Heart Disease Prediction and Analysis

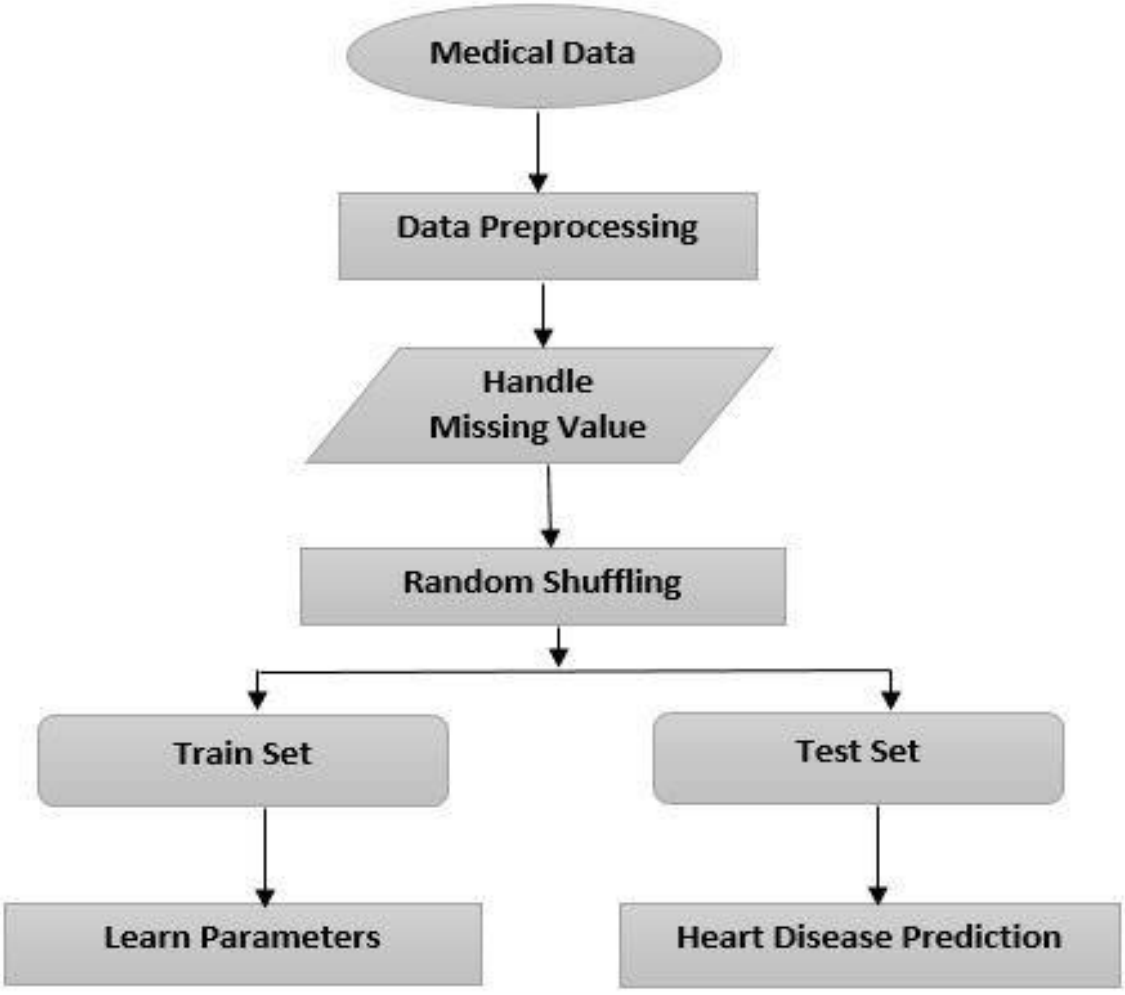
Introduction :

- **Heart attack is one of the main sources of death around the world and it is imperative to predict the disease at a premature phase.The computer aided systems help the doctor as a tool for predicting and diagnosing heart disease.**
- **Clinicians and patients need reliable information about an individual's risk of developing Heart Disease. Ideally, they would have entirely accurate data and would be able to use a perfect model to estimate risk. Such a model would be able to categorize people with heart disease and others.**

Problem Statement:

- **The objective of this project is to widespread about Heart related cardiovascular disease and to brief about existing decision support systems for the prediction of heart disease supported by data mining and machine learning techniques.**

Block Diagram :

