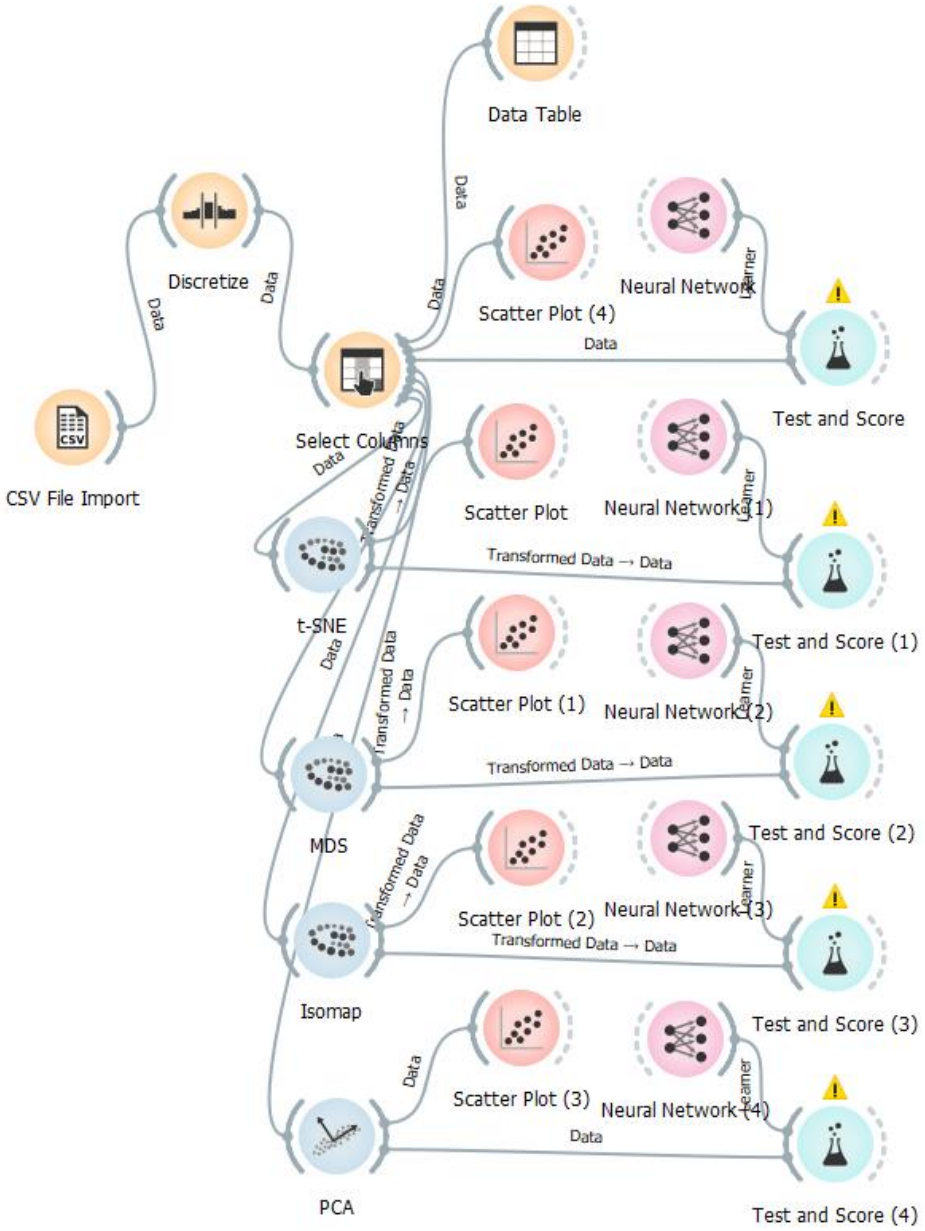
Anastasia Psarou - Assignment 6 – Machine Learning

This is my dataflow:



So, we are going to compare the values of Classification Accuracy in the different Test and Score widgets in order to find if these methods can help in obtaining better classification results by using NN.

