Ex:14 Tool Study

