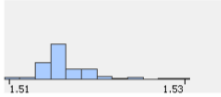
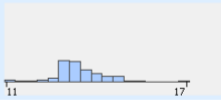
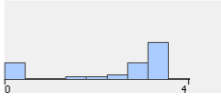
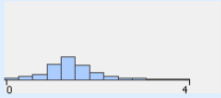
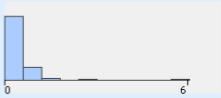
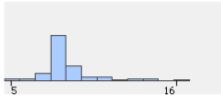
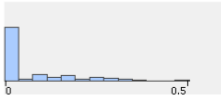
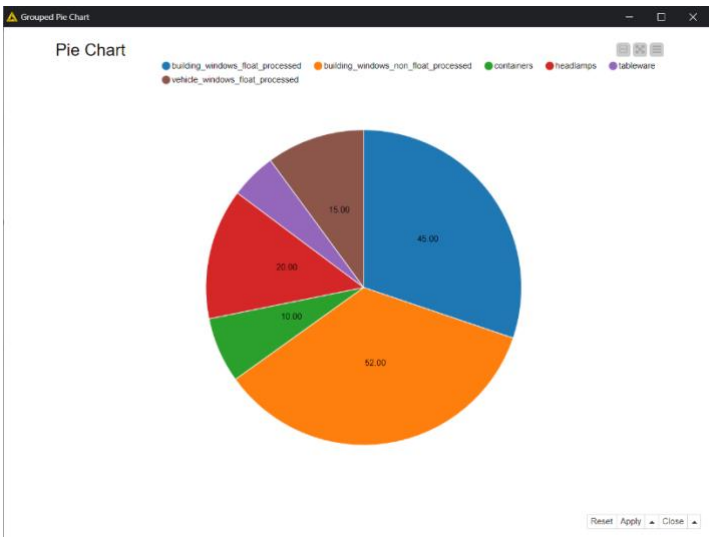
MACHINE LEARNING LABORATORY – ONSPOT EXERCISE
LAB – 03 – KNIME TOOL

Various algorithm using Knime tool:-

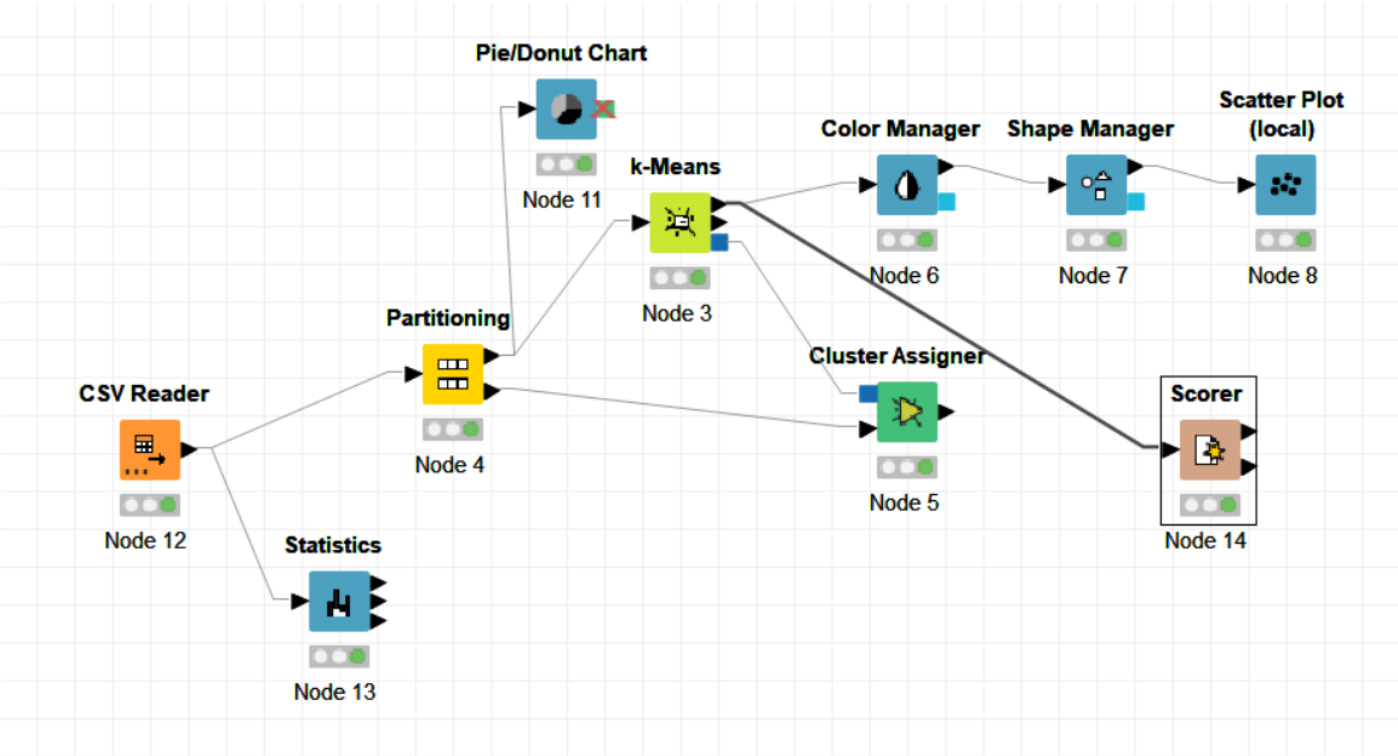
Dataset – [Index of /ml/machine-learning-databases/glass (uci.edu)]

