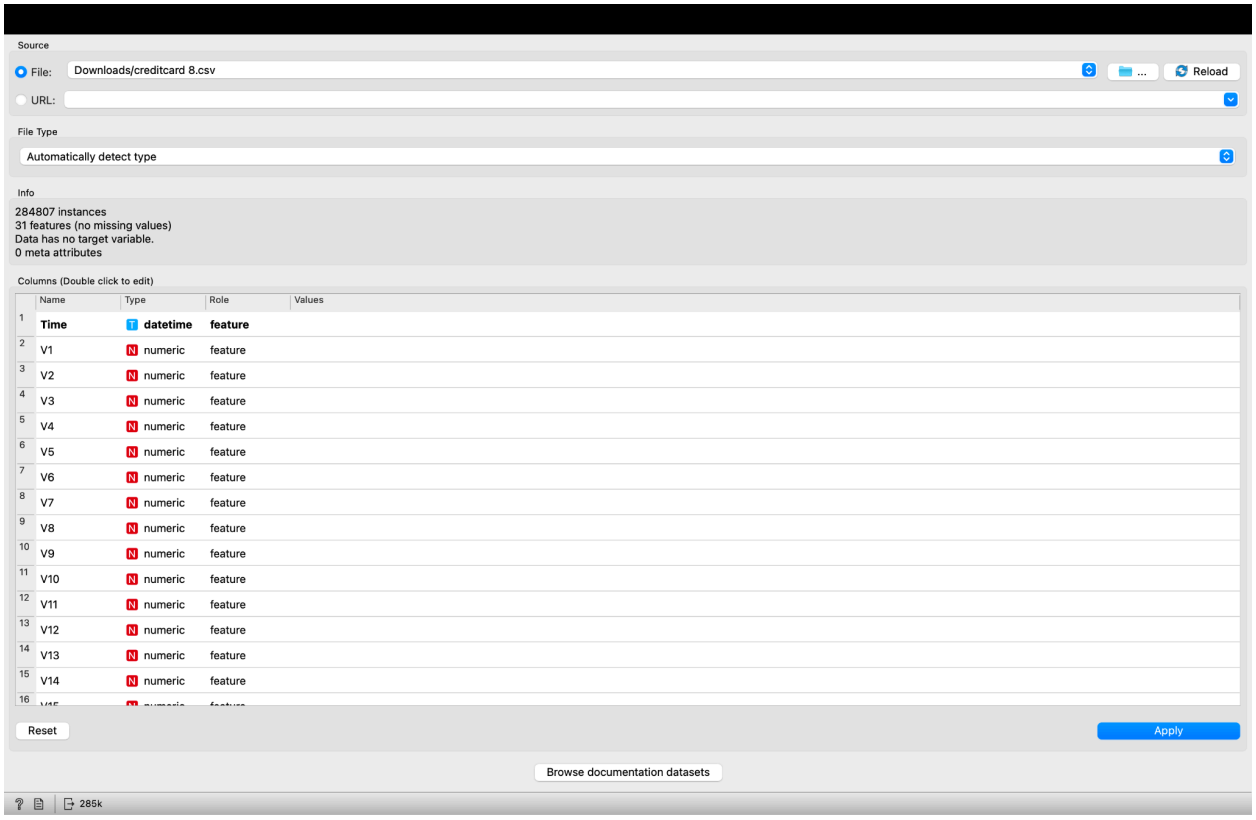
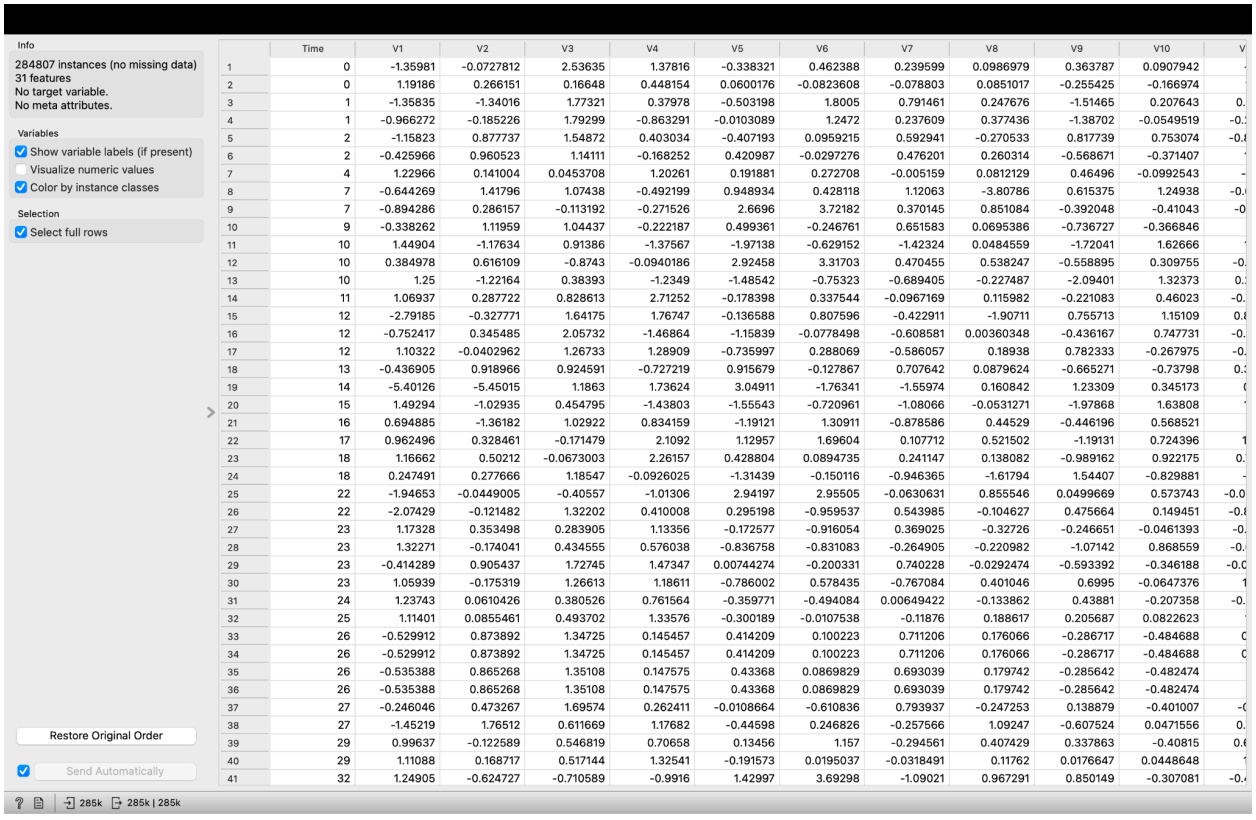
INTRODUCTION TO THE CASE

For the credit card fraud detection data i have used the data mining techniques and software i have worked for the data mining is Orange data mining which is build on the python code

EXPLORATORY DATA ANALYSIS:

I have loaded the csv file into the orange data mining to find out the statistical information about the data. After loading the source code to the software it is found that data contains 284807 instances of the data and there are 31 features out of which one is a categorical feature (class). I have verified the data types and i have rectified the data type of the time feature to be of datetime data type. It is also found that there are no missing values as per the software. Once verified with the data types, i have proceeded further to do the model engineering.