Firstly, we are going to compare the CA values having a NN with 100 neurons in hidden layers.

Feature Selection:

Model	Train time [s]	Test time [s]	AUC	CA	F1	Precision	Recall	LogLoss	Specificity
Neural Network	26.420	0.130	0.993	0.905	0.904	0.904	0.905		0.989

t-SNE

Model	Train time [s]	Test time [s]	AUC	CA	F1	Precision	Recall	LogLoss	Specificity
Neural Network	95.591	0.060	0.981	0.911	0.911	0.912	0.911		0.990

MDS

Model	Train time [s]	Test time [s]	AUC	CA	F1	Precision	Recall	LogLoss	Specificity
Neural Network	41.642	0.071	0.990	0.879	0.879	0.879	0.879		0.986

Isomap

Model	Train time [s]	Test time [s]	AUC	CA	F1	Precision	Recall	LogLoss	Specificity
Neural Network	37.828	0.069	0.990	0.903	0.903	0.904	0.903		0.989

PCA

Model	Train time [s]	Test time [s]	AUC	CA	F1	Precision	Recall	LogLoss	Specificity
Neural Network	25.103	0.139	0.993	0.905	0.904	0.904	0.905		0.989

From these results we can observe that t-SNE is the method that generates the best CA for 100 neurons in hidden layers. Also, method t-SNE is the only one that produces results better compared to do not having feature engineering algorithms.

Now we are gone use 50 neurons in hidden layers and see if we have any change in the results.

Feature Selection:

Model	Train time [s]	Test time [s]	AUC	CA	F1	Precision	Recall	LogLoss	Specificity
Neural Network	24.295	0.100	0.992	0.888	0.887	0.887	0.888		0.987

t-SNE

Model	Train time [s]	Test time [s]	AUC	CA	F1	Precision	Recall	LogLoss	Specificity
Neural Network	46.313	0.034	0.980	0.907	0.907	0.908	0.907		0.990

MDS

Model	Train time [s]	Test time [s]	AUC	CA	F1	Precision	Recall	LogLoss	Specificity
Neural Network	23.077	0.048	0.990	0.876	0.876	0.876	0.876		0.986

Isomap

Model	Train time [s]	Test time [s]	AUC	CA	F1	Precision	Recall	LogLoss	Specificity
Neural Network	23.024	0.048	0.990	0.901	0.901	0.902	0.901		0.989

PCA

Model	Train time [s]	Test time [s]	AUC	CA	F1	Precision	Recall	LogLoss	Specificity
Neural Network	23.911	0.103	0.992	0.888	0.887	0.887	0.888		0.987

From these results we can observe that t-SNE is the method that generates the best CA for 50 neurons in hidden layers. Also, methods t-SNE and Isomap are the ones that produce results better compared to do not having feature engineering algorithms.

Now we are gone use 200 neurons in hidden layers and see if we have any change in the results.

Feature Selection:

Model	Train time [s]	Test time [s]	AUC	CA	F1	Precision	Recall	LogLoss	Specificity
Neural Network	30.652	0.136	0.994	0.908	0.907	0.908	0.908		0.990

t-SNE

Model	Train time [s]	Test time [s]	AUC	CA	F1	Precision	Recall	LogLoss	Specificity
Neural Network	141.176	0.072	0.981	0.914	0.913	0.914	0.914		0.990

MDS

Model	Train time [s]	Test time [s]	AUC	CA	F1	Precision	Recall	LogLoss	Specificity
Neural Network	75.366	0.107	0.990	0.883	0.883	0.883	0.883		0.987

Isomap

Model	Train time [s]	Test time [s]	AUC	CA	F1	Precision	Recall	LogLoss	Specificity
Neural Network	76.125	0.096	0.990	0.903	0.903	0.904	0.903		0.989