Statistics:-

Column	Min	Mean	Median	Max	Std. Dev.	Skewness	Kurtosis	No. Missing	No. +∞	No. -∞	Histogram
RI	1.5111	1.5184	?	1.5339	0.003	1.6254	4.9317	0	0	0	
Na	10.73	13.4079	?	17.38	0.8166	0.4542	3.0522	0	0	0	
Mg	0.0	2.6845	?	4.49	1.4424	-1.1526	-0.4103	0	0	0	
Al	0.29	1.4449	?	3.5	0.4993	0.9073	2.0606	0	0	0	
Si	69.81	72.6509	?	75.41	0.7745	-0.7304	2.9679	0	0	0	
K	0.0	0.4971	?	6.21	0.6522	6.5516	54.6897	0	0	0	
Ca	5.43	8.957	?	16.19	1.4232	2.0471	6.682	0	0	0	
Ba	0.0	0.175	?	3.15	0.4972	3.4164	12.5411	0	0	0	
Fe	0.0	0.057	?	0.51	0.0974	1.7543	2.662	0	0	0	



1. **Clustering algorithm (K-means) :-**



Cluster assigner (some rows are shown):-

Console | Node Monitor

Node: Cluster Assigner (5:5)

State: EXECUTED

Port Output Port 0 Load data Rows: 65, Columns: 11

ID	RI	Na	Mg	Al	Si	K	Ca	Ba	Fe	Type	Cluster
2	1.51761	13.89	3.6	1.36	72.73	0.48	7.83	0.0	0.0	building_windows_float_processed	cluster_1
3	1.51618	13.53	3.55	1.54	72.99	0.39	7.78	0.0	0.0	building_windows_float_processed	cluster_1
9	1.51918	14.04	3.58	1.37	72.08	0.56	8.3	0.0	0.0	building_windows_float_processed	cluster_1
13	1.51589	12.88	3.43	1.4	73.28	0.69	8.05	0.0	0.24	building_windows_float_processed	cluster_1
15	1.51763	12.61	3.59	1.31	73.29	0.58	8.5	0.0	0.0	building_windows_float_processed	cluster_1
17	1.51784	12.68	3.67	1.16	73.11	0.61	8.7	0.0	0.0	building_windows_float_processed	cluster_1
18	1.52196	14.36	3.85	0.89	71.36	0.15	9.15	0.0	0.0	building_windows_float_processed	cluster_1
25	1.5172	13.38	3.5	1.15	72.85	0.5	8.43	0.0	0.0	building_windows_float_processed	cluster_1

ID	RI	Na	Mg	Al	Si	K	Ca	Ba	Fe	Type	Cluster
165	1.51915	12.73	1.85	1.86	72.69	0.6	10...	0.0	0.0	containers	cluster_1
167	1.52151	11.03	1.71	1.56	73.44	0.58	11...	0.0	0.0	containers	cluster_0
174	1.52043	13.38	0.0	1.4	72.25	0.33	12.5	0.0	0.0	containers	cluster_0
183	1.51916	14.15	0.0	2.09	72.74	0.0	10...	0.0	0.0	tableware	cluster_0
184	1.51969	14.56	0.0	0.56	73.48	0.0	11...	0.0	0.0	tableware	cluster_0
186	1.51131	13.69	3.2	1.81	72.81	1.76	5.43	1.19	0.0	headlamps	cluster_1
187	1.51838	14.32	3.26	2.22	71.25	1.46	5.79	1.63	0.0	headlamps	cluster_1
188	1.52315	13.44	3.34	1.23	72.38	0.6	8.83	0.0	0.0	headlamps	cluster_1
190	1.52365	15.79	1.83	1.31	70.43	0.31	8.61	1.68	0.0	headlamps	cluster_2
191	1.51613	13.88	1.78	1.79	73.1	0.0	8.67	0.76	0.0	headlamps	cluster_2
196	1.51545	14.14	0.0	2.68	73.39	0.08	9.07	0.61	0.05	headlamps	cluster_2
210	1.51623	14.14	0.0	2.88	72.61	0.08	9.18	1.06	0.0	headlamps	cluster_2
211	1.51685	14.92	0.0	1.99	73.06	0.0	8.4	1.59	0.0	headlamps	cluster_2

Clustering output

Console | Node Monitor

Node: Scorer (5:14)

State: EXECUTED

Port Output Port 0 | Load data | Rows: 9, Columns: 9

ID	cluster_0	cluster_1	cluster_2
building_windows_float_processed	0	45	0
building_windows_non_float_processed	10	42	0
vehicle_windows_float_processed	0	15	0
containers	5	2	3
tableware	0	4	3
headlamps	0	1	19

Changing number of clusters to 6 and number of iterations to 200:-

Console | Node Monitor

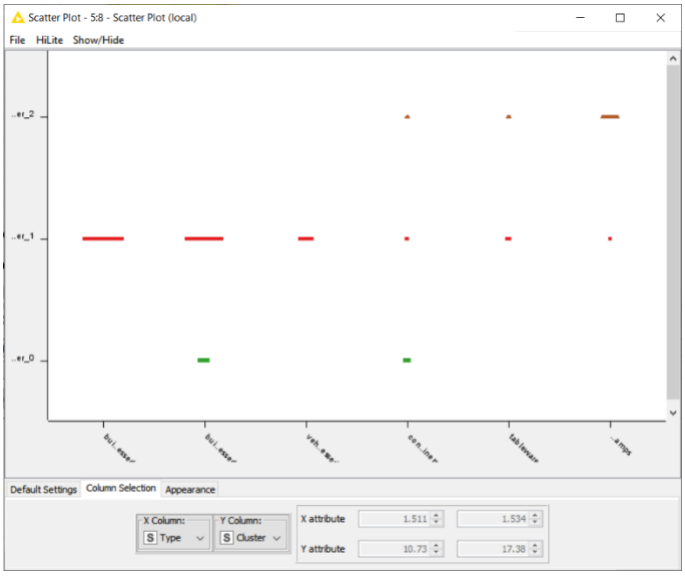
Node: Scorer (5:14)