Info

284807 instances (no missing data)
31 features
No target variable.
No meta attributes.

Variables

- [x] Show variable labels (if present)
- [ ] Visualize numeric values
- [x] Color by instance classes

Selection

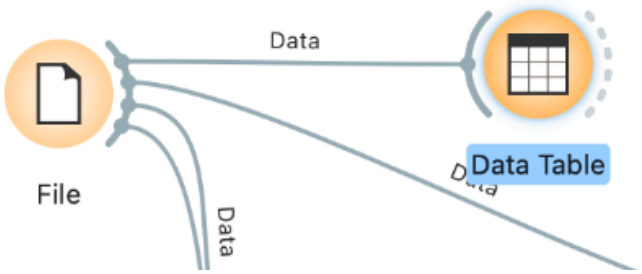
- [x] Select full rows

Restore Original Order

- [x] Send Automatically

285k | 285k | 285k

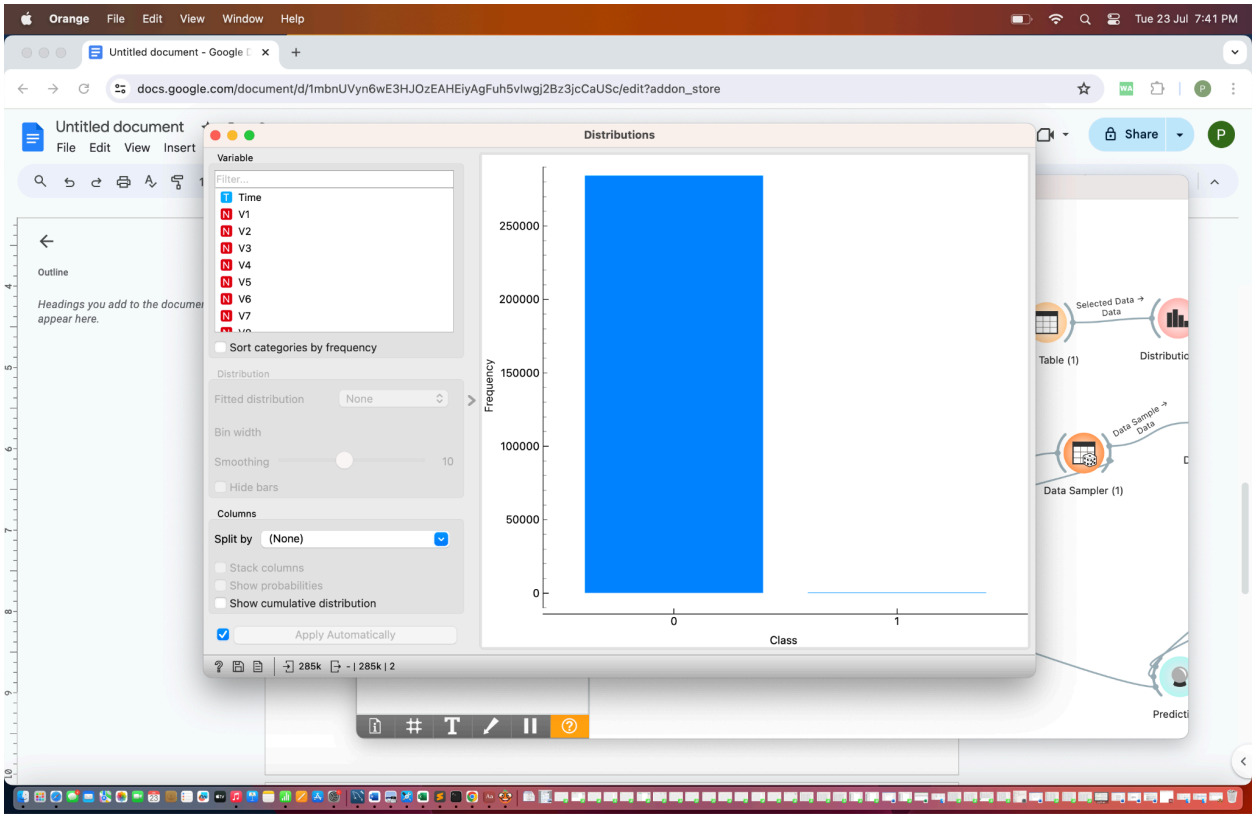
	Time	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	V
1	0	-1.35981	-0.0727812	2.53635	1.37816	-0.338321	0.462388	0.239599	0.0986979	0.363787	0.0907942	
2	0	1.19186	0.266151	0.16648	0.448154	0.0600176	-0.0823608	-0.078803	0.0851017	-0.255425	-0.166974	
3	1	-1.35835	-1.34016	1.77321	0.37978	-0.503198	1.8005	0.791461	0.247676	-1.51465	0.207643	0.
4	1	-0.966272	-0.185226	1.79299	-0.863291	-0.0103089	1.2472	0.237609	0.377436	-1.38702	-0.0549519	-0.
5	2	-1.15823	0.877737	1.54872	0.403034	-0.407193	0.0959215	0.592941	-0.270533	0.817739	0.753074	-0.
6	2	-0.425966	0.960523	1.14111	-0.168252	0.420987	-0.0297276	0.476201	0.260314	-0.568671	-0.371407	
7	4	1.22966	0.141004	0.0453708	1.20261	0.191881	0.272708	-0.005159	0.0812129	0.46496	-0.0992543	
8	7	-0.644269	1.41796	1.07438	-0.492199	0.948934	0.428118	1.12063	-3.80786	0.615375	1.24938	-0.
9	7	-0.894286	0.286157	-0.113192	-0.271526	2.6696	3.72182	0.370145	0.851084	-0.392048	-0.41043	-0
10	9	-0.338262	1.11959	1.04437	-0.222187	0.499361	-0.246761	0.651583	0.0695386	-0.736727	-0.366846	
11	10	1.44904	-1.17634	0.91386	-1.37567	-1.97138	-0.629152	-1.42324	0.0484559	-1.72041	1.62666	
12	10	0.384978	0.616109	-0.8743	-0.0940186	2.92458	3.31703	0.470455	0.538247	-0.558895	0.309755	-0.
13	10	1.25	-1.22164	0.38393	-1.2349	-1.48542	-0.75323	-0.689405	-0.227487	-2.09401	1.32373	0.
14	11	1.06937	0.287722	0.828613	2.71252	-0.178398	0.337544	-0.0967169	0.115982	-0.221083	0.46023	-0.
15	12	-2.79185	-0.327771	1.64175	1.76747	-0.136588	0.807596	-0.422911	-1.90711	0.755713	1.15109	0.
16	12	-0.752417	0.345485	2.05732	-1.46864	-1.15839	-0.0778498	-0.608581	0.00360348	-0.436167	0.747731	-0.
17	12	1.10322	-0.0402962	1.26733	1.28909	-0.735997	0.288069	-0.586057	0.18938	0.782333	-0.267975	-0.
18	13	-0.436905	0.918966	0.924591	-0.727219	0.915679	-0.127867	0.707642	0.0879624	-0.665271	-0.73798	0.
19	14	-5.40126	-5.45015	1.1863	1.73624	3.04911	-1.76341	-1.55974	0.160842	1.23309	0.345173	
20	15	1.49294	-1.02935	0.454795	-1.43803	-1.55543	-0.720961	-1.08066	-0.0531271	-1.97868	1.63808	
