



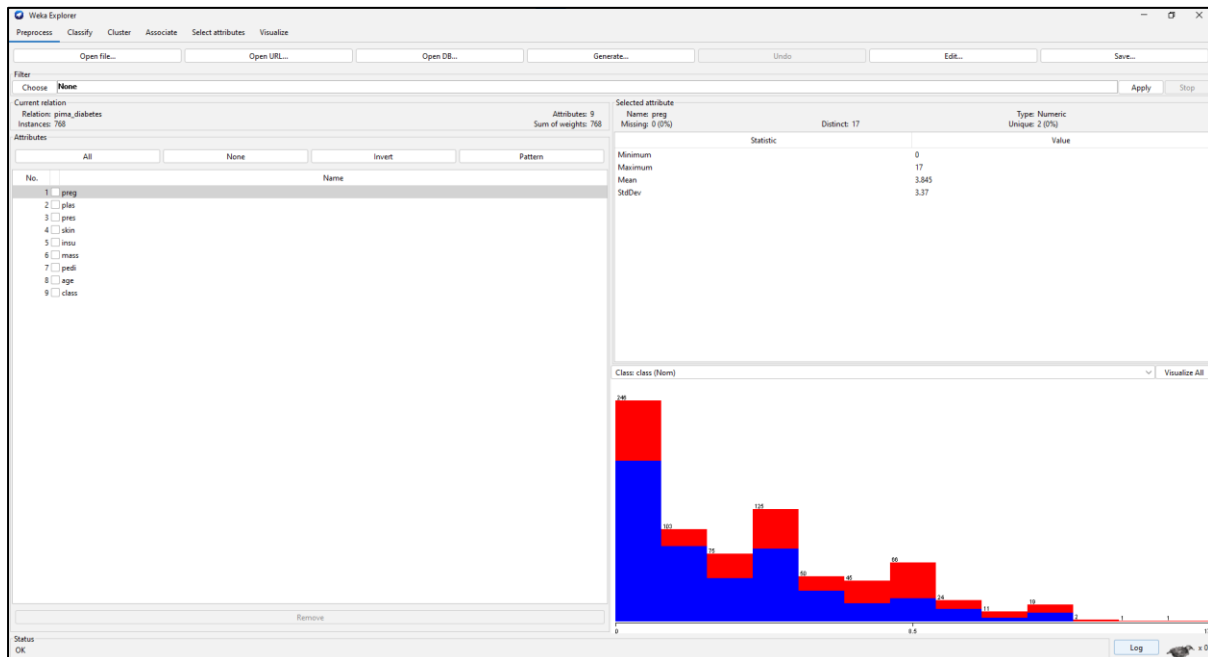
A.Y. 2022-2023

Subject: Data Mining and Warehousing

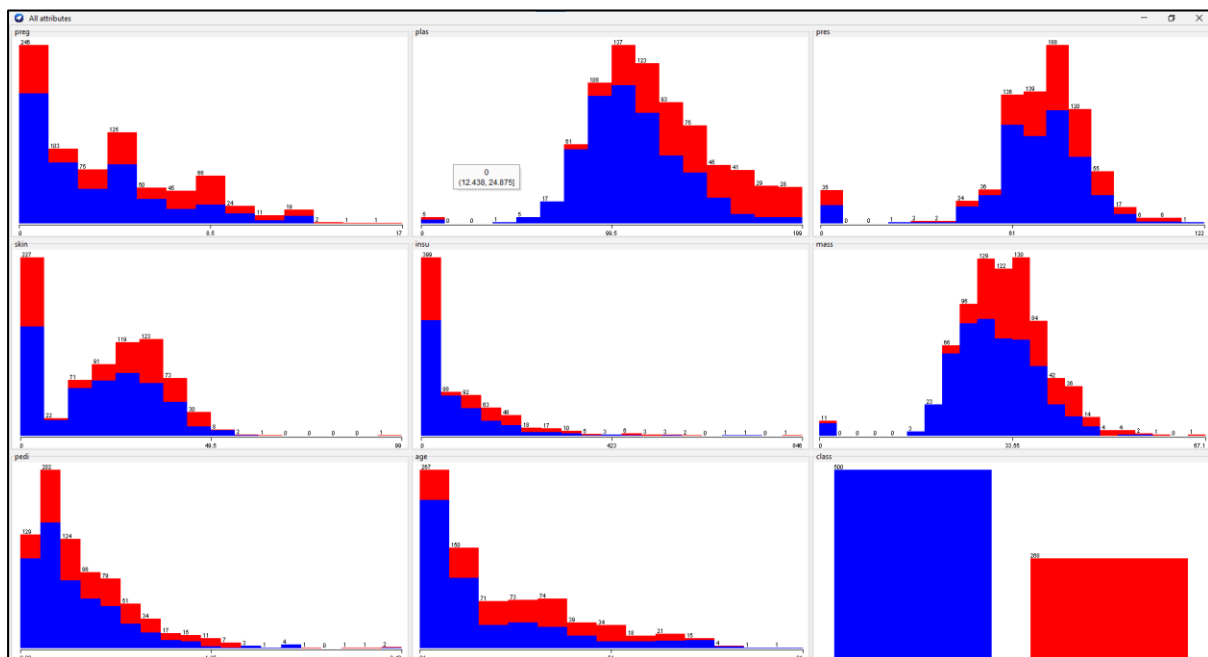
SAP ID: 60004220253 – Devansh Mehta

Experiment 01

Preprocess



Visualizing All Attributes



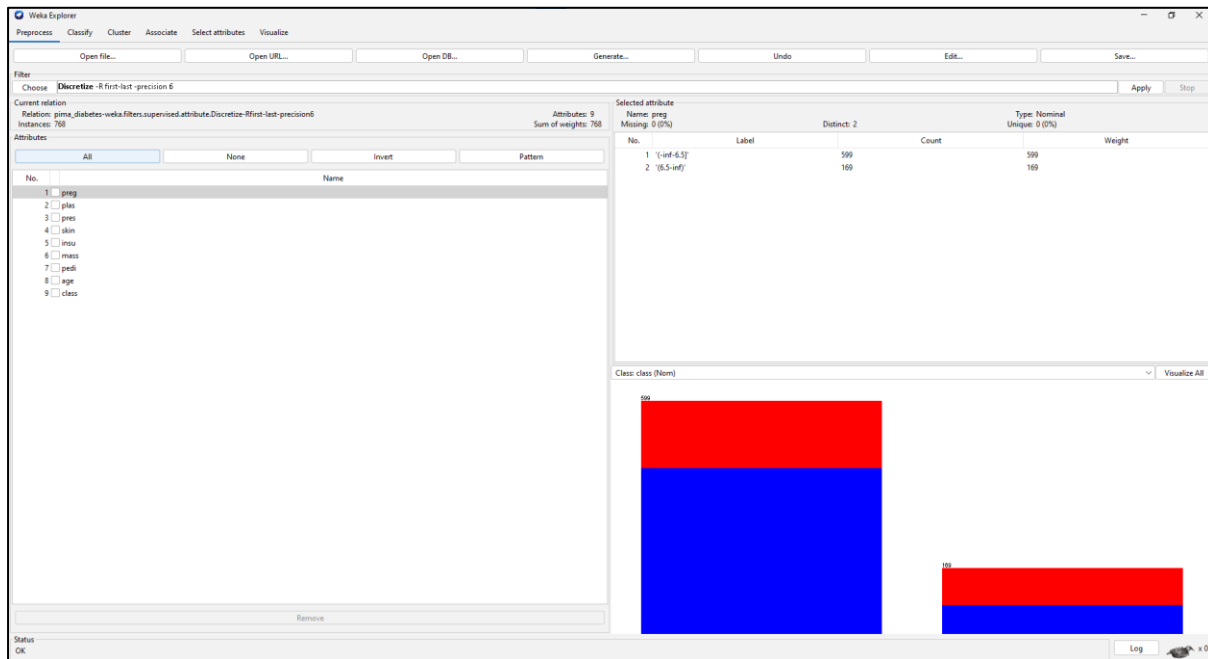


A.Y. 2022-2023

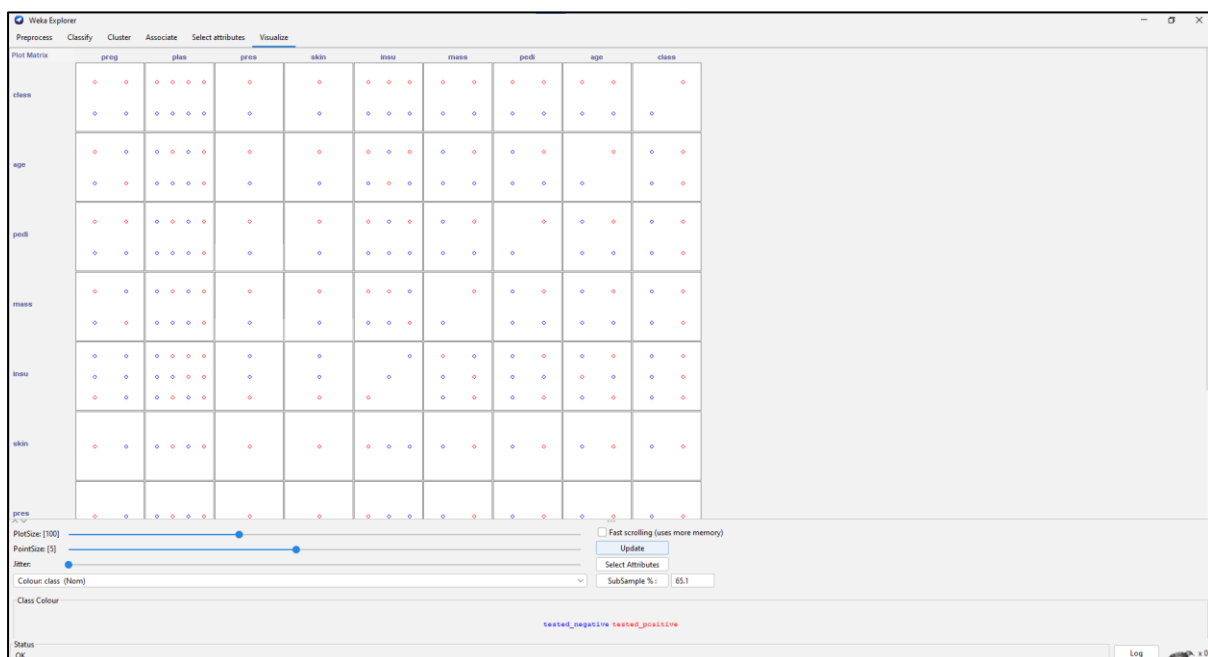
Subject: Data Mining and Warehousing

SAP ID: 60004220253 – Devansh Mehta

Applying filter – Discretize

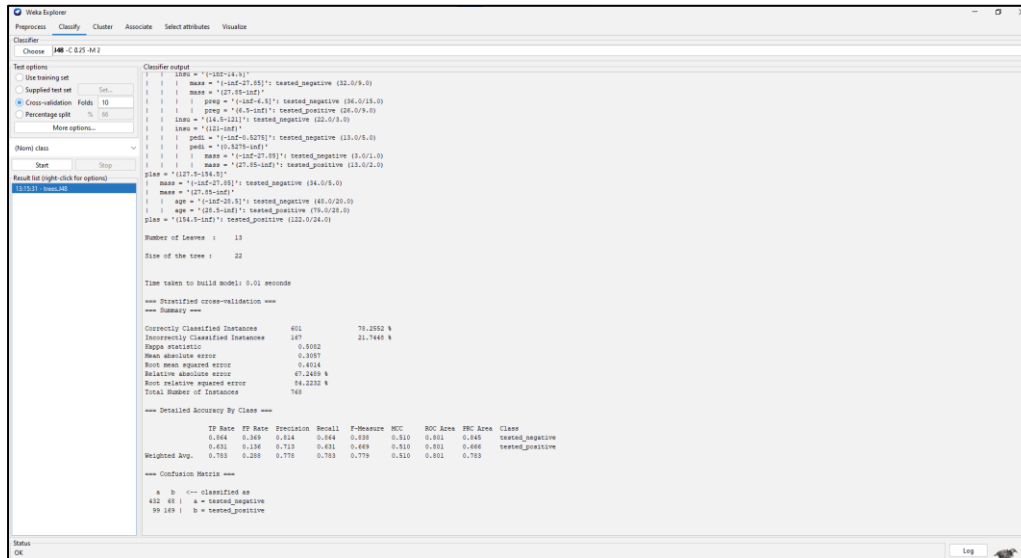


Visualize – Scatterplot Point size = 5

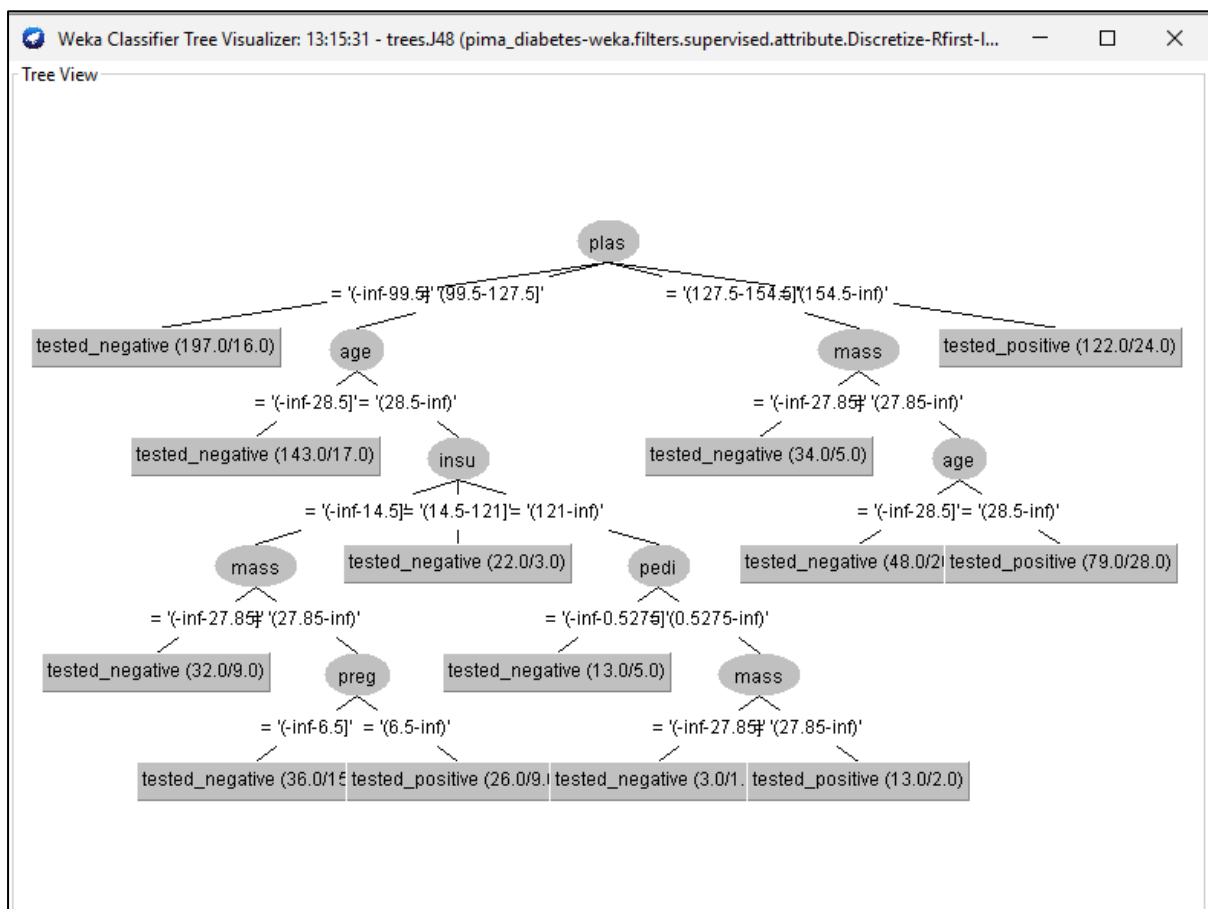