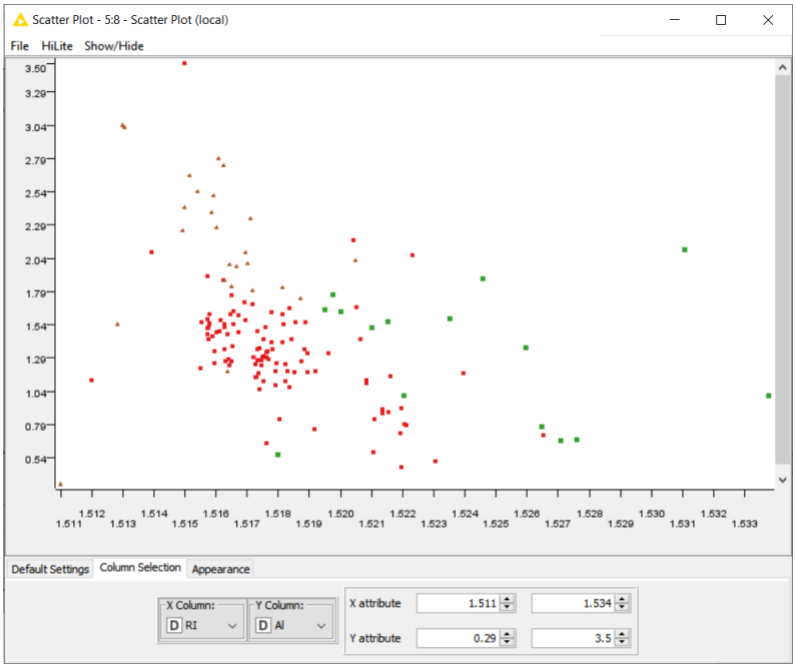
State: EXECUTED

Port Output Port 0 | Load data | Rows: 12, Columns: 12

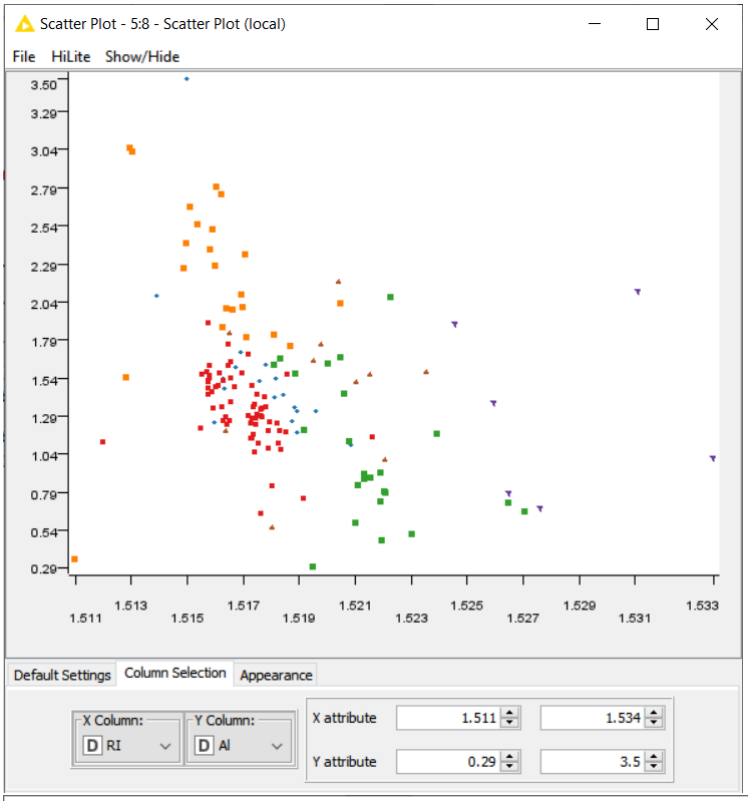
ID	cluster_0	cluster_1	cluster_2	cluster_3	cluster_4	cluster_5
building_windows_float_processed	11	29	0	0	5	0
building_windows_non_float_processed	5	30	2	6	9	0
vehicle_windows_float_processed	3	9	0	0	3	0
containers	0	0	7	0	1	2
tableware	4	0	0	0	0	3
headlamps	1	0	1	0	0	18