21	16	0.694885	-1.36182	1.02922	0.834159	-1.19121	1.30911	-0.878586	0.44529	-0.446196	0.568521	
22	17	0.962496	0.328461	-0.171479	2.1092	1.12957	1.69604	0.107712	0.521502	-1.19131	0.724396	1
23	18	1.16662	0.50212	-0.0673003	2.26157	0.428804	0.0894735	0.241147	0.138082	-0.989162	0.922175	0.
24	18	0.247491	0.277666	1.18547	-0.0926025	-1.31439	-0.150116	-0.946365	-1.61794	1.54407	-0.829881	
25	22	-1.94653	-0.0449005	-0.40557	-1.01306	2.94197	2.95505	-0.0630631	0.855546	0.0499669	0.573743	-0.0
26	22	-2.07429	-0.121482	1.32202	0.410008	0.295198	-0.959537	0.543985	-0.104627	0.475664	0.149451	-0.
27	23	1.17328	0.353498	0.283905	1.13356	-0.172577	-0.916054	0.369025	-0.32726	-0.246651	-0.0461393	-0.
28	23	1.32271	-0.174041	0.434555	0.576038	-0.836758	-0.831083	-0.264905	-0.220982	-1.07142	0.868559	-0.
29	23	-0.414289	0.905437	1.72745	1.47347	0.00744274	-0.200331	0.740228	-0.0292474	-0.593392	-0.346188	-0.0
30	23	1.05939	-0.175319	1.26613	1.18611	-0.786002	0.578435	-0.767084	0.401046	0.6995	-0.0647376	1
31	24	1.23743	0.0610426	0.380526	0.761564	-0.359771	-0.494084	0.00649422	-0.133862	0.43881	-0.207358	-0.
32	25	1.11401	0.0855461	0.493702	1.33576	-0.300189	-0.0107538	-0.11876	0.188617	0.205687	0.0822623	
33	26	-0.529912	0.873892	1.34725	0.145457	0.414209	0.100223	0.711206	0.176066	-0.286717	-0.484688	0
34	26	-0.529912	0.873892	1.34725	0.145457	0.414209	0.100223	0.711206	0.176066	-0.286717	-0.484688	0
35	26	-0.535388	0.865268	1.35108	0.147575	0.43368	0.0869829	0.693039	0.179742	-0.285642	-0.482474	
36	26	-0.535388	0.865268	1.35108	0.147575	0.43368	0.0869829	0.693039	0.179742	-0.285642	-0.482474	
37	27	-0.246046	0.473267	1.69574	0.262411	-0.0108664	-0.610836	0.793937	-0.247253	0.138879	-0.401007	-0
38	27	-1.45219	1.76512	0.611669	1.17682	-0.44598	0.246826	-0.257566	1.09247	-0.607524	0.0471556	0.
39	29	0.99637	-0.122589	0.546819	0.70658	0.13456	1.157	-0.294561	0.407429	0.337863	-0.40815	0.
40	29	1.11088	0.168717	0.517144	1.32541	-0.191573	0.0195037	-0.0318491	0.11762	0.0176647	0.0448648	1
41	32	1.24905	-0.624727	-0.710589	-0.9916	1.42997	3.69298	-1.09021	0.967291	0.850149	-0.307081	-0.

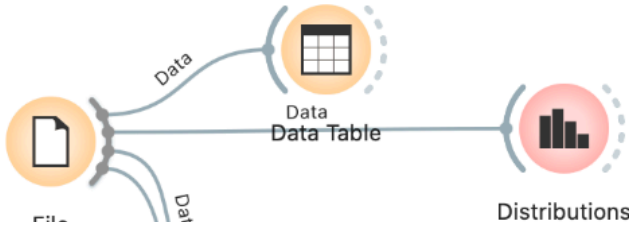


DATA CLEANING

Since there are no missing values in the data, data cleaning was not necessary.

After doing the exploratory data analysis it has been found that the data is highly imbalance as we have 284315 transactions which are found to be good transactions.which translates to 99.83% and we have 492 fradulent transactions which is of around 0.17% .