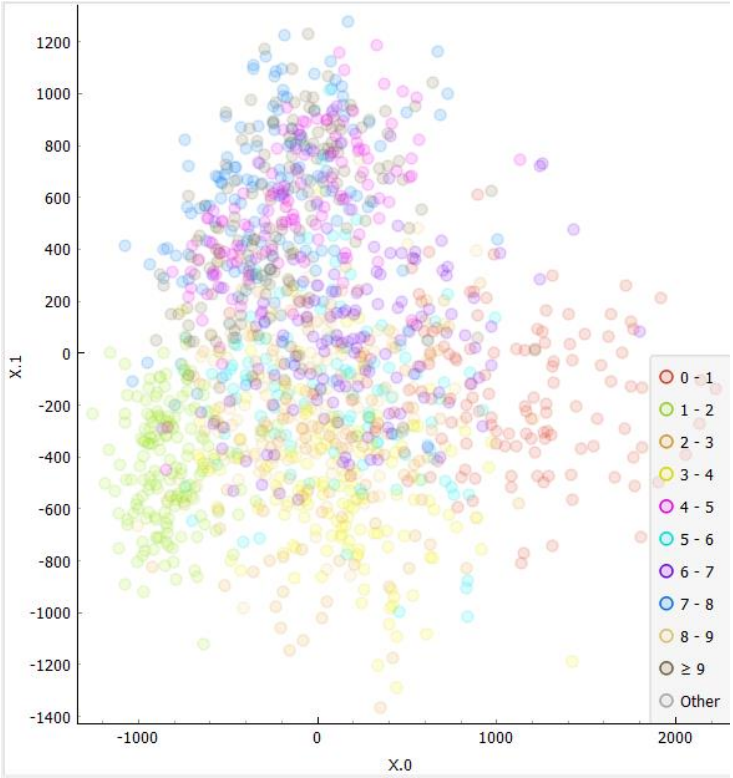
PCA

Model	Train time [s]	Test time [s]	AUC	CA	F1	Precision	Recall	LogLoss	Specificity
Neural Network	33.769	0.160	0.994	0.908	0.907	0.908	0.908		0.990

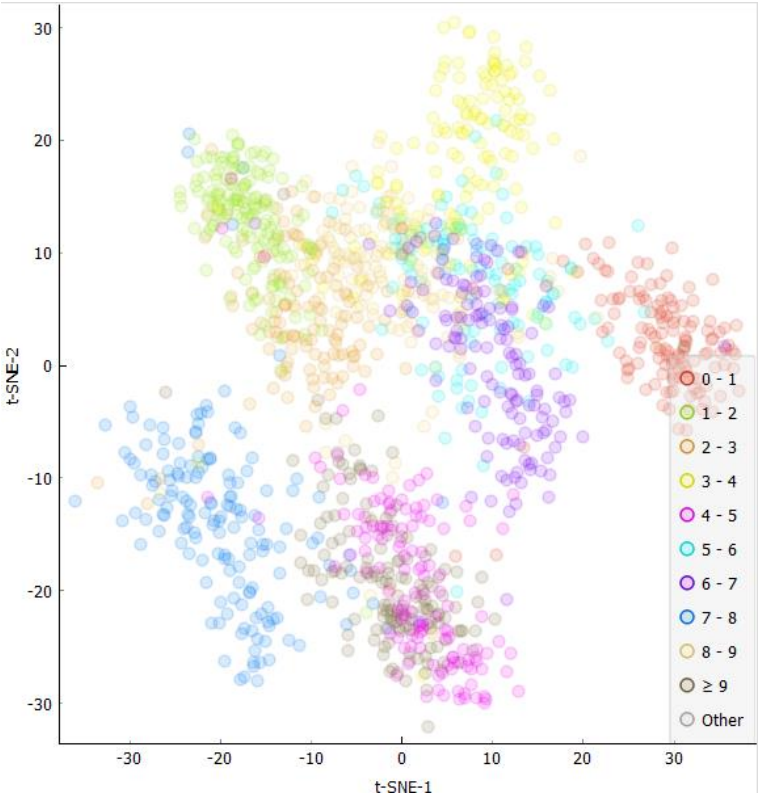
From these results we can observe that t-SNE is the method that generates the best CA for 200 neurons in hidden layers. Also, only method t-SNE is the one that produces results better compared to do not having feature engineering algorithms.