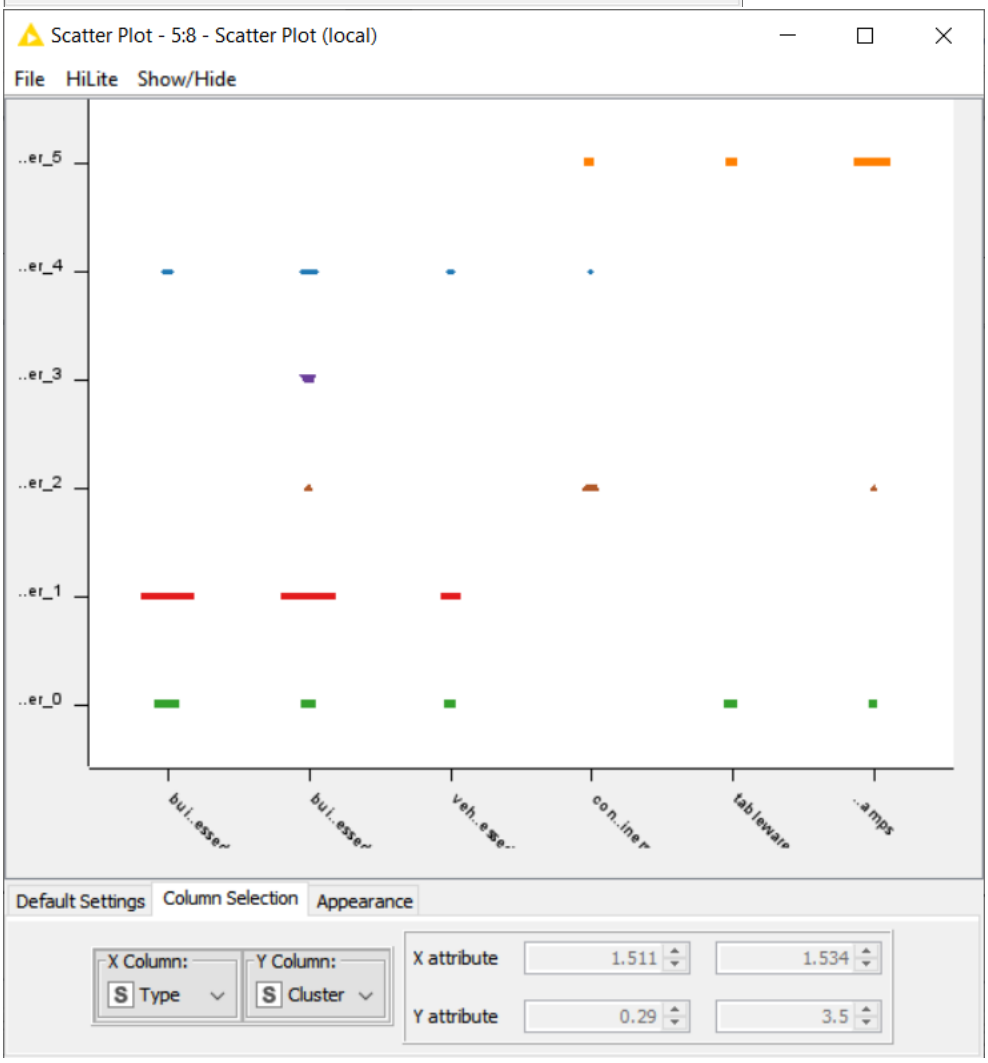
Old Scatter Plot:-



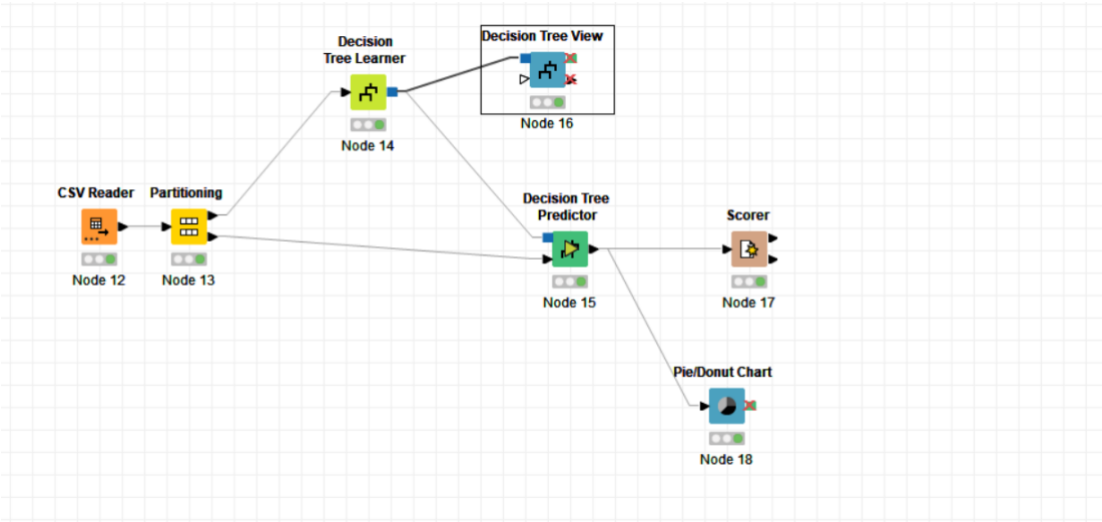


New scatter plot:-

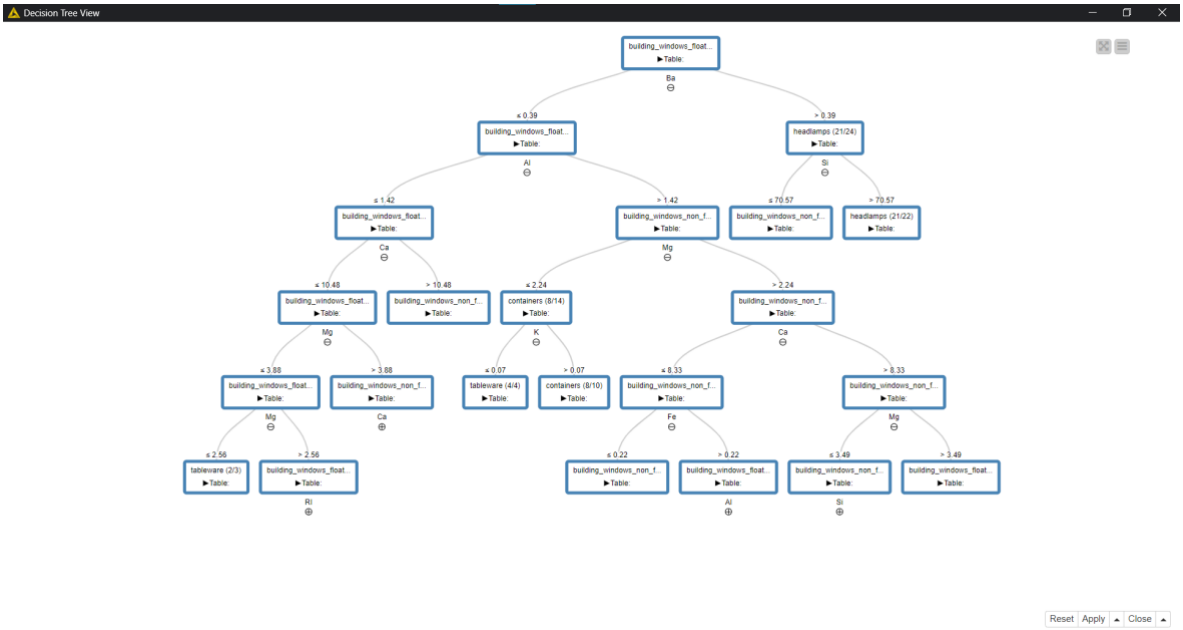

Scatter Plot - 5:8 - Scatter Plot (local)
File HiLite Show/Hide
3.50
3.29
3.04
2.79
2.54
2.29
2.04
1.79
1.54
1.29
1.04
0.79
0.54
0.29
1.511 1.513 1.515 1.517 1.519 1.521 1.523 1.525 1.527 1.529 1.531 1.533
Default Settings Column Selection Appearance
X Column: D RI
Y Column: D Al
X attribute 1.511 1.534
Y attribute 0.29 3.5


Scatter Plot - 5:8 - Scatter Plot (local)
File HiLite Show/Hide
..er_5
..er_4
..er_3
..er_2
..er_1
..er_0
bui..esse
bui..esse
veh..esse
con..iner
tablewar
..amps
Default Settings Column Selection Appearance
X Column: S Type
Y Column: S Cluster
X attribute 1.511 1.534
Y attribute 0.29 3.5

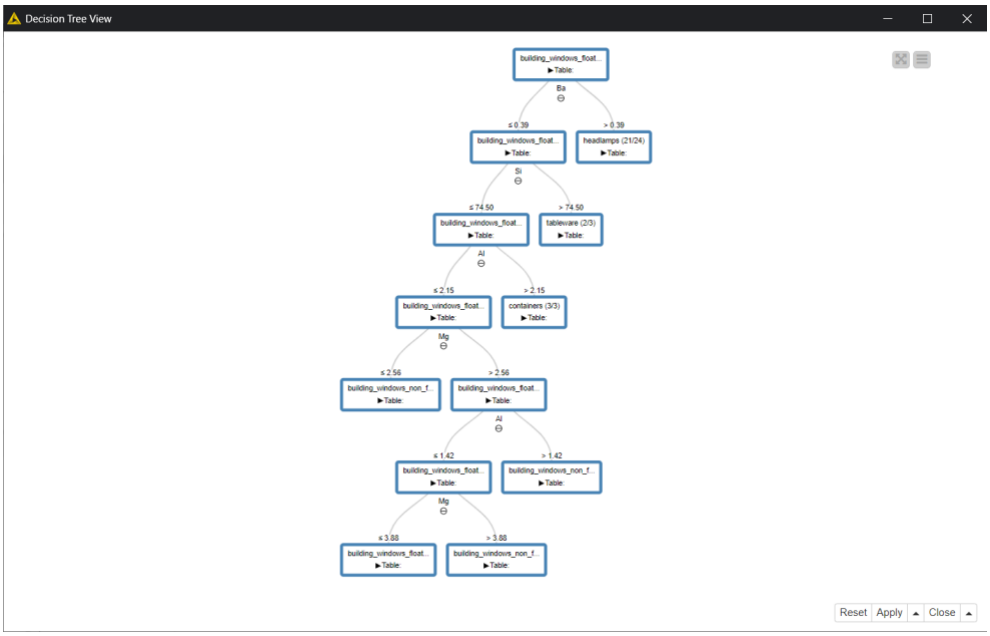
2. Classification Algorithms (Decision tree)



Old Desision tree view:-



Changing Pruning method to MDL, and quality measure to Gain ratio:-



Prediction value:-

Console | Node Monitor

Node: Decision Tree Predictor (6:15)

State: EXECUTED

Port Output Port 0 Load data Rows: 43, Columns: 11