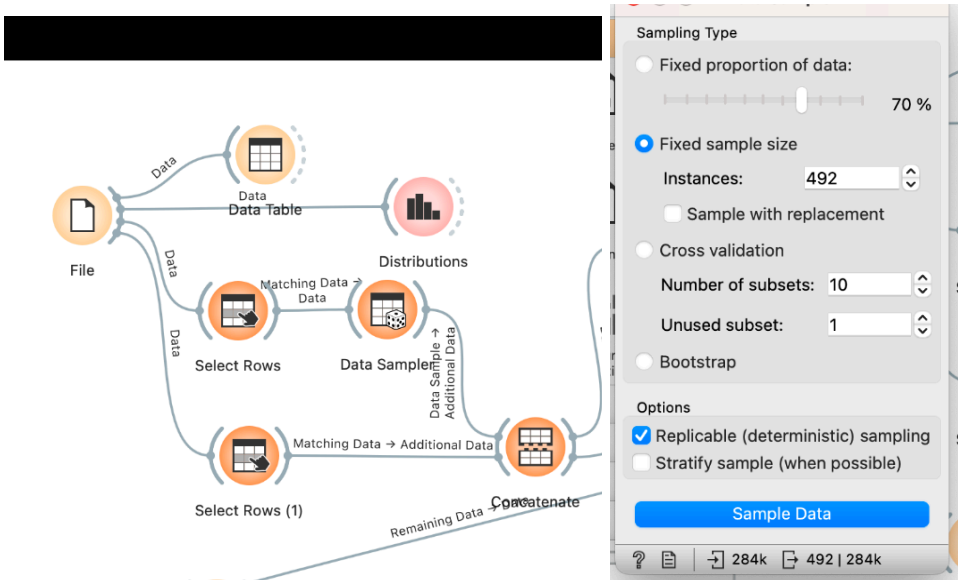


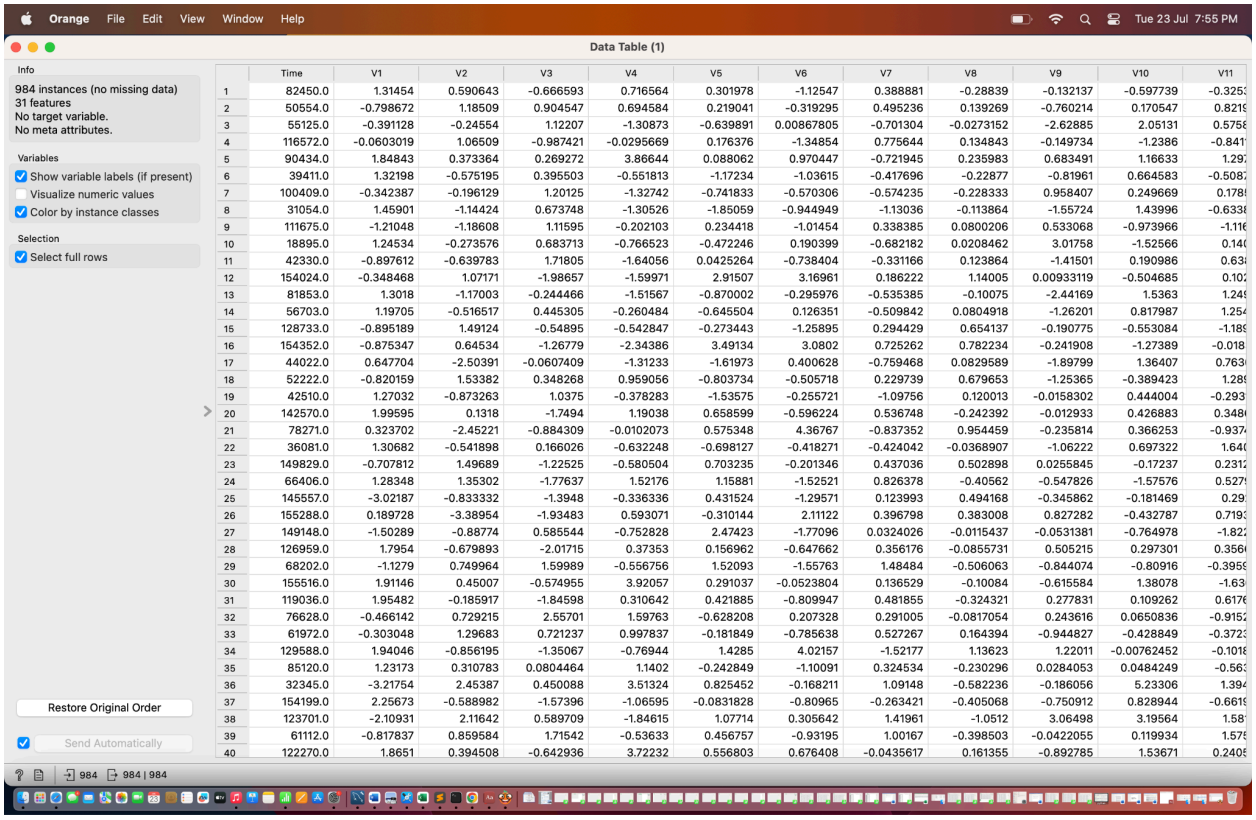


Since the data is highly imbalanced it is necessary to balance the data before doing any further analysis

DEALING WITH IMBALANCED DATA

To balance the data we have two options. One is two under populate the majority class or over populate the minority class. The option i have chosen is to under populate the majority class to reduce the complexity of computation. For this i have used the select rows widget and filtered out only the good transactions and used the data sampler to reduce it to 492 transactions. I have further obtained the fraudulent transactions and merged the underpopulated good transactions with the fraudulent transactions.
The final balanced data set contains 984 records which is a mix of both good and fraudulent transactions.