SAP ID: 60004220253 – Devansh Mehta

Classify – J48 Cross Validation



Tree



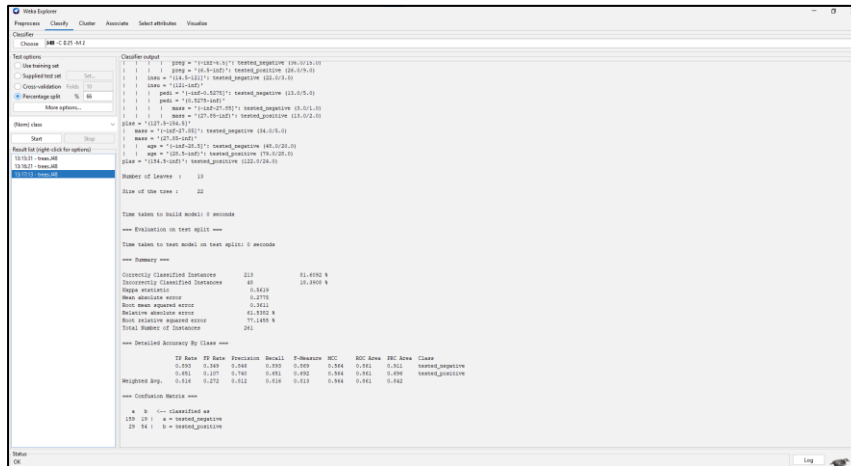


A.Y. 2022-2023

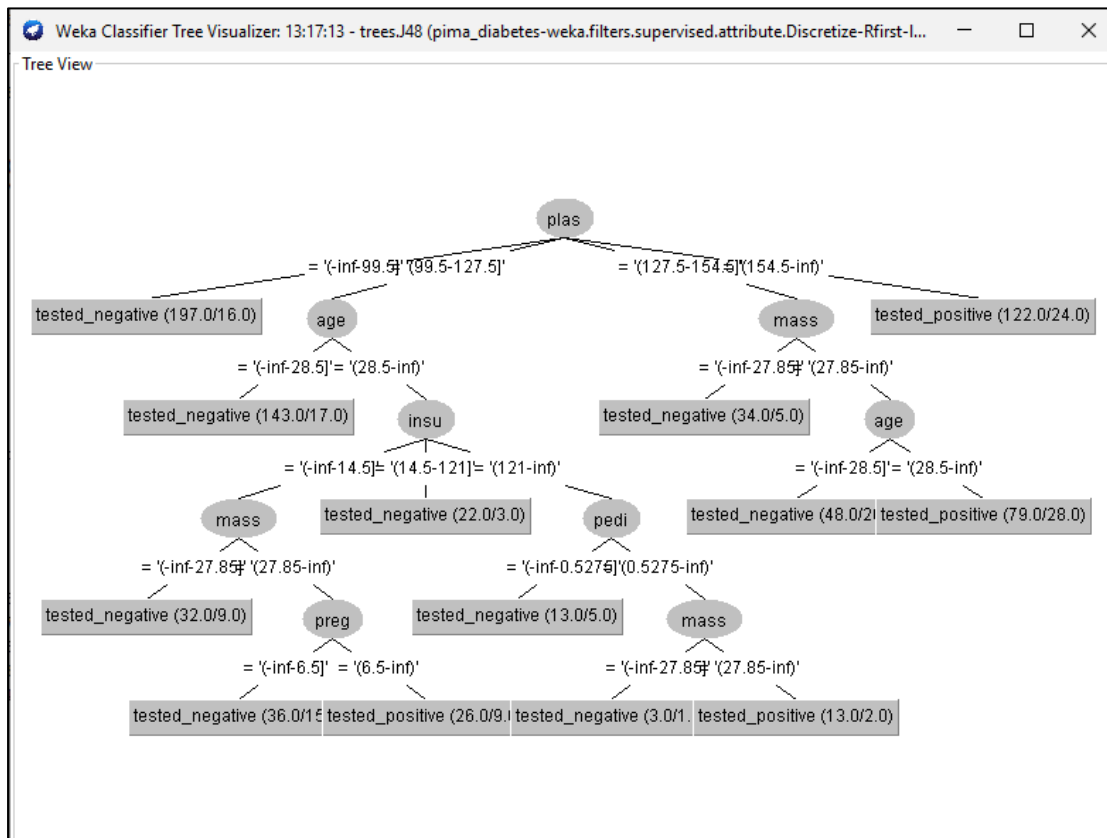
Subject: Data Mining and Warehousing

SAP ID: 60004220253 – Devansh Mehta

Classify – J48 Percentage Split



Tree





A.Y. 2022-2023

Subject: Data Mining and Warehousing

SAP ID: 60004220253 – Devansh Mehta

Classify - NaiveBayes - Cross Validation

Classifier
Choose: **NaiveBayes**

Test options
☐ Use training set
☐ Supplied test set
☒ Cross-validation Folds: 10