ID	RI	Na	Mg	Al	Si	K	Ca	Ba	Fe	Type	Prediction (Type)
64	1.52227	14.17	3.81	0.78	71.35	0.0	9.69	0.0	0.0	building_windows_float_processed	building_windows_float_processed
72	1.51848	13.64	3.87	1.27	71.96	0.54	8.32	0.0	0.32	building_windows_non_float_processed	building_windows_non_float_processed
73	1.51593	13.09	3.59	1.52	73.1	0.67	7.83	0.0	0.0	building_windows_non_float_processed	building_windows_non_float_processed
75	1.51596	13.02	3.56	1.54	73.11	0.72	7.9	0.0	0.0	building_windows_non_float_processed	building_windows_non_float_processed
76	1.5159	13.02	3.58	1.51	73.12	0.69	7.96	0.0	0.0	building_windows_non_float_processed	building_windows_non_float_processed
79	1.51613	13.92	3.52	1.25	72.88	0.37	7.94	0.0	0.14	building_windows_non_float_processed	vehicle_windows_float_processed
82	1.51593	13.25	3.45	1.43	73.17	0.61	7.86	0.0	0.0	building_windows_non_float_processed	building_windows_non_float_processed
83	1.51646	13.41	3.55	1.25	72.81	0.68	8.1	0.0	0.0	building_windows_non_float_processed	vehicle_windows_float_processed
88	1.51645	13.4	3.49	1.52	72.65	0.67	8.08	0.0	0.1	building_windows_non_float_processed	building_windows_non_float_processed