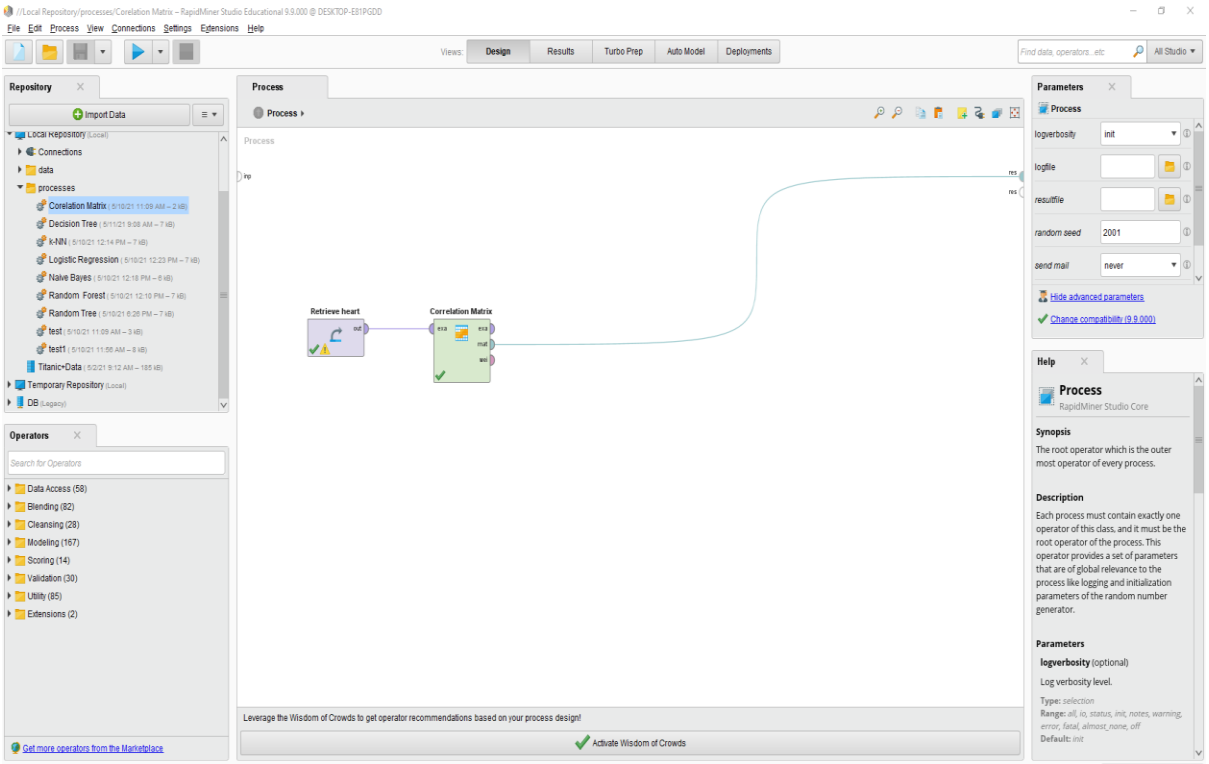


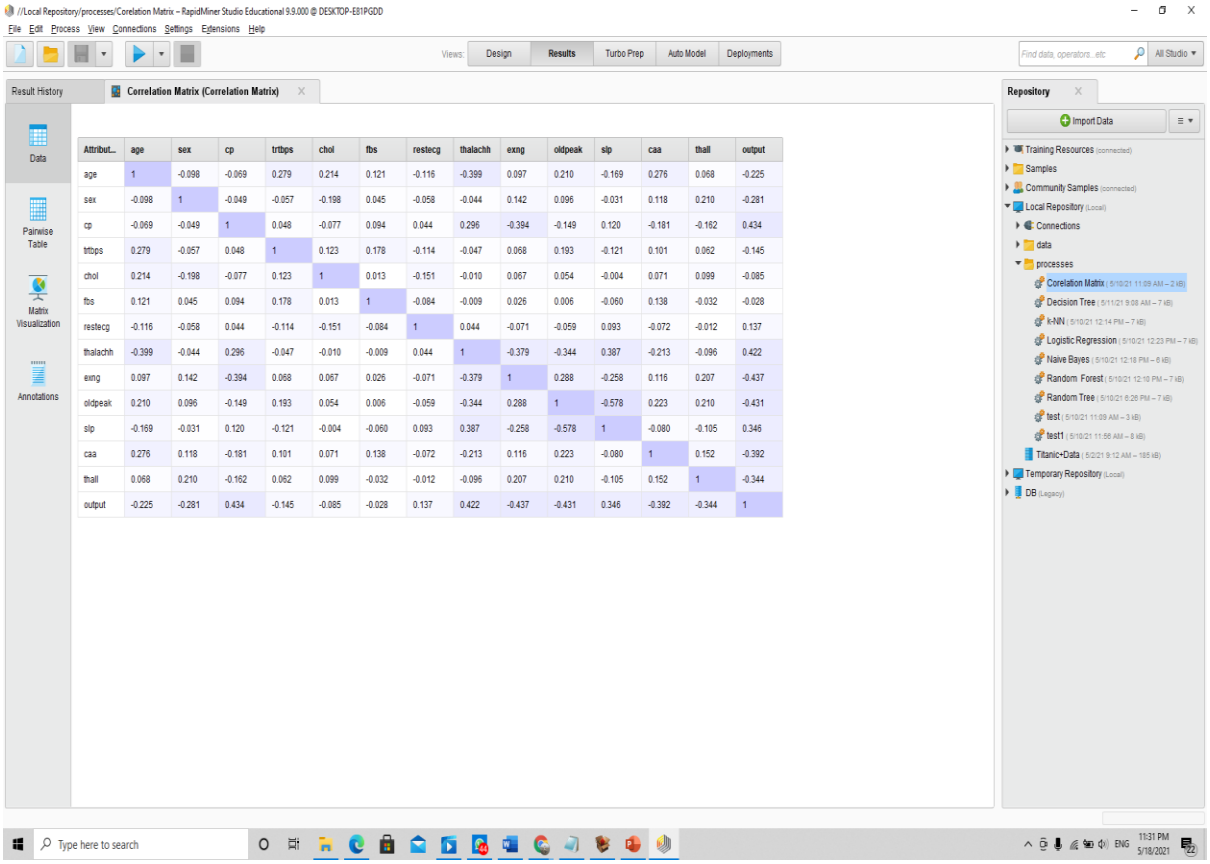
Dataset :

age	sex	cp	trtbps	chol	fbs	restecg	thalachh	exng	oldpeak	slp	caa	thall	output
63	1	3	145	233	1	0	150	0	2.3	0	0	1	1
37	1	2	130	250	0	1	187	0	3.5	0	0	2	1
41	0	1	130	204	0	0	172	0	1.4	2	0	2	1
56	1	1	120	236	0	1	178	0	0.8	2	0	2	1
57	0	0	120	354	0	1	163	1	0.6	2	0	2	1
57	1	0	140	192	0	1	148	0	0.4	1	0	1	1
56	0	1	140	294	0	0	153	0	1.3	1	0	2	1
44	1	1	120	263	0	1	173	0	0	2	0	3	1
52	1	2	172	199	1	1	162	0	0.5	2	0	3	1
57	1	2	150	168	0	1	174	0	1.6	2	0	2	1
54	1	0	140	239	0	1	160	0	1.2	2	0	2	1
48	0	2	130	275	0	1	139	0	0.2	2	0	2	1
49	1	1	130	266	0	1	171	0	0.6	2	0	2	1
64	1	3	110	211	0	0	144	1	1.8	1	0	2	1
58	0	3	150	283	1	0	162	0	1	2	0	2	1
50	0	2	120	219	0	1	158	0	1.6	1	0	2	1
58	0	2	120	340	0	1	172	0	0	2	0	2	1
66	0	3	150	226	0	1	114	0	2.6	0	0	2	1
43	1	0	150	247	0	1	171	0	1.5	2	0	2	1
69	0	3	140	239	0	1	151	0	1.8	2	2	2	1
59	1	0	135	234	0	1	161	0	0.5	1	0	3	1
44	1	2	130	233	0	1	179	1	0.4	2	0	2	1
42	1	0	140	226	0	1	178	0	0	2	0	2	1
61	1	2	150	243	1	1	137	1	1	1	0	2	1
40	1	3	140	199	0	1	178	1	1.4	2	0	3	1
71	0	1	160	302	0	1	162	0	0.4	2	2	2	1
59	1	2	150	212	1	1	157	0	1.6	2	0	2	1
51	1	2	110	175	0	1	123	0	0.6	2	0	2	1
65	0	2	140	417	1	0	157	0	0.8	2	1	2	1
53	1	2	130	197	1	0	152	0	1.2	0	0	2	1
41	0	1	105	198	0	1	168	0	0	2	1	2	1
65	1	0	120	177	0	1	140	0	0.4	2	0	3	1
44	1	1	130	219	0	0	188	0	0	2	0	2	1
54	1	2	125	273	0	0	152	0	0.5	0	1	2	1
51	1	3	125	213	0	0	125	1	1.4	2	1	2	1
46	0	2	142	177	0	0	160	1	1.4	0	0	2	1
54	0	2	135	304	1	1	170	0	0	2	0	2	1
54	1	2	150	232	0	0	165	0	1.6	2	0	3	1
65	0	2	155	269	0	1	148	0	0.8	2	0	2	1
65	0	2	160	360	0	0	151	0	0.8	2	0	2	1
51	0	2	140	308	0	0	142	0	1.5	2	1	2	1
48	1	1	130	245	0	0	180	0	0.2	1	0	2	1
45	1	0	104	208	0	0	148	1	3	1	0	2	1
53	0	0	130	264	0	0	143	0	0.4	1	0	2	1
39	1	2	140	321	0	0	182	0	0	2	0	2	1
52	1	1	120	325	0	1	172	0	0.2	2	0	2	1
44	1	2	140	235	0	0	180	0	0	2	0	2	1
47	1	2	138	257	0	0	156	0	0	2	0	2	1
53	0	2	128	216	0	0	115	0	0	2	0	0	1
53	0	0	138	234	0	0	160	0	0	2	0	2	1
51	0	2	130	256	0	0	149	0	0.5	2	0	2	1
66	1	0	120	302	0	0	151	0	0.4	1	0	2	1
62	1	2	130	231	0	1	146	0	1.8	1	3	3	1
44	0	2	108	141	0	1	175	0	0.6	1	0	2	1
63	0	2	135	252	0	0	172	0	0	2	0	2	1
52	1	1	134	201	0	1	158	0	0.8	2	1	2	1
48	1	0	122	222	0	0	186	0	0	2	0	2	1
45	1	0	115	260	0	0	185	0	0	2	0	2	1
34	1	3	118	182	0	0	174	0	0	2	0	2	1
57	0	0	128	303	0	0	159	0	0	2	1	2	1
71	0	2	110	265	1	0	130	0	0	2	1	2	1

Rapid Miner Design :

