# Introduction

The following report contains the work of the orange data mining tool. With the help of these data mining tools, data analysis and visualisation are done based on the two datasets, iris and sailing. The two datasets are imported in the orange file. The visualisation and data analysis work is done by the use of the two models that is neural networking and random forest, and the report contains the result of the following model inside.

# Scaling dataset visualisation

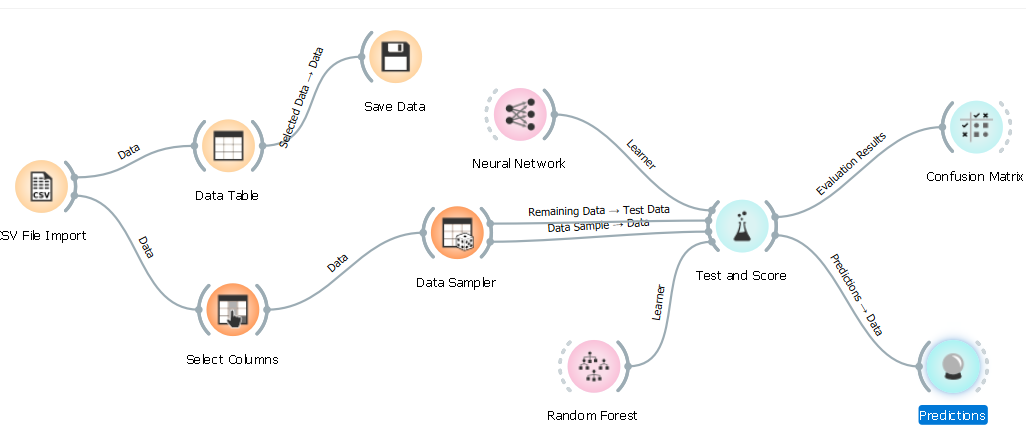


Figure 1 Orange model for scaling dataset

This is the orange data mining model for the dataset scaling. The orange data mining model contains the select column, data table, data sampler, random forest model, neural network, test and score, confusion matrix and the prediction models and tools inside it.

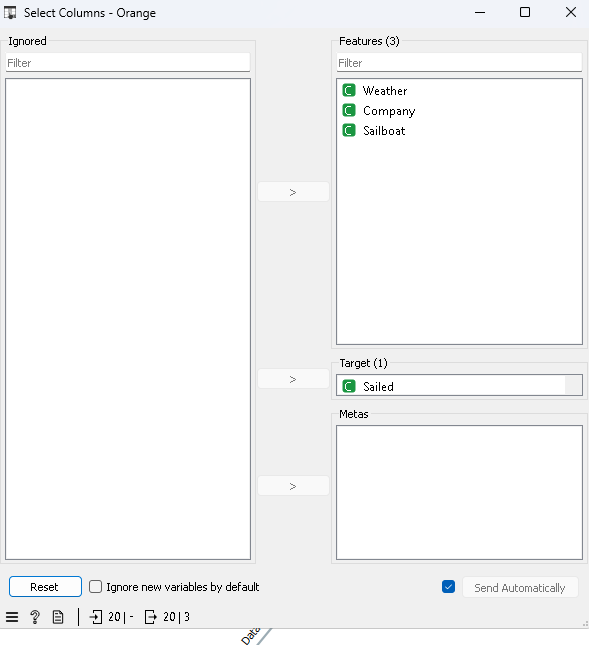


Figure 2 Select column

The select column contains the various attributes inside it like weather, company, sailboat and sailed and in them, sailed is taken in the targeted variable, and others are taken in the features column.

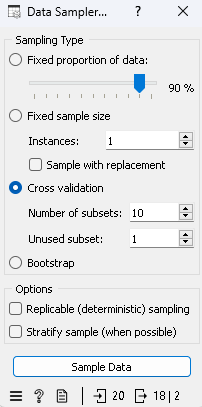


Figure 3 Data sampler

The cross-validation is the selected column in the data sampler. In it, the number of subsets Is ten, and the unused subset is 1.

## Testing and prediction

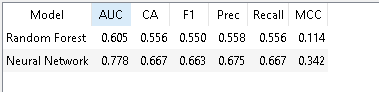


Figure 4 test and score for the models

The test and score shown in the models are as follows. The precision score for the random forest is 0.558, and for neural networking is 0.675. the CA value is 0.556 for random forest, and the Neural networking is 0.667. The AUC is 0.605 and 0.778 for the random forest and the neural network.

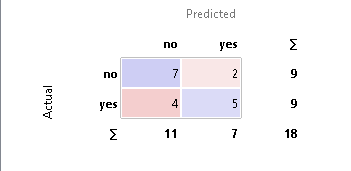


Figure 5 Confusion matrix

The confusion matrix shows the total result is 18, where the complete yes is 5, the complete no is 7, and the other is 6.