Confusion matrix:-

Confusion Matrix - 6:17 - Scorer

File Hilite

Prediction (Type) \ Type	building_windows_float_processed	building_windows_non_float_processed	vehicle_windows_float_processed	containers	headlamps	tableware
building_windows_float_processed	7	2	1	0	0	0
building_windows_non_float_pro...	0	12	0	1	1	3
vehicle_windows_float_processed	0	3	4	0	0	0
containers	0	1	0	3	1	0
headlamps	0	0	0	0	4	0
tableware	0	0	0	0	0	0

Correct classified: 30 Wrong classified: 13

Accuracy: 69.767 % Error: 30.233 %

Cohen's kappa (κ) 0.599

New confusion matrix:-

Confusion Matrix - 6:17 - Scorer

File Hilite

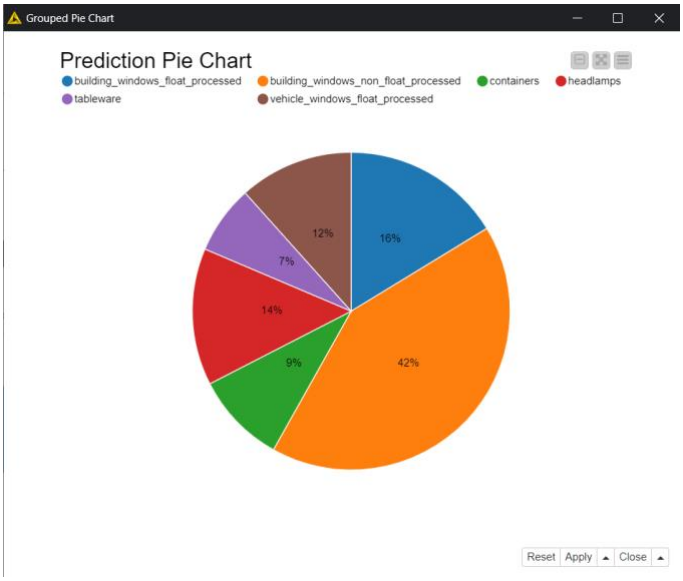
Prediction ...	building_windows_float...	building_windows_non_float_...	headlamps	vehicle_wi...	containers	tableware
building_win...	7	7	0	4	0	0
building_win...	0	11	1	1	4	3
headlamps	0	0	5	0	0	0
vehicle_win...	0	0	0	0	0	0
containers	0	0	0	0	0	0
tableware	0	0	0	0	0	0

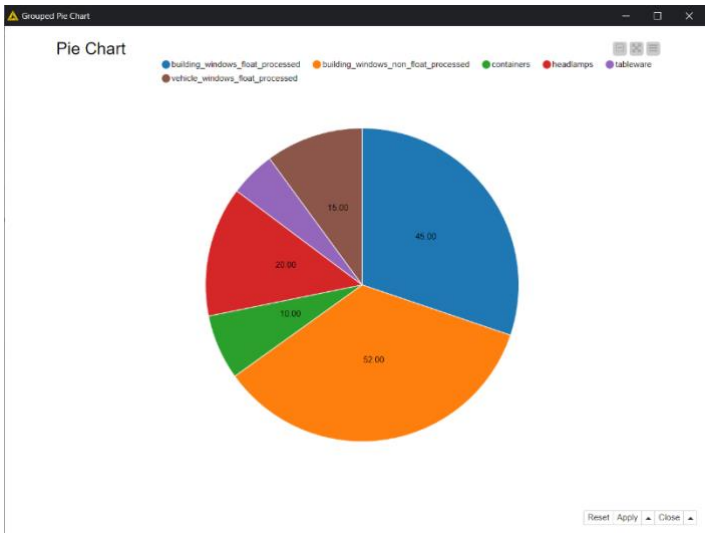
Correct classified: 23 Wrong classified: 20