	Time	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	V11
1	82450.0	1.31454	0.590643	-0.666593	0.716564	0.301978	-1.12547	0.388881	-0.28839	-0.132137	-0.597739	-0.325
2	50554.0	-0.798672	1.18509	0.904547	0.694584	0.219041	-0.319295	0.495236	0.139269	-0.760214	0.170547	0.821
3	55125.0	-0.391128	-0.24554	1.12207	-1.30873	-0.639891	0.00867805	-0.701304	-0.0273152	-2.62885	2.05131	0.575
4	116572.0	-0.0603019	1.06509	-0.987421	-0.0295669	0.176376	-1.34854	0.775644	0.134843	-0.149734	-1.2386	-0.841
5	90434.0	1.84843	0.373364	0.269272	3.86644	0.088062	0.970447	-0.721945	0.235983	0.683491	1.16633	1.29
6	39411.0	1.32198	-0.575195	0.395503	-0.551813	-1.17234	-1.03615	-0.417696	-0.22877	-0.81961	0.664583	-0.508
7	100409.0	-0.342387	-0.196129	1.20125	-1.32742	-0.741833	-0.570306	-0.574235	-0.228333	0.958407	0.249669	0.178
8	31054.0	1.45901	-1.14424	0.673748	-1.30526	-1.85059	-0.944949	-1.13036	-0.113864	-1.55724	1.43996	-0.633
9	111675.0	-1.21048	-1.18608	1.11595	-0.202103	0.234418	-1.01454	0.338385	0.0800206	0.533068	-0.973966	-1.116
10	18895.0	1.24534	-0.273576	0.683713	-0.766523	-0.472246	0.190399	-0.682182	0.0208462	3.01758	-1.52566	0.14
11	42330.0	-0.897612	-0.639783	1.71805	-1.64056	0.0425264	-0.738404	-0.331166	0.123864	-1.41501	0.190986	0.63
12	154024.0	-0.348468	1.07171	-1.98657	-1.59971	2.91507	3.16961	0.186222	1.14005	0.00933119	-0.504685	0.10
13	81853.0	1.3018	-1.17003	-0.244466	-1.51567	-0.870002	-0.295976	-0.535385	-0.10075	-2.44169	1.5363	1.24
14	56703.0	1.19705	-0.516517	0.445305	-0.260484	-0.645504	0.126351	-0.509842	0.0804918	-1.26201	0.817987	1.25
15	128733.0	-0.895189	1.49124	-0.54895	-0.542847	-0.273443	-1.25895	0.294429	0.654137	-0.190775	-0.553084	-1.18
16	154352.0	-0.875347	0.64534	-1.26779	-2.34386	3.49134	3.0802	0.725262	0.782234	-0.241908	-1.27389	-0.018
17	44022.0	0.647704	-2.50391	-0.0607409	-1.31233	-1.61973	0.400628	-0.759468	0.0829589	-1.89799	1.36407	0.763
18	52222.0	-0.820159	1.53382	0.348268	0.959056	-0.803734	-0.505718	0.229739	0.679653	-1.25365	-0.389423	1.28
19	42510.0	1.27032	-0.873263	1.0375	-0.378283	-1.53575	-0.255721	-1.09756	0.120013	-0.0158302	0.444004	-0.293
20	142570.0	1.99595	0.1318	-1.7494	1.19038	0.658599	-0.596224	0.536748	-0.242392	-0.012933	0.426883	0.348
21	78271.0	0.323702	-2.45221	-0.884309	-0.0102073	0.575348	4.36767	-0.837352	0.954459	-0.235814	0.366253	-0.937
22	36081.0	1.30682	-0.541898	0.166026	-0.632248	-0.698127	-0.418271	-0.424042	-0.0368907	-1.06222	0.697322	1.64
23	149829.0	-0.707812	1.49689	-1.22525	-0.580504	0.703235	-0.201346	0.437036	0.502898	0.0255845	-0.17237	0.231
24	66406.0	1.28348	1.35302	-1.77637	1.52176	1.15881	-1.52521	0.826378	-0.40562	-0.547826	-1.57576	0.527
25	145557.0	-3.02187	-0.833332	-1.3948	-0.336336	0.431524	-1.29571	0.123993	0.494168	-0.345862	-0.181469	0.29
26	155288.0	0.189728	-3.38954	-1.93483	0.593071	-0.310144	2.11122	0.396798	0.383008	0.827282	-0.432787	0.719
27	149148.0	-1.50289	-0.88774	0.585544	-0.752828	2.47423	-1.77096	0.0324026	-0.0115437	-0.0531381	-0.764978	-1.82
28	126959.0	1.7954	-0.679893	-2.01715	0.37353	0.156962	-0.647662	0.356176	-0.0855731	0.505215	0.297301	0.356
29	68202.0	-1.1279	0.749964	1.59989	-0.556756	1.52093	-1.55763	1.48484	-0.506063	-0.844074	-0.80916	-0.3959
30	155516.0	1.91146	0.45007	-0.574955	3.92057	0.291037	-0.0523804	0.136529	-0.10084	-0.615584	1.38078	-1.63
31	119036.0	1.95482	-0.185917	-1.84598	0.310642	0.421885	-0.809947	0.481855	-0.324321	0.277831	0.109262	0.617
32	76628.0	-0.466142	0.729215	2.55701	1.59763	-0.628208	0.207328	0.291005	-0.0817054	0.243616	0.0650836	-0.915
33	61972.0	-0.303048	1.29683	0.721237	0.997837	-0.181849	-0.785638	0.527267	0.164394	-0.944827	-0.428849	-0.372
34	129588.0	1.94046	-0.856195	-1.35067	-0.76944	1.4285	4.02157	-1.52177	1.13623	1.22011	-0.00762452	-0.101
35	85120.0	1.23173	0.310783	0.0804464	1.1402	-0.242849	-1.10091	0.324534	-0.230296	0.0284053	0.0484249	-0.56
36	32345.0	-3.21754	2.45387	0.450088	3.51324	0.825452	-0.168211	1.09148	-0.582236	-0.186056	5.23306	1.394
37	154199.0	2.25673	-0.588982	-1.57396	-1.06595	-0.0831828	-0.80965	-0.263421	-0.405068	-0.750912	0.828944	-0.661
38	123701.0	-2.10931	2.11642	0.589709	-1.84615	1.07714	0.305642	1.41961	-1.0512	3.06498	3.19564	1.58
39	61112.0	-0.817837	0.859584	1.71542	-0.53633	0.456757	-0.93195	1.00167	-0.398503	-0.0422055	0.119934	1.575
40	122270.0	1.8651	0.394508	-0.642936	3.72232	0.556803	0.676408	-0.0435617	0.161355	-0.892785	1.53671	0.240