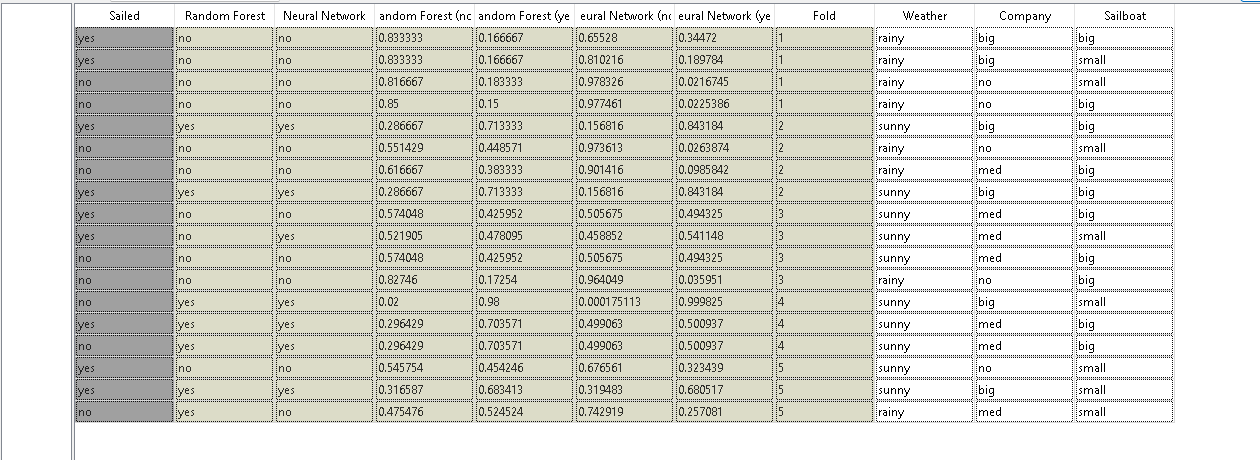


Figure 6 Prediction of the scaling

This is the prediction done by the following model on the sailing dataset.

From these results, it can be noticed that the neural network gives more precise results than the random forest, as the neural network is more effective in the bulky dataset, which is the scaling dataset.

# Iris dataset visualisation

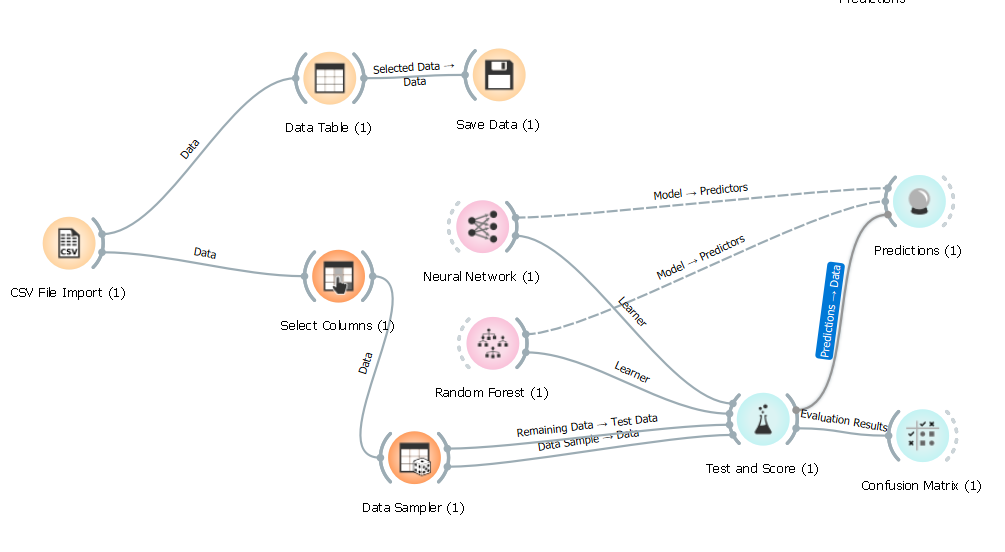


Figure 7 orange data mining model for iris dataset

This is the orange data mining model for the dataset iris. These models take the same tools and models for the visualisation and evaluation of the dataset.

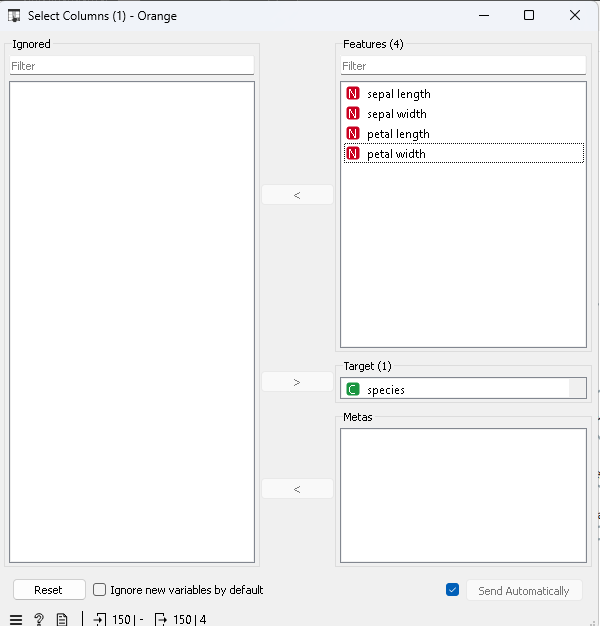


Figure 8 Select the column for the iris.