Accuracy: 53.488 % Error: 46.512 %

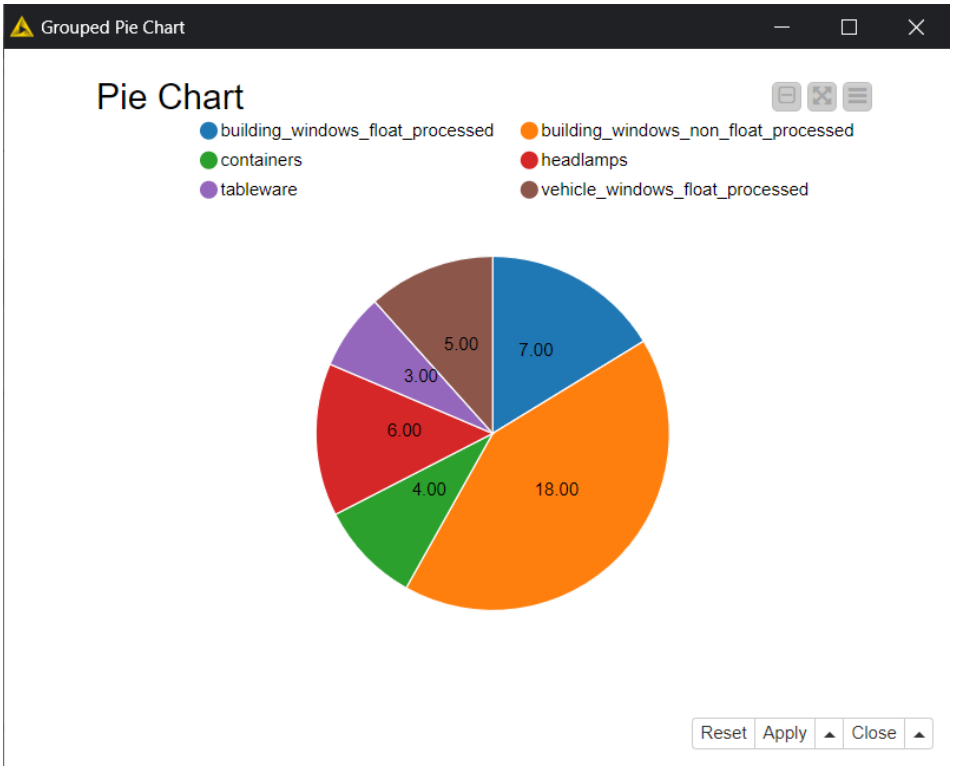
Cohen's kappa (κ) 0.355

Prediction pie chart vs normal pie chart:-

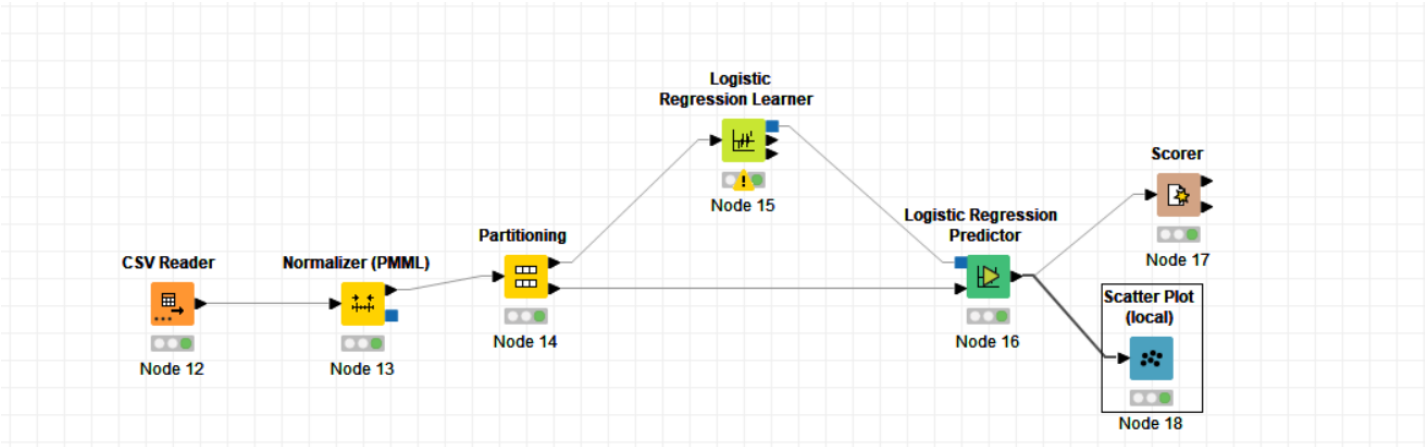




New pie chart:-



3. Logistic Regression



Confusion matrix:-

Confusion Matrix - 7:17 - Scorer

File Hilite

Type \ Pre...	building_windows_float_processed	building_windows_non_float_processed	vehicle_windows_float_processed	containers	tableware	headlamps
building_win...	10	5	1	0	0	0
building_win...	4	17	0	0	0	1
vehicle_win...	3	1	0	0	0	0
containers	0	1	0	2	0	1
tableware	0	0	0	0	1	0
headlamps	0	1	0	0	0	6

Correct classified: 36 Wrong classified: 18

Accuracy: 66.667 % Error: 33.333 %

Cohen's kappa (κ) 0.52

New confusion matrix on changing epsilon to 2*10^-5, Learning strategy = Line search and Prior = gauss

Confusion Matrix - 7:17 - Scorer

File Hilite

Type \ Pre...	building_wi...	building_wi...	vehicle_wi...	containers	tableware	headlamps
building_win...	12	4	0	0	0	0
building_win...	9	13	0	0	0	0
vehicle_win...	3	1	0	0	0	0
containers	0	3	0	0	0	1
tableware	0	0	0	0	0	1
headlamps	0	1	0	0	0	6