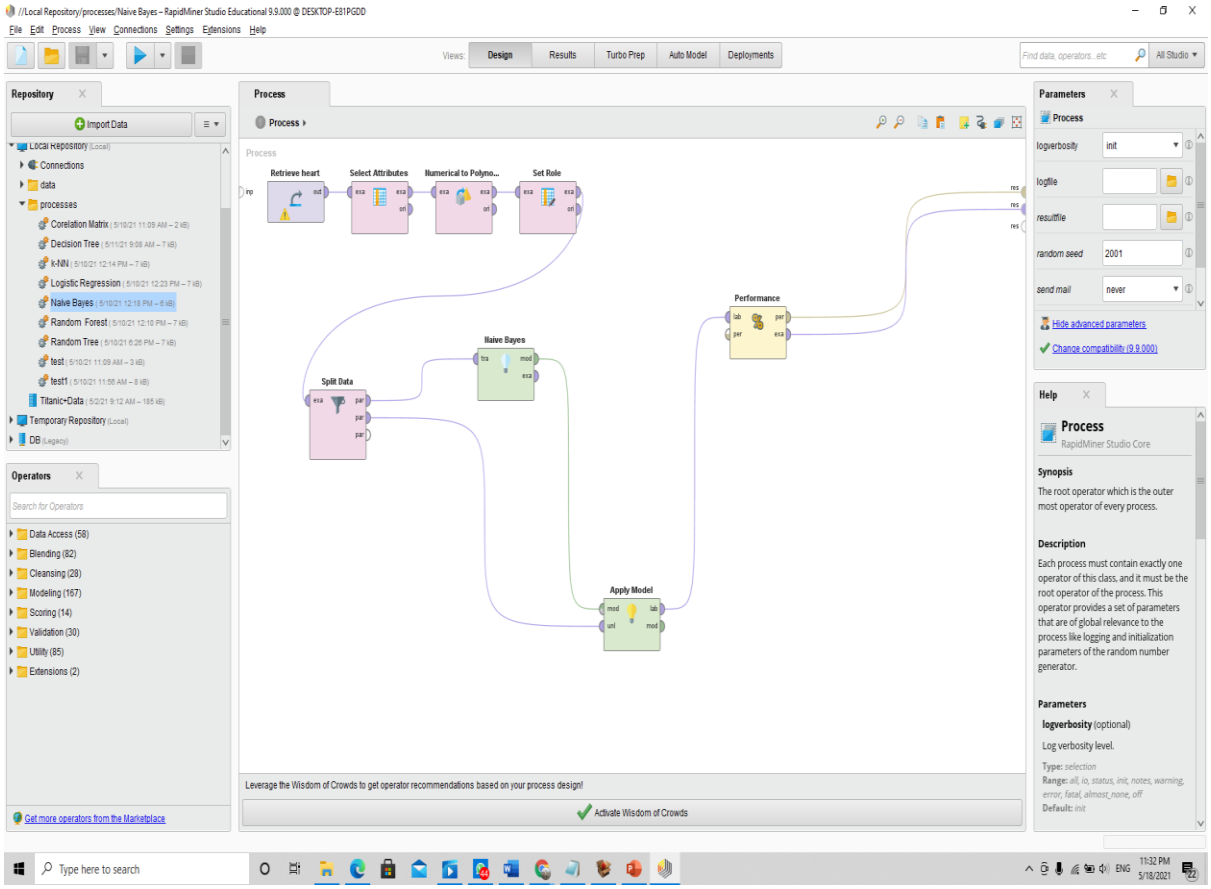
Corelation Matrix



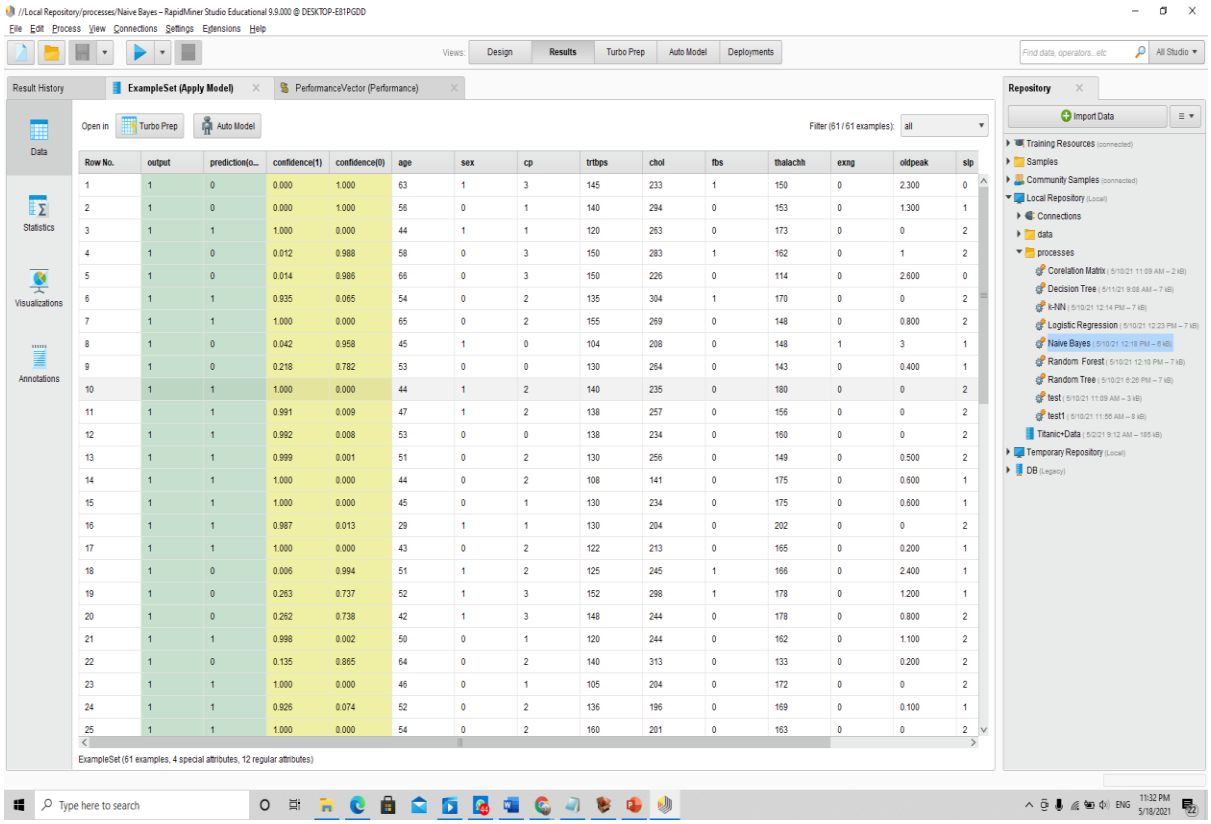


Attribut...	age	sex	cp	trtbps	chol	fbs	restecg	thalachh	exng	oldpeak	slp	caa	thall	output
age	1	-0.098	-0.069	0.279	0.214	0.121	-0.116	-0.399	0.097	0.210	-0.169	0.276	0.068	-0.225
sex	-0.098	1	-0.049	-0.057	-0.198	0.045	-0.058	-0.044	0.142	0.096	-0.031	0.118	0.210	-0.281
cp	-0.069	-0.049	1	0.048	-0.077	0.094	0.044	0.296	-0.394	-0.149	0.120	-0.181	-0.162	0.434
trtbps	0.279	-0.057	0.048	1	0.123	0.178	-0.114	-0.047	0.068	0.193	-0.121	0.101	0.062	-0.145
chol	0.214	-0.198	-0.077	0.123	1	0.013	-0.151	-0.010	0.067	0.054	-0.004	0.071	0.099	-0.085
fbs	0.121	0.045	0.094	0.178	0.013	1	-0.084	-0.009	0.026	0.006	-0.060	0.138	-0.032	-0.028
restecg	-0.116	-0.058	0.044	-0.114	-0.151	-0.084	1	0.044	-0.071	-0.059	0.093	-0.072	-0.012	0.137
thalachh	-0.399	-0.044	0.296	-0.047	-0.010	-0.009	0.044	1	-0.379	-0.344	0.387	-0.213	-0.096	0.422
exng	0.097	0.142	-0.394	0.068	0.067	0.026	-0.071	-0.379	1	0.288	-0.258	0.116	0.207	-0.437
oldpeak	0.210	0.096	-0.149	0.193	0.054	0.006	-0.059	-0.344	0.288	1	-0.578	0.223	0.210	-0.431
slp	-0.169	-0.031	0.120	-0.121	-0.004	-0.060	0.093	0.387	-0.258	-0.578	1	-0.080	-0.105	0.346
caa	0.276	0.118	-0.181	0.101	0.071	0.138	-0.072	-0.213	0.116	0.223	-0.080	1	0.152	-0.392
thall	0.068	0.210	-0.162	0.062	0.099	-0.032	-0.012	-0.096	0.207	0.210	-0.105	0.152	1	-0.344
output	-0.225	-0.281	0.434	-0.145	-0.085	-0.028	0.137	0.422	-0.437	-0.431	0.346	-0.392	-0.344	1