☐ Percentage split % 65
More options...

Classifier output

Class	Count	Percentage
131531 - trees.J48	237.0	145.0
131621 - trees.J48	145.0	25.0
131713 - trees.J48	101.0	103.0
[total]	503.0	271.0

mass

Class	Count	Percentage
131531 - trees.J48	196.0	25.0
131621 - trees.J48	306.0	242.0
[total]	502.0	270.0

Result Set (right-click for options)

- 131531 - trees.J48
- 131621 - trees.J48
- 131713 - trees.J48
- 131531 - trees.J48

Summary

Measure	Value
Correctly Classified Instances	598
Incorrectly Classified Instances	170
Kappa statistic	0.5128
Mean absolute error	0.2737
Root mean squared error	0.3905
Relative absolute error	60.2252 %
Root relative squared error	51.5274 %
Total Number of Instances	768

Detailed Accuracy By Class

TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	PRC Area	Class
0.830	0.217	0.830	0.830	0.830	0.513	0.846	tested_negative
0.483	0.170	0.483	0.483	0.483	0.513	0.846	tested_positive
Weighted Avg.	0.779	0.266	0.779	0.779	0.513	0.846	0.850

Confusion Matrix

a	b	classified as
415	85	a = tested_negative
85	183	b = tested_positive

Classify – NaiveBayes – Percentage Split

Classifier