FEATURE ENGINEERING

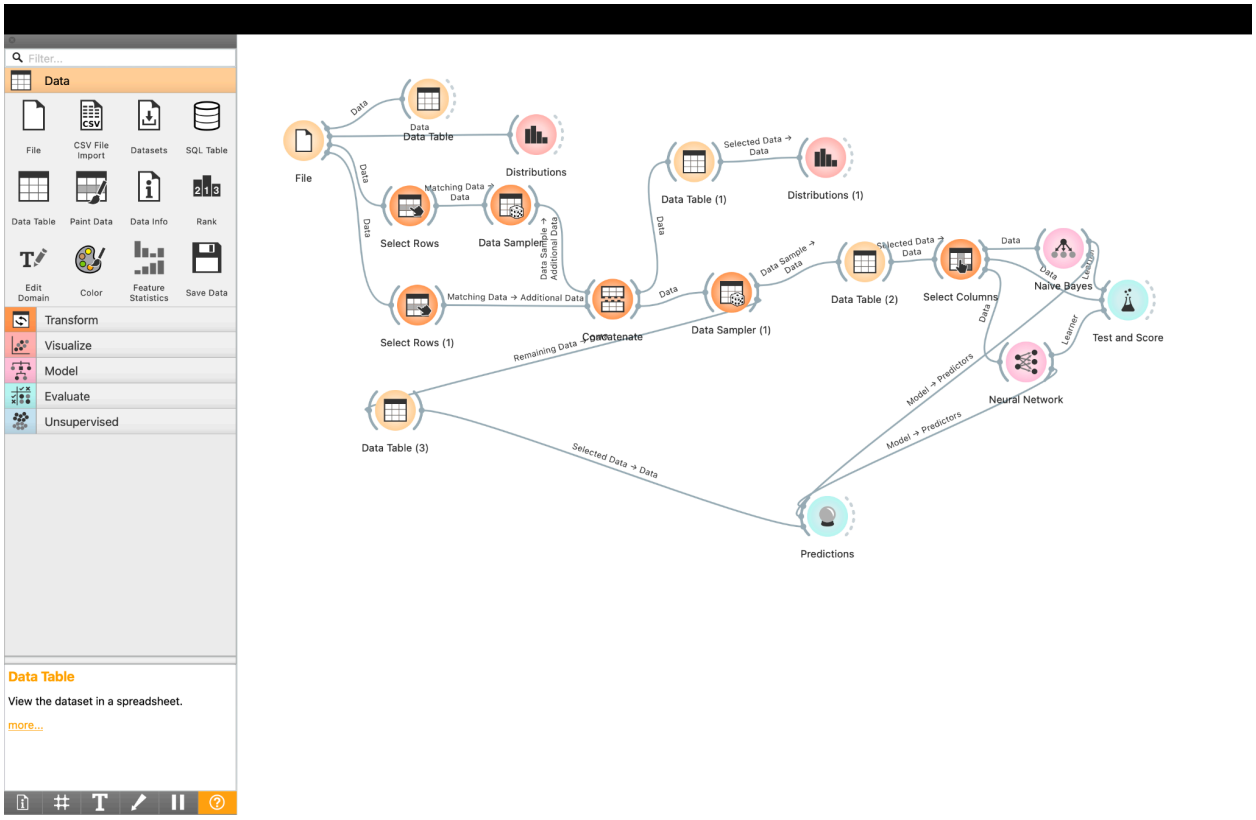
Since the data contains 31 features out of which, 28 features are from PCA. Remaining 3 features can be used for model selection and model building

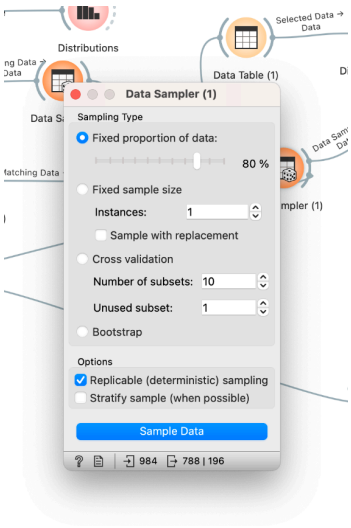
MODEL SELECTION

For this project, i have used two models which are good at classification. The models that i have used are neural networks and naive bayes. These two models have been tested further and validated accordingly

MODEL TRAINING

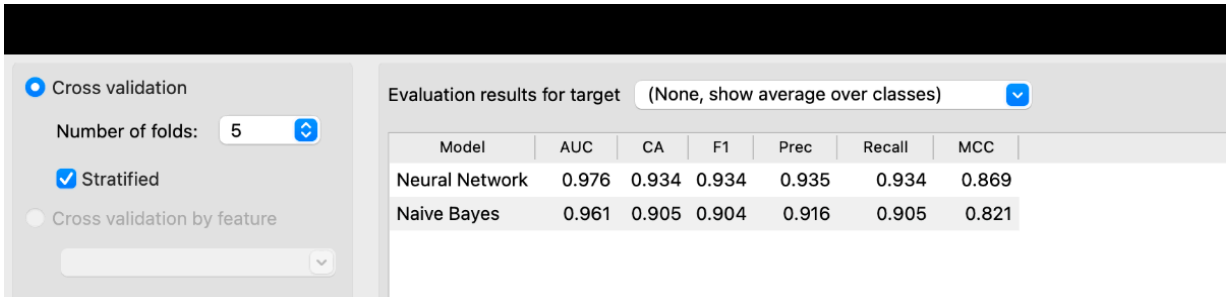
I have further used the balanced data set to split the data into training data set and test data set. The training data set has been used for model building and validation which is further tested on the test data. I have split the data as 80:20 ratio where 80% is the training data set and 20% is the testing data set.
The below screenshots indicate the same





MODEL VALIDATION

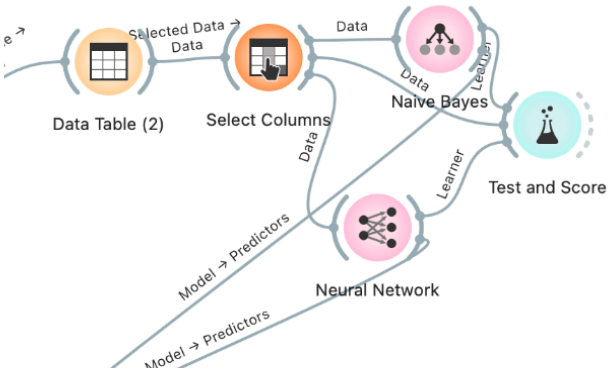
I have validated the model which i have built and tested for its accuracy using the test and score widget.