Naïve-Bayes Algorithm Design:

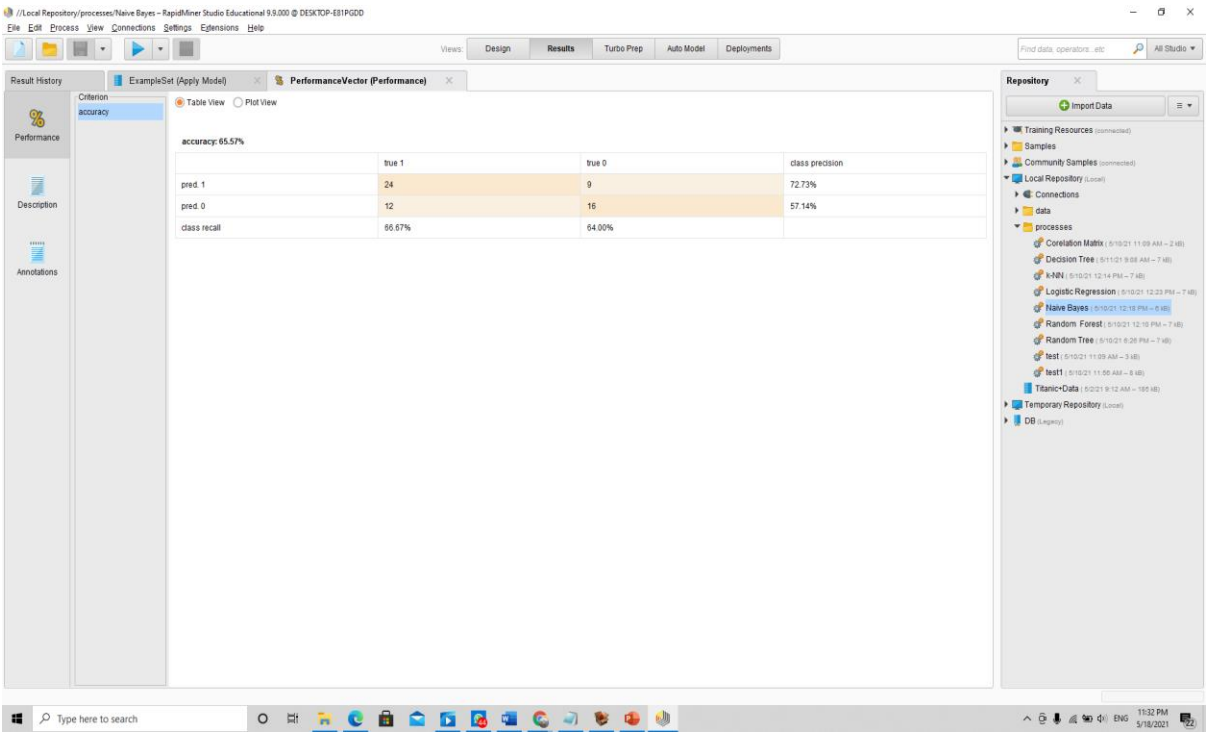


Output for Naïve Bayes

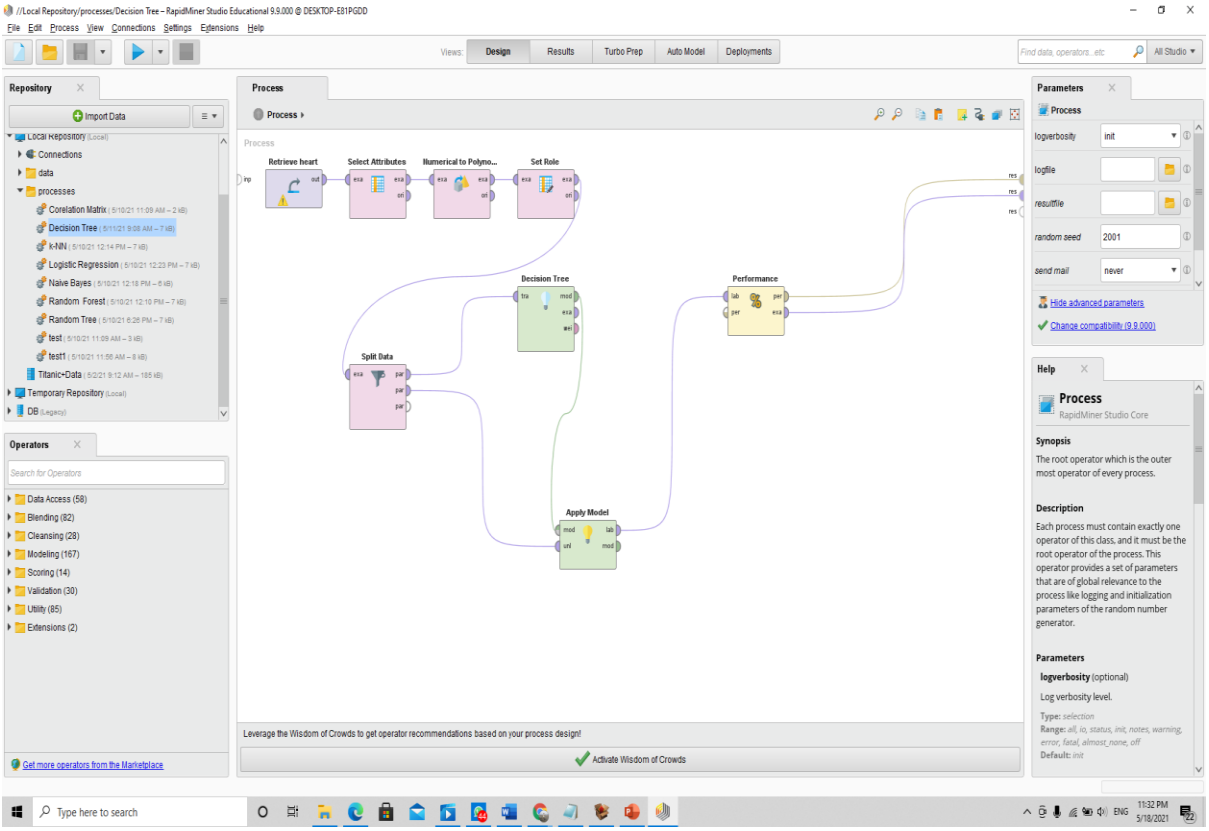


Row No.	output	prediction(o...	confidence(1)	confidence(0)	age	sex	cp	trtbps	chol	fbs	thalachh	exng	oldpeak	slp
1	1	0	0.000	1.000	63	1	3	145	233	1	150	0	2.300	0
2	1	0	0.000	1.000	56	0	1	140	294	0	153	0	1.300	1
3	1	1	1.000	0.000	44	1	1	120	263	0	173	0	0	2
4	1	0	0.012	0.988	58	0	3	150	283	1	162	0	1	2
5	1	0	0.014	0.986	66	0	3	150	226	0	114	0	2.600	0
6	1	1	0.935	0.065	54	0	2	135	304	1	170	0	0	2
7	1	1	1.000	0.000	65	0	2	155	269	0	148	0	0.800	2
8	1	0	0.042	0.958	45	1	0	104	208	0	148	1	3	1
9	1	0	0.218	0.782	53	0	0	130	264	0	143	0	0.400	1
10	1	1	1.000	0.000	44	1	2	140	235	0	180	0	0	2
11	1	1	0.991	0.009	47	1	2	138	257	0	156	0	0	2
12	1	1	0.992	0.008	53	0	0	138	234	0	160	0	0	2