The select column contains the various attributes inside it like sepal length and width and species, and in them, species are taken in the targeted variable, and others are taken in the features column.

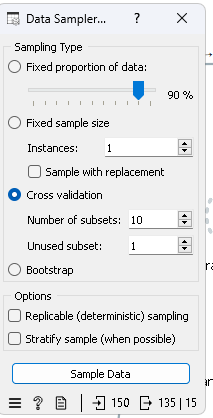


Figure 9 Data sampler

The cross-validation is the selected column in the data sampler. In it, the number of subsets Is ten, and the unused subset is 1.

## Testing and prediction

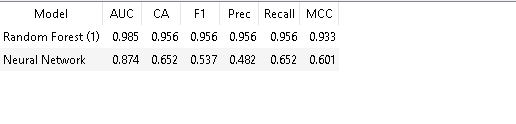


Figure 10 test and score of the models

The test and score shown in the models are as follows. The precision score for the random forest is 0.956, and for neural networking is 0.537. the CA value is 0.956 for random forest, and the Neural networking is 0.652. The AUC is 0.985 and 0.874 for the random forest and the neural network.

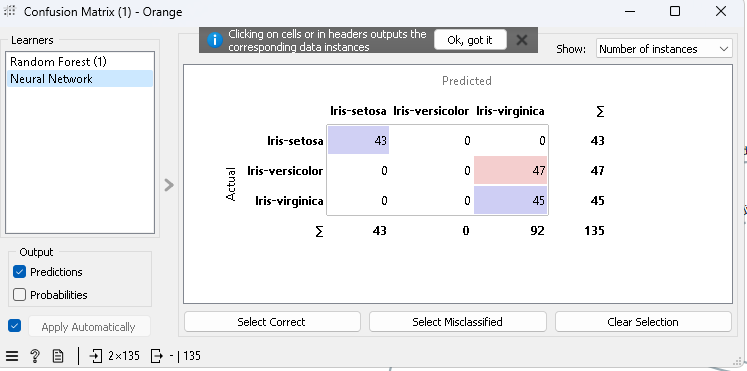


Figure 11 Confusion matrix

The confusion matrix shows the total result is 135, where the iris-setosa is 43, kris-versicolor is 47, and iris-virginica is 45.

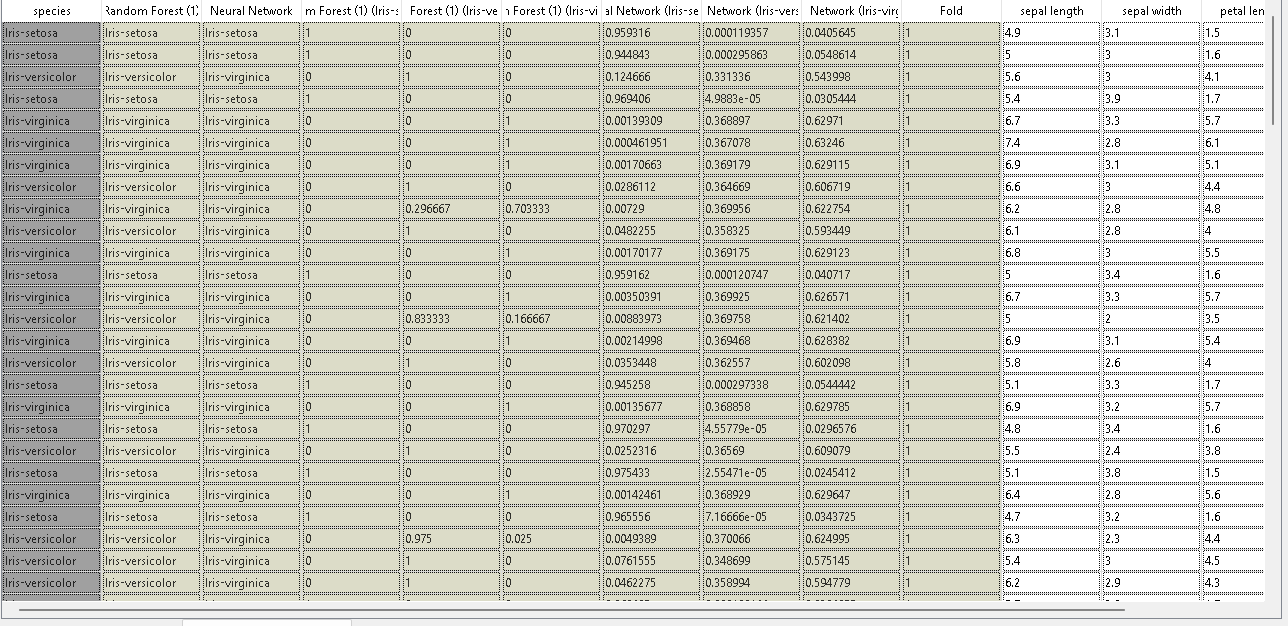


Figure 12 prediction

This is the prediction done by the following model on the iris dataset.

From these results, it can be noticed that the neural network gives less precise results than the random forest, as the neural network is more effective in the bulky dataset, which is not the iris dataset.