Correct classified: 31 Wrong classified: 23

Accuracy: 57.407 % Error: 42.593 %

Cohen's kappa (κ) 0.377

Normalized output:-

Console Node Monitor

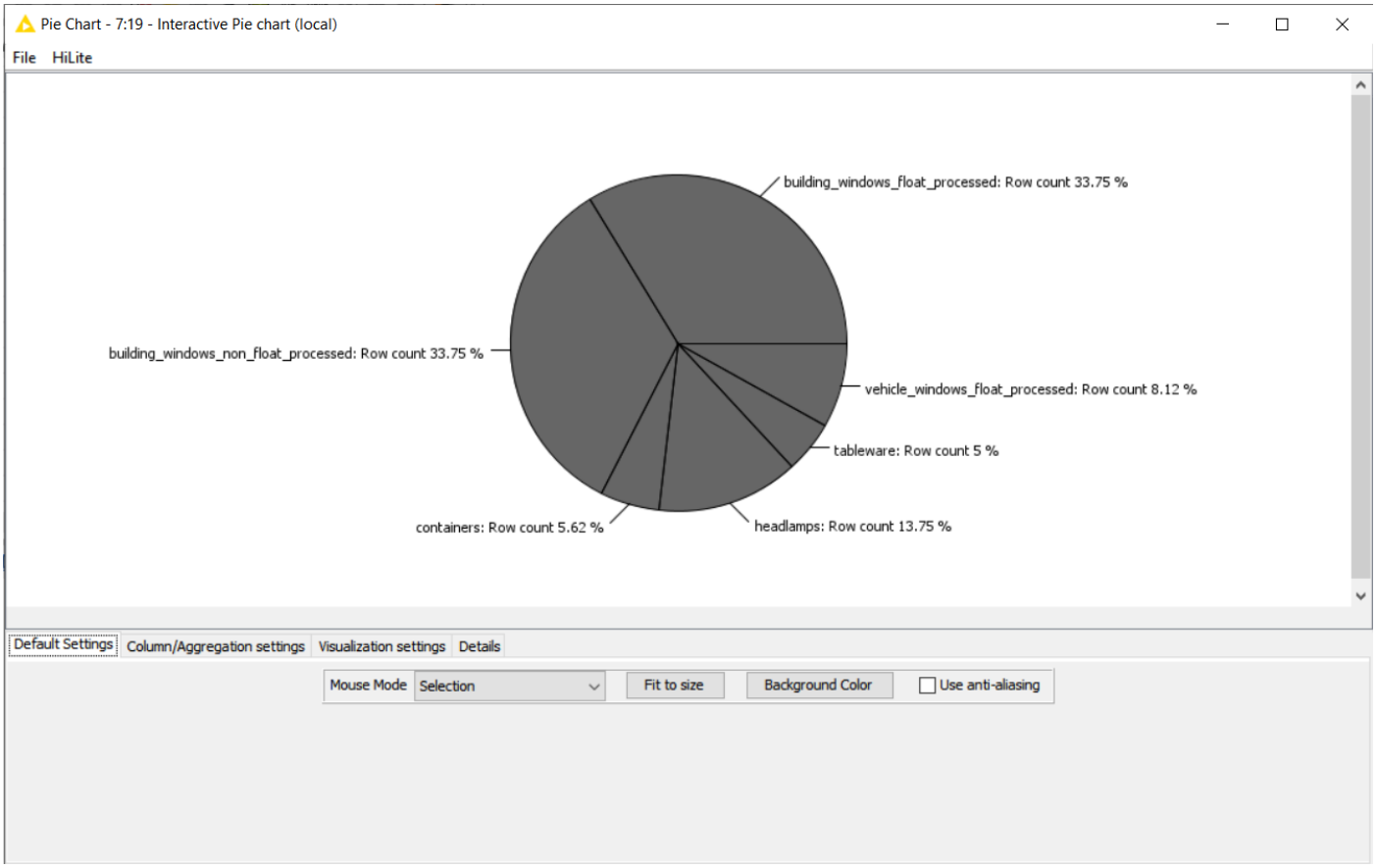
Node: Normalizer (PMML) (7:13)

State: EXECUTED

Port Output Port 0 Load data Rows: 214, Columns: 10

ID	RI	Na	Mg	Al	Si	K	Ca	Ba	Fe	Type
1	0.8708258...	0.28428670...	1.251703737...	-0.6908221...	-1.12444555935...	-0.670134221...	-0.145425366...	-0.3520513842541744	-0.5850790848193707	building_windows_float_processed
2	-0.248750...	0.59043281...	0.634679915...	-0.1700614...	0.102079723490...	-0.026151928...	-0.791877072...	-0.3520513842541744	-0.5850790848193707	building_windows_float_processed
3	-0.719630...	0.14958241...	0.600015656...	0.19046512...	0.437760327216...	-0.164148134...	-0.827010317...	-0.3520513842541744	-0.5850790848193707	building_windows_float_processed
4	-0.232285...	-0.2422846...	0.697075583...	-0.3102662...	-0.05284978592...	0.1118442766...	-0.517837761...	-0.3520513842541744	-0.5850790848193707	building_windows_float_processed
5	-0.311314...	-0.1688095...	0.648545619...	-0.4104125...	0.553957459275...	0.0811784532...	-0.623237496...	-0.3520513842541744	-0.5850790848193707	building_windows_float_processed
6	-0.792073...	-0.7566100...	0.641612767...	0.35069918...	0.411938742314...	0.2191746588...	-0.623237496...	-0.3520513842541744	2.083265200373234	building_windows_float_processed
7	-0.308021...	-0.1320720...	0.634679915...	-0.6107051...	0.566868251726...	0.1271771884...	-0.552971006...	-0.3520513842541744	-0.5850790848193707	building_windows_float_processed
8	0.265214...	0.2157506...	0.641612767...	0.7000684...	0.760530128401...	0.1118442766...	0.502784464...	-0.3520513842541744	-0.5850790848193707	building_windows_float_processed

Pie chart:-


Pie Chart - 7:19 - Interactive Pie chart (local)
File
HiLite
building_windows_float_processed: Row count 33.75 %
building_windows_non_float_processed: Row count 33.75 %
vehicle_windows_float_processed: Row count 8.12 %
tableware: Row count 5 %
containers: Row count 5.62 %
headlamps: Row count 13.75 %
Default Settings
Column/Aggregation settings
Visualization settings
Details
Mouse Mode
Selection
Fit to size
Background Color
Use anti-aliasing