13	1	1	0.999	0.001	51	0	2	130	256	0	149	0	0.500	2
14	1	1	1.000	0.000	44	0	2	108	141	0	175	0	0.600	1
15	1	1	1.000	0.000	45	0	1	130	234	0	175	0	0.600	1
16	1	1	0.987	0.013	29	1	1	130	204	0	202	0	0	2
17	1	1	1.000	0.000	43	0	2	122	213	0	165	0	0.200	1
18	1	0	0.006	0.994	51	1	2	125	245	1	166	0	2.400	1
19	1	0	0.263	0.737	52	1	3	152	298	1	178	0	1.200	1
20	1	0	0.262	0.738	42	1	3	148	244	0	178	0	0.800	2
21	1	1	0.998	0.002	50	0	1	120	244	0	162	0	1.100	2
22	1	0	0.135	0.865	64	0	2	140	313	0	133	0	0.200	2
23	1	1	1.000	0.000	46	0	1	105	204	0	172	0	0	2
24	1	1	0.926	0.074	52	0	2	136	196	0	169	0	0.100	1
25	1	1	1.000	0.000	54	0	2	160	201	0	163	0	0	2

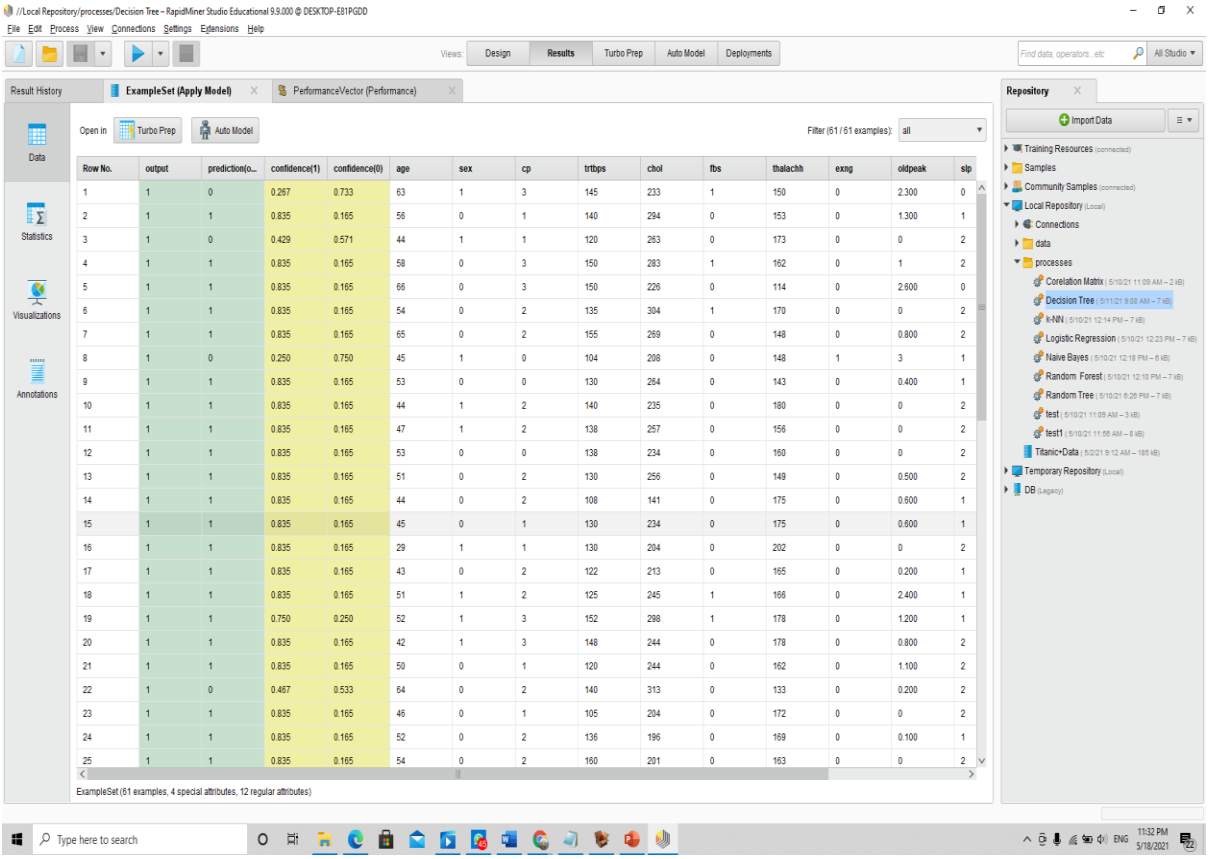
Performance for Naïve-Bayes algorithm of accuracy = 65.57%



Decision Tree Algorithm Design :

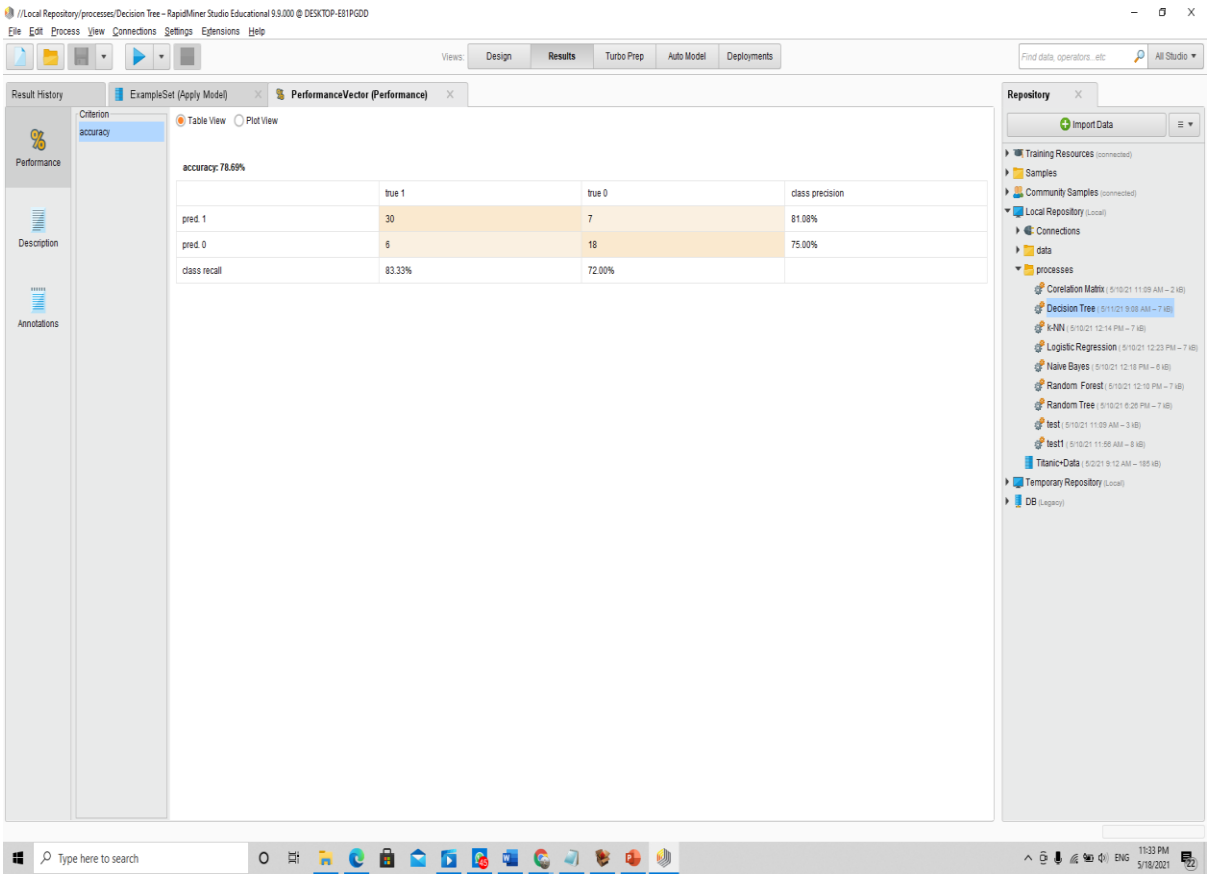


Output for Decision Tree Algorithm:



Row No.	output	prediction(o...	confidence(1)	confidence(0)	age	sex	cp	trtbps	chol	fbs	thalachh	exng	oldpeak	slp
1	1	0	0.267	0.733	63	1	3	145	233	1	150	0	2.300	0
2	1	1	0.835	0.165	56	0	1	140	294	0	153	0	1.300	1
3	1	0	0.429	0.571	44	1	1	120	263	0	173	0	0	2
4	1	1	0.835	0.165	58	0	3	150	283	1	162	0	1	2
5	1	1	0.835	0.165	66	0	3	150	226	0	114	0	2.600	0
6	1	1	0.835	0.165	54	0	2	135	304	1	170	0	0	2
7	1	1	0.835	0.165	65	0	2	155	269	0	148	0	0.800	2
8	1	0	0.250	0.750	45	1	0	104	208	0	148	1	3	1
9	1	1	0.835	0.165	53	0	0	130	264	0	143	0	0.400	1
10	1	1	0.835	0.165	44	1	2	140	235	0	180	0	0	2
11	1	1	0.835	0.165	47	1	2	138	257	0	156	0	0	2
12	1	1	0.835	0.165	53	0	0	138	234	0	160	0	0	2
13	1	1	0.835	0.165	51	0	2	130	256	0	149	0	0.500	2
14	1	1	0.835	0.165	44	0	2	108	141	0	175	0	0.600	1
15	1	1	0.835	0.165	45	0	1	130	234	0	175	0	0.600	1
16	1	1	0.835	0.165	29	1	1	130	204	0	202	0	0	2
17	1	1	0.835	0.165	43	0	2	122	213	0	165	0	0.200	1
18	1	1	0.835	0.165	51	1	2	125	245	1	166	0	2.400	1
19	1	1	0.750	0.250	52	1	3	152	298	1	178	0	1.200	1
20	1	1	0.835	0.165	42	1	3	148	244	0	178	0	0.800	2
21	1	1	0.835	0.165	50	0	1	120	244	0	162	0	1.100	2
22	1	0	0.467	0.533	64	0	2	140	313	0	133	0	0.200	2
23	1	1	0.835	0.165	46	0	1	105	204	0	172	0	0	2
24	1	1	0.835	0.165	52	0	2	136	196	0	169	0	0.100	1
25	1	1	0.835	0.165	54	0	2	160	201	0	163	0	0	2

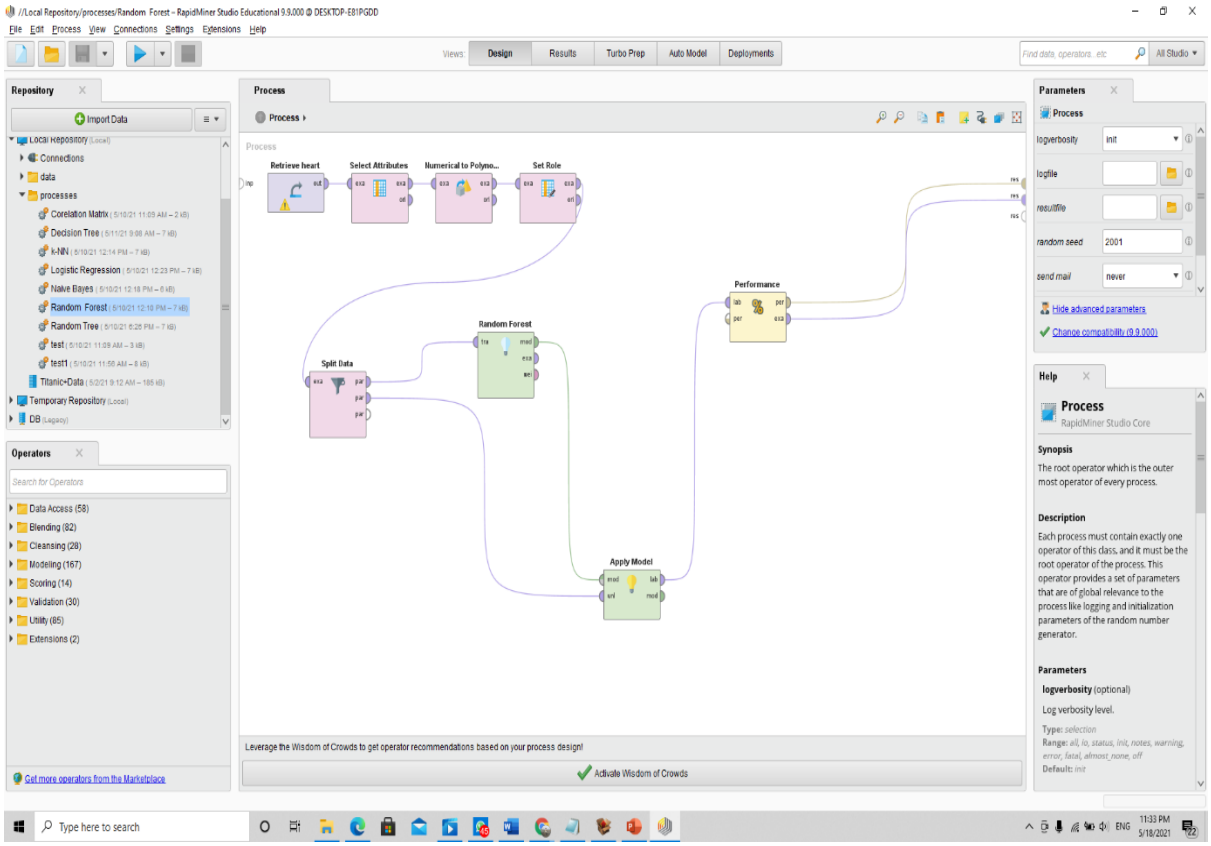
Performance for Decision Tree matrix of Accuracy = 78.69%