Choose: **NaiveBayes**

Test options
☐ Use training set
☐ Supplied test set
☒ Cross-validation Folds: 10
☒ Percentage split % 80
More options...

Classifier output

Class	Count	Percentage
131531 - trees.J48	101.0	103.0
131621 - trees.J48	145.0	25.0
131713 - trees.J48	101.0	103.0
[total]	503.0	271.0

mass

Class	Count	Percentage
131531 - trees.J48	196.0	25.0
131621 - trees.J48	306.0	242.0
[total]	502.0	270.0

Result Set (right-click for options)

- 131531 - trees.J48
- 131621 - trees.J48
- 131713 - trees.J48
- 131531 - trees.J48

Summary

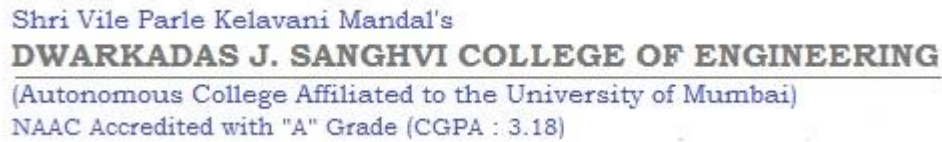
Measure	Value
Correctly Classified Instances	132
Incorrectly Classified Instances	22
Kappa statistic	0.6671
Mean absolute error	0.2352
Root mean squared error	0.3414
Relative absolute error	52.4936 %
Root relative squared error	73.0255 %
Total Number of Instances	154

Detailed Accuracy By Class

TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	PRC Area	Class
0.808	0.245	0.808	0.808	0.808	0.667	0.890	tested_negative
0.755	0.085	0.755	0.755	0.755	0.667	0.890	tested_positive
Weighted Avg.	0.857	0.197	0.856	0.857	0.856	0.667	0.890

Confusion Matrix

a	b	classified as
95	10	a = tested_negative
12	37	b = tested_positive

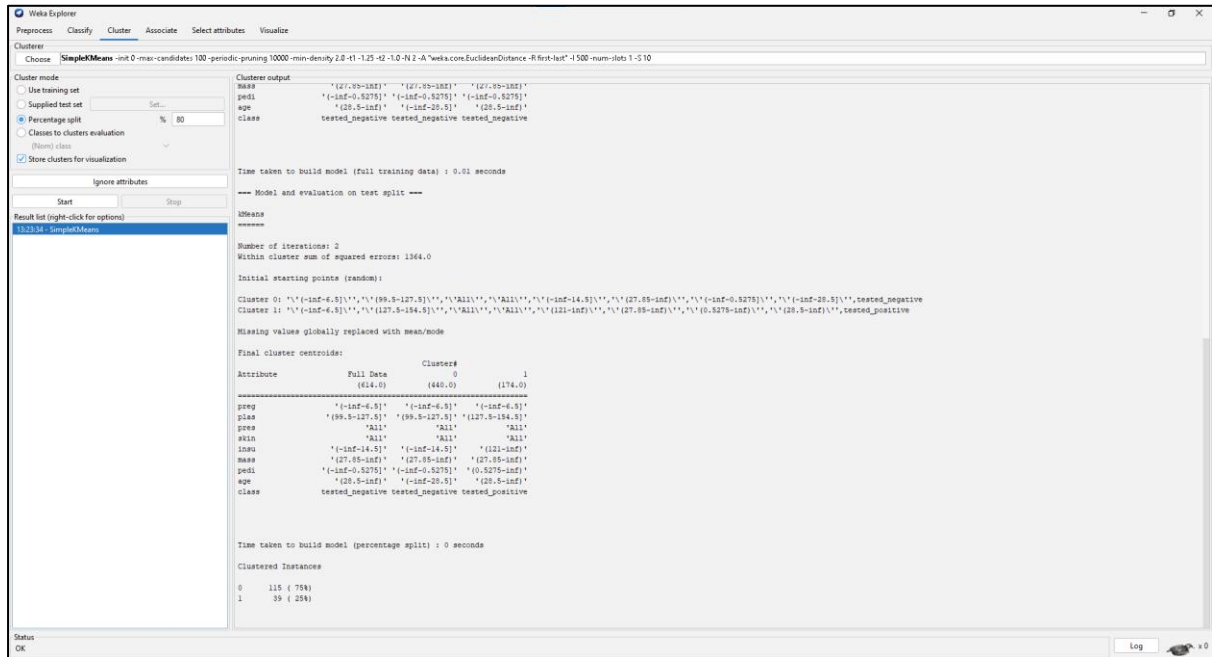


A.Y. 2022-2023

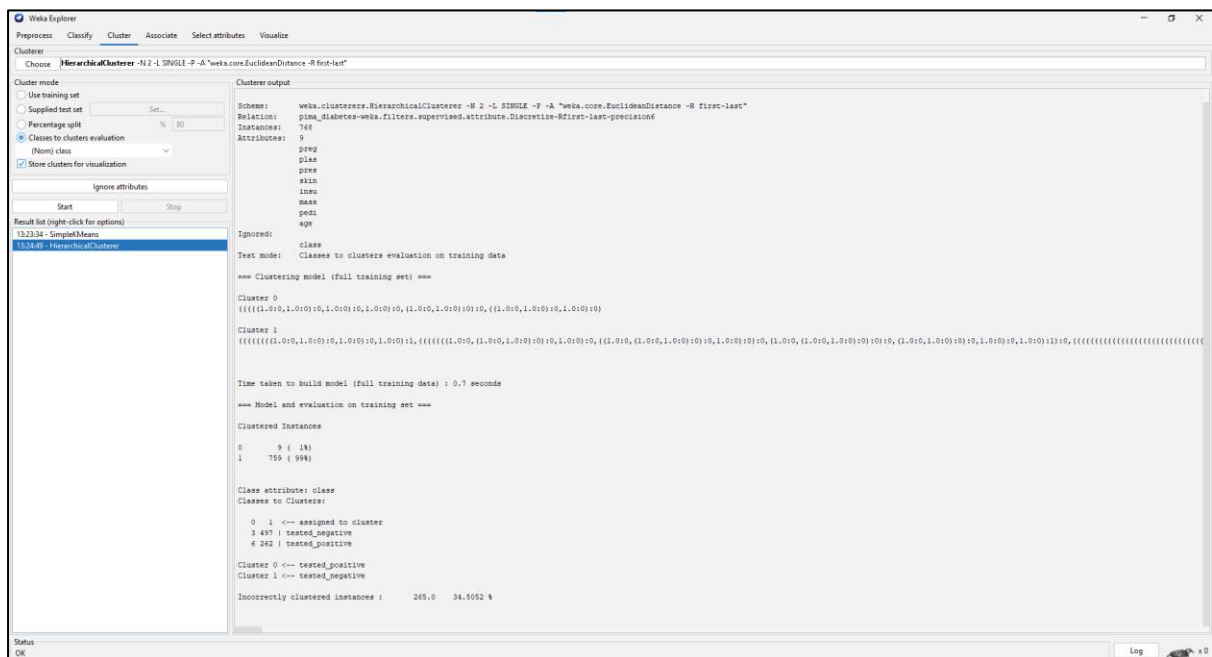
Subject: Data Mining and Warehousing

SAP ID: 60004220253 – Devansh Mehta

Cluster – SimpleKMeans



Cluster – Hierarchical Cluster



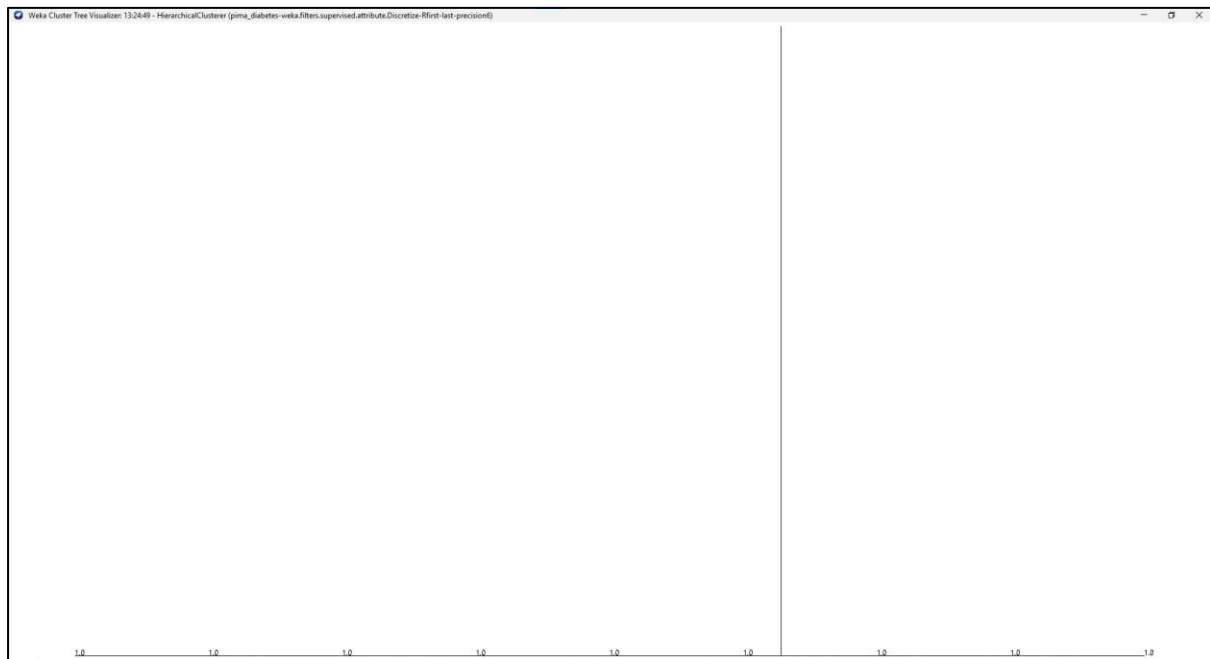


A.Y. 2022-2023

Subject: Data Mining and Warehousing

SAP ID: 60004220253 – Devansh Mehta

Tree



Associate

