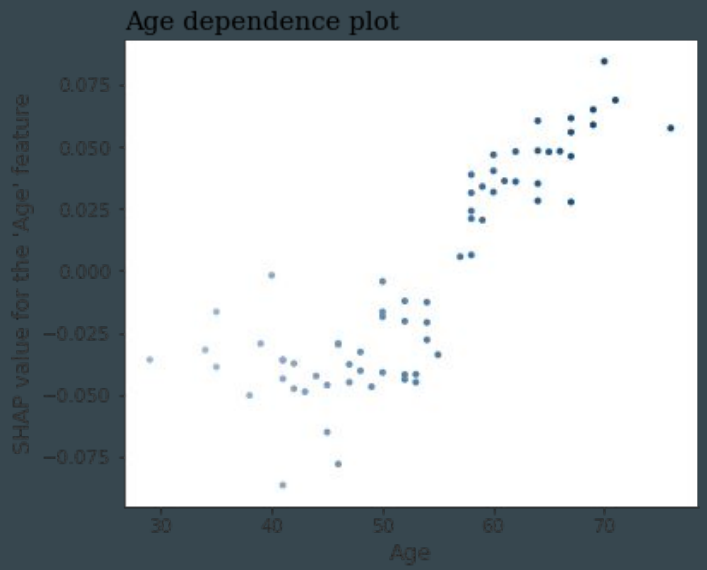

Model Accuracy - FTRL
accuracy
epoch
train
test


model Loss - Adagrad
loss
epoch
train
test


Model Loss - FTRL
loss
epoch
train
test

EXPLAINABLE AI





Weight	Feature
0.1231 ± 0.1586	Max HR
0.1219 ± 0.1882	Thallium
0.1212 ± 0.1293	ST depression
0.1089 ± 0.1693	Chest pain type
0.1089 ± 0.1444	Number of vessels fluro
0.0935 ± 0.1100	Age
0.0837 ± 0.0710	Cholesterol
0.0779 ± 0.0783	BP
0.0490 ± 0.0930	Slope of ST
0.0483 ± 0.1083	Exercise angina
0.0333 ± 0.0545	Sex
0.0209 ± 0.0357	EKG results
0.0095 ± 0.0242	FBS over 120

SHAPELY ADDITIVE EXPLANATIONS - SHAP