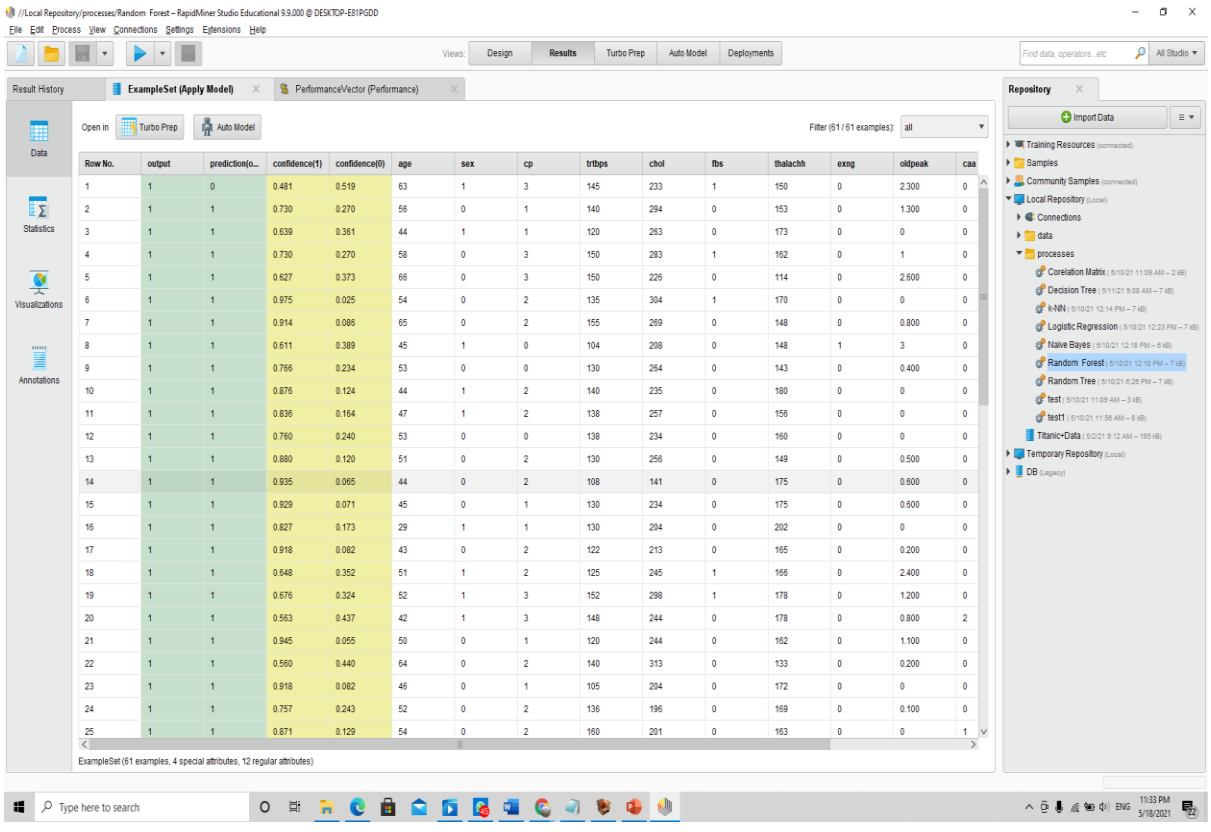
accuracy: 78.69%

	true 1	true 0	class precision
pred. 1	30	7	81.08%
pred. 0	6	18	75.00%
class recall	83.33%	72.00%	

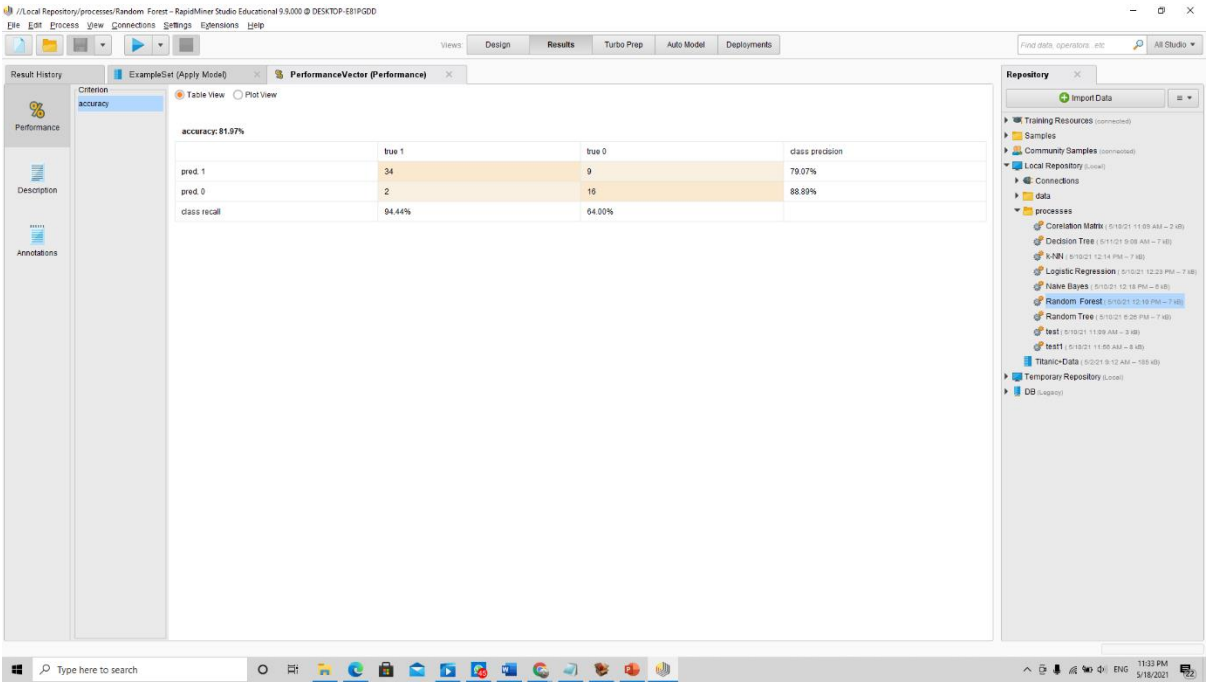
Random Forest Design Algorithm :



Output for Random Forest Algorithm



Performance for Random Forest Algorithm of accuracy 81.97%



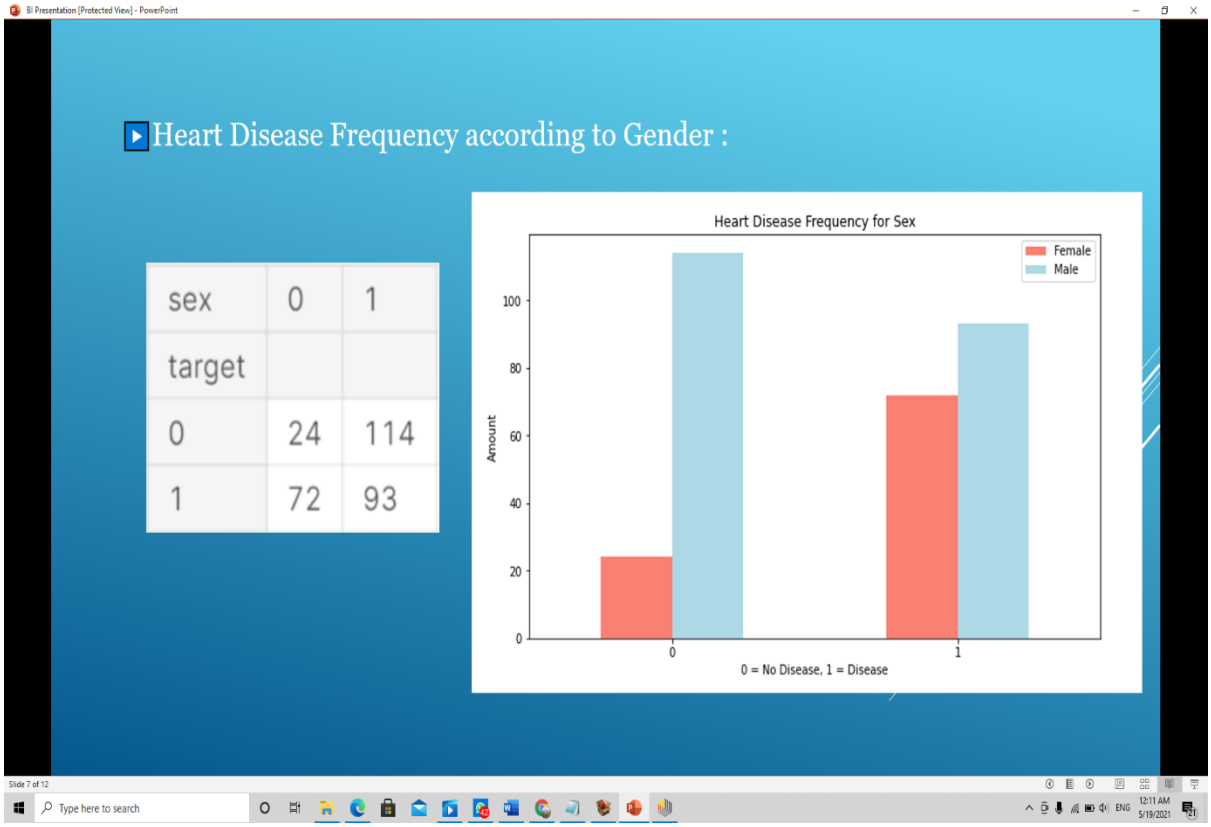
Algorithm / Classifiers which we used for heart disease analysis:

Naive Bayes classifiers are a collection of classification algorithms based on Bayes' Theorem. It is not a single algorithm but a family of algorithms where all of them share a common principle, i.e. every pair of features being classified is independent of each other.