Model	AUC	CA	F1	Prec	Recall	MCC
Neural Network	0.976	0.934	0.934	0.935	0.934	0.869
Naive Bayes	0.961	0.905	0.904	0.916	0.905	0.821

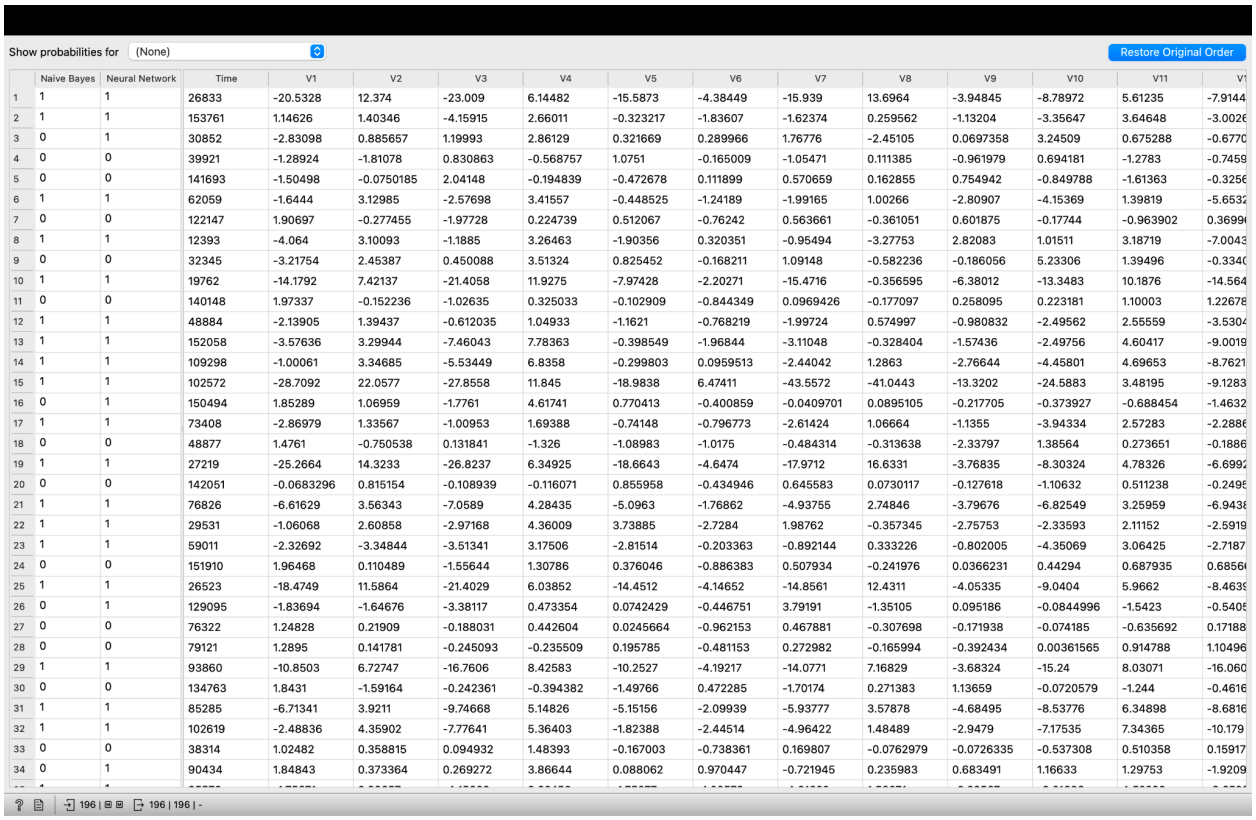
It is found that both the classifiers have more than 90% accuracy. Which can be further used to test my test data and deploy the model

tributions (1)



MODEL DEPLOYMENT

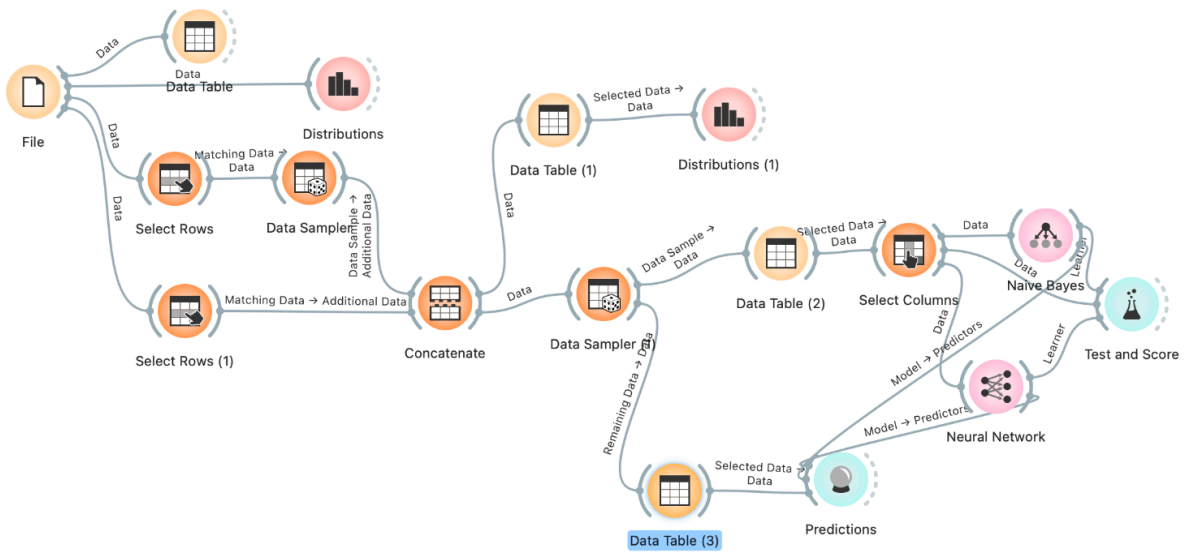
Now i have built the model, i can verify the model using the test data set. I have used the predictions widget to predict which transactions are good and which transactions are fraudulent. It has been found that the model has worked as per the requirement and has successfully classified the good transactions and fraudulent transactions