Lastly, we are going to take a look at the scatter plots.

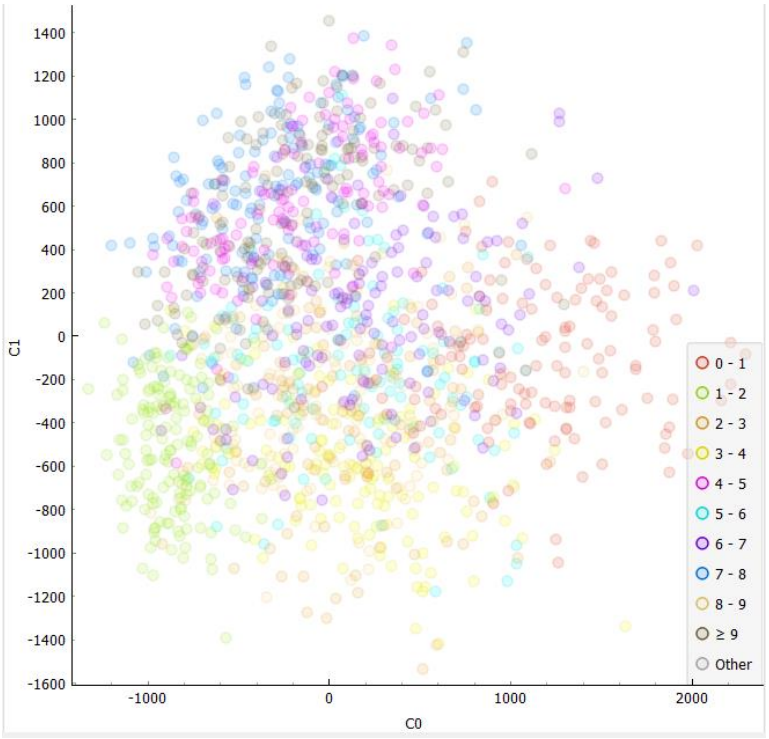
Feature Selection:



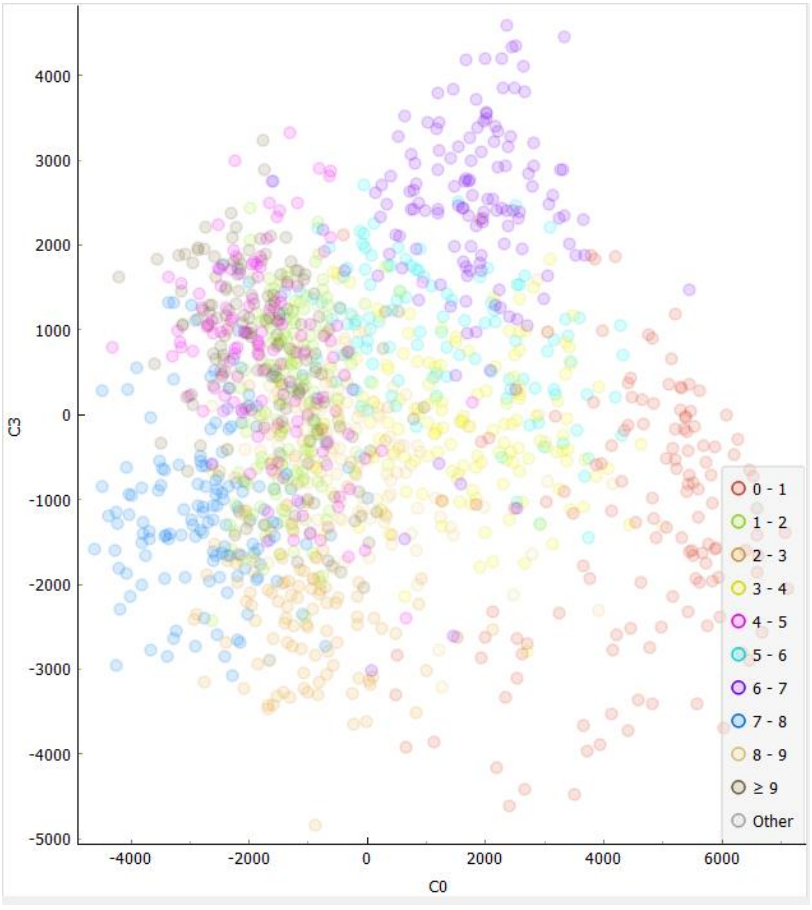
t-SNE



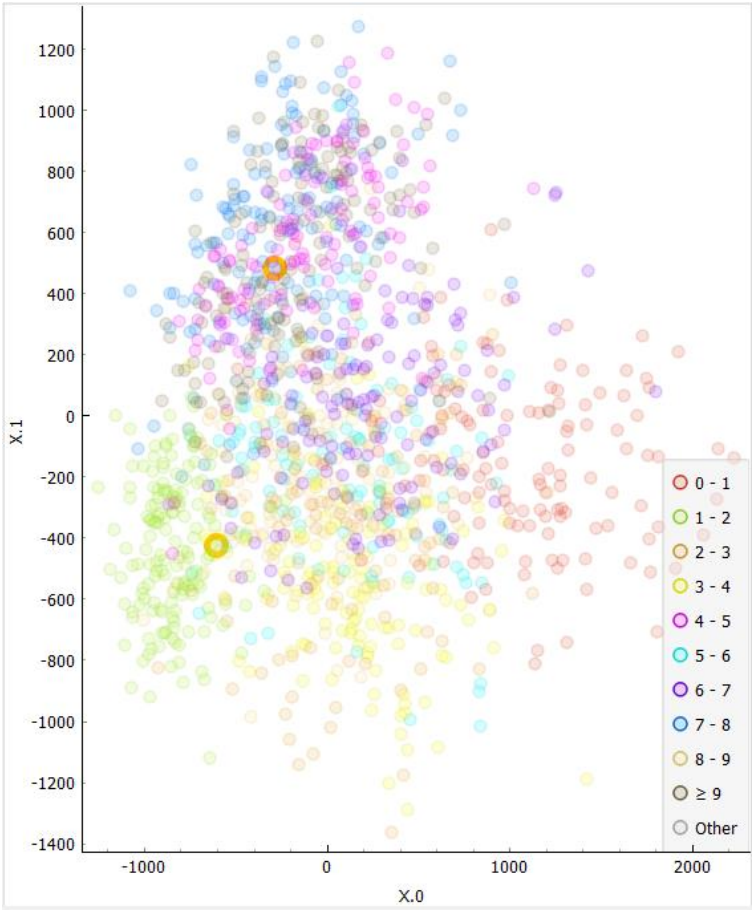
MDS


1400
1200
1000
800
600
400
200
0
-200
-400
-600
-800
-1000
-1200
-1400
-1600
C1
-1000
0
1000
2000
C0
0 - 1
1 - 2
2 - 3
3 - 4
4 - 5
5 - 6
6 - 7
7 - 8
8 - 9
≥ 9
Other

Isomap


4000
3000
2000
1000
0
-1000
-2000
-3000
-4000
-5000
C3
-4000
-2000
0
2000
4000
6000
C0
0 - 1
1 - 2
2 - 3
3 - 4
4 - 5
5 - 6
6 - 7
7 - 8
8 - 9
≥ 9
Other

PCA



From these Scatter Plots we can see that t-SNE is the feature engineering algorithm that has the more discrete values, compared to PCA, Isomap, MDS.

Finally, we should point out that PCA algorithm seems a good algorithm as far as classification is concerned as in the above experiments it generates the same CA value as not using a feature engineering algorithm. Although, t-SNE produces even better results so it is the only algorithm from above that can help in obtaining better classification results by using NN.