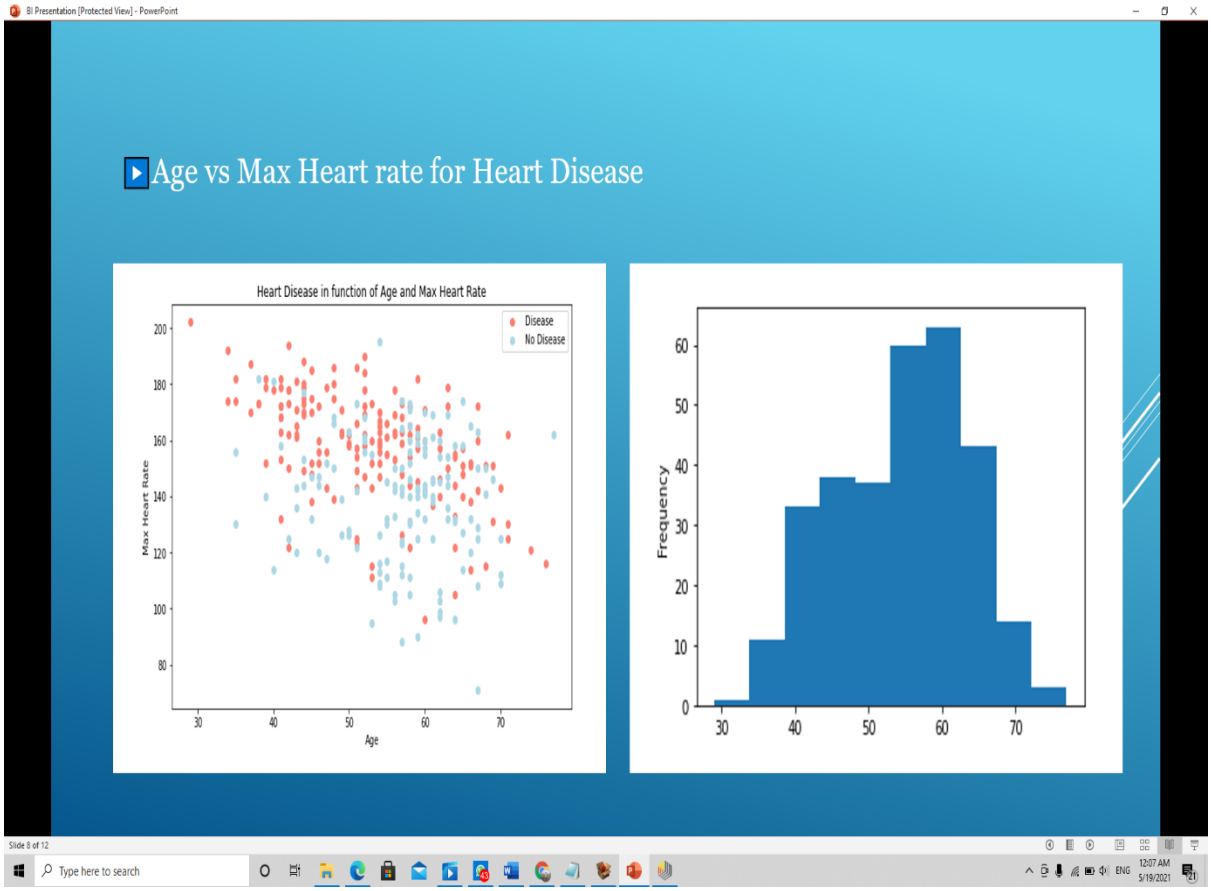
Decision tree is the most powerful and popular tool for classification and prediction. A Decision tree is a flowchart like tree structure, where each internal node denotes a test on an attribute, each branch represents an outcome of the test, and each leaf node (terminal node) holds a class label.

Random forest is a supervised learning algorithm. The "forest" it builds, is an ensemble of decision trees, usually trained with the "bagging" method. The general idea of the bagging method is that a combination of learning models increases the overall result.

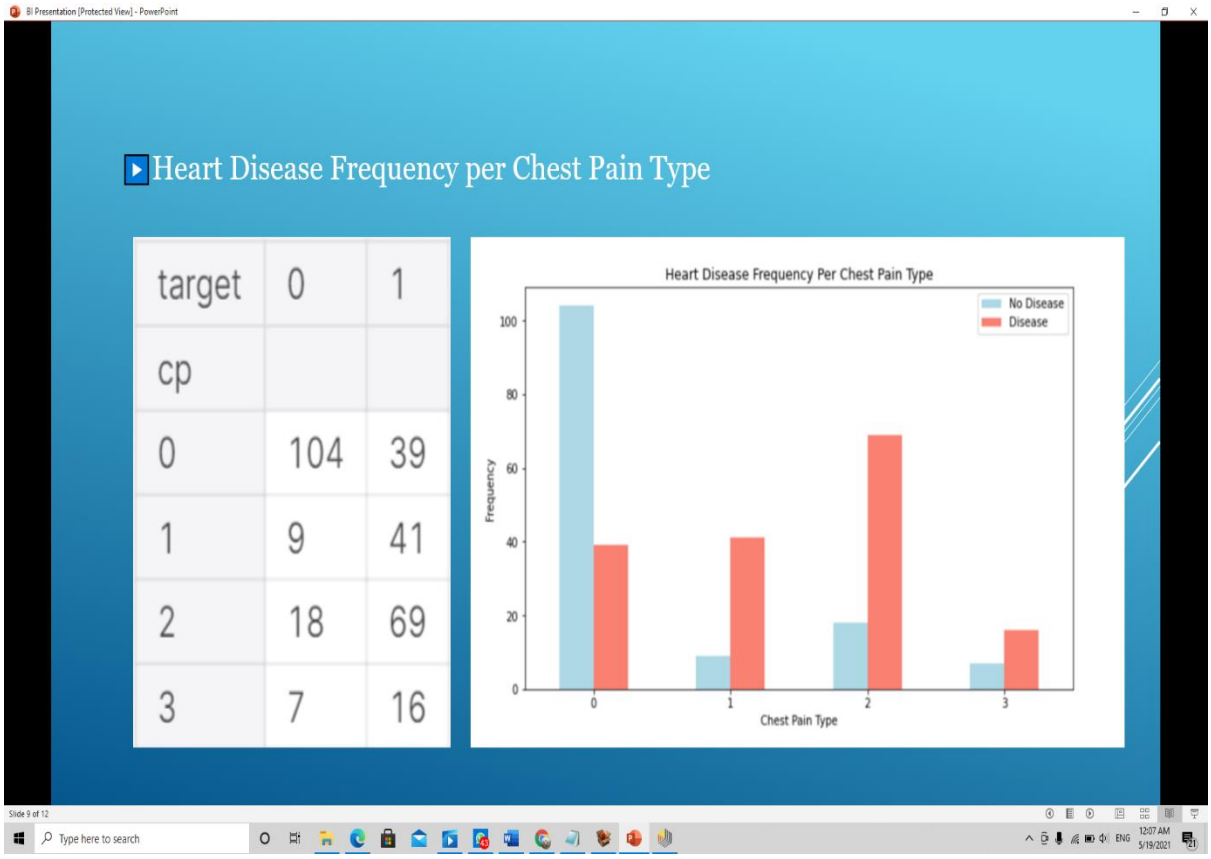
Heart Disease Frequency vs Gender:



Age vs Max Heart Rate :



Heart Disease Frequency vs Chest Pain Type:



Dashboard Analysis :

Conclusion:

- **High Blood Pressure, High Cholestrol and High Heart Rate leads to high chance of heart attack.**
- **Age from 40-60 years have the high chance of heart attack.**
- **Male gender has more chance of heart attack compared to female ones.**
- **Highly Correlated factors in this dataset are :**
  - **Age and trtbps (blood pressure rate)**
  - **Age and chol (cholestrol level)**