Show probabilities for (None) Restore Original Order

	Naive Bayes	Neural Network	Time	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	V11	V1
1	1	1	26833	-20.5328	12.374	-23.009	6.14482	-15.5873	-4.38449	-15.939	13.6964	-3.94845	-8.78972	5.61235	-7.9144
2	1	1	153761	1.14626	1.40346	-4.15915	2.66011	-0.323217	-1.83607	-1.62374	0.259562	-1.13204	-3.35647	3.64648	-3.0026
3	0	1	30852	-2.83098	0.885657	1.19993	2.86129	0.321669	0.289966	1.76776	-2.45105	0.0697358	3.24509	0.675288	-0.6770
4	0	0	39921	-1.28924	-1.81078	0.830863	-0.568757	1.0751	-0.165009	-1.05471	0.111385	-0.961979	0.694181	-1.2783	-0.7459
5	0	0	141693	-1.50498	-0.0750185	2.04148	-0.194839	-0.472678	0.111899	0.570659	0.162855	0.754942	-0.849788	-1.61363	-0.3256
6	1	1	62059	-1.6444	3.12985	-2.57698	3.41557	-0.448525	-1.24189	-1.99165	1.00266	-2.80907	-4.15369	1.39819	-5.6532
7	0	0	122147	1.90697	-0.277455	-1.97728	0.224739	0.512067	-0.76242	0.563661	-0.361051	0.601875	-0.17744	-0.963902	0.3699
8	1	1	12393	-4.064	3.10093	-1.1885	3.26463	-1.90356	0.320351	-0.95494	-3.27753	2.82083	1.01511	3.18719	-7.0043
9	0	0	32345	-3.21754	2.45387	0.450088	3.51324	0.825452	-0.168211	1.09148	-0.582236	-0.186056	5.23306	1.39496	-0.3340
10	1	1	19762	-14.1792	7.42137	-21.4058	11.9275	-7.97428	-2.20271	-15.4716	-0.356595	-6.38012	-13.3483	10.1876	-14.564
11	0	0	140148	1.97337	-0.152236	-1.02635	0.325033	-0.102909	-0.844349	0.0969426	-0.177097	0.258095	0.223181	1.10003	1.22678
12	1	1	48884	-2.13905	1.39437	-0.612035	1.04933	-1.1621	-0.768219	-1.99724	0.574997	-0.980832	-2.49562	2.55559	-3.5304
13	1	1	152058	-3.57636	3.29944	-7.46043	7.78363	-0.398549	-1.96844	-3.11048	-0.328404	-1.57436	-2.49756	4.60417	-9.0019
14	1	1	109298	-1.00061	3.34685	-5.53449	6.8358	-0.299803	0.0959513	-2.44042	1.2863	-2.76644	-4.45801	4.69653	-8.7621
15	1	1	102572	-28.7092	22.0577	-27.8558	11.845	-18.9838	6.47411	-43.5572	-41.0443	-13.3202	-24.5883	3.48195	-9.1283
16	0	1	150494	1.85289	1.06959	-1.7761	4.61741	0.770413	-0.400859	-0.0409701	0.0895105	-0.217705	-0.373927	-0.688454	-1.4632
17	1	1	73408	-2.86979	1.33567	-1.00953	1.69388	-0.74148	-0.796773	-2.61424	1.06664	-1.1355	-3.94334	2.57283	-2.2886
18	0	0	48877	1.4761	-0.750538	0.131841	-1.326	-1.08983	-1.0175	-0.484314	-0.313638	-2.33797	1.38564	0.273651	-0.1886
19	1	1	27219	-25.2664	14.3233	-26.8237	6.34925	-18.6643	-4.6474	-17.9712	16.6331	-3.76835	-8.30324	4.78326	-6.6992
20	0	0	142051	-0.0683296	0.815154	-0.108939	-0.116071	0.855958	-0.434946	0.645583	0.0730117	-0.127618	-1.10632	0.511238	-0.2495
21	1	1	76826	-6.61629	3.56343	-7.0589	4.28435	-5.0963	-1.76862	-4.93755	2.74846	-3.79676	-6.82549	3.25959	-6.9438
22	1	1	29531	-1.06068	2.60858	-2.97168	4.36009	3.73885	-2.7284	1.98762	-0.357345	-2.75753	-2.33593	2.11152	-2.5919
23	1	1	59011	-2.32692	-3.34844	-3.51341	3.17506	-2.81514	-0.203363	-0.892144	0.333226	-0.802005	-4.35069	3.06425	-2.7187
24	0	0	151910	1.96468	0.110489	-1.55644	1.30786	0.376046	-0.886383	0.507934	-0.241976	0.0366231	0.44294	0.687935	0.6856
25	1	1	26523	-18.4749	11.5864	-21.4029	6.03852	-14.4512	-4.14652	-14.8561	12.4311	-4.05335	-9.0404	5.9662	-8.4639
26	0	1	129095	-1.83694	-1.64676	-3.38117	0.473354	0.0742429	-0.446751	3.79191	-1.35105	0.095186	-0.0844996	-1.5423	-0.5405
27	0	0	76322	1.24828	0.21909	-0.188031	0.442604	0.0245664	-0.962153	0.467881	-0.307698	-0.171938	-0.074185	-0.635692	0.17188
28	0	0	79121	1.2895	0.141781	-0.245093	-0.235509	0.195785	-0.481153	0.272982	-0.165994	-0.392434	0.00361565	0.914788	1.10496
29	1	1	93860	-10.8503	6.72747	-16.7606	8.42583	-10.2527	-4.19217	-14.0771	7.16829	-3.68324	-15.24	8.03071	-16.060
30	0	0	134763	1.8431	-1.59164	-0.242361	-0.394382	-1.49766	0.472285	-1.70174	0.271383	1.13659	-0.0720579	-1.244	-0.4616
31	1	1	85285	-6.71341	3.9211	-9.74668	5.14826	-5.15156	-2.09939	-5.93777	3.57878	-4.68495	-8.53776	6.34898	-8.6816
32	1	1	102619	-2.48836	4.35902	-7.77641	5.36403	-1.82388	-2.44514	-4.96422	1.48489	-2.9479	-7.17535	7.34365	-10.179
33	0	0	38314	1.02482	0.358815	0.094932	1.48393	-0.167003	-0.738361	0.169807	-0.0762979	-0.0726335	-0.537308	0.510358	0.15917
34	0	1	90434	1.84843	0.373364	0.269272	3.86644	0.088062	0.970447	-0.721945	0.235983	0.683491	1.16633	1.29753	-1.9209

196 | 196 | 196 | -


Data
Data Table
File
Distributions
Data
Matching Data → Data
Select Rows
Data Sampler
Data Sample → Additional Data
Data
Data Table (1)
Selected Data → Data
Distributions (1)
Data
Matching Data → Additional Data
Select Rows (1)
Concatenate
Data
Data Sampler (2)
Data Sample → Data
Remaining Data → Data
Data Table (2)
Selected Data → Data
Select Columns
Data
Naive Bayes
Learner
Test and Score
Data
Model → Predictors
Learner
Neural Network
Model → Predictors
Selected Data → Data
Predictions
Data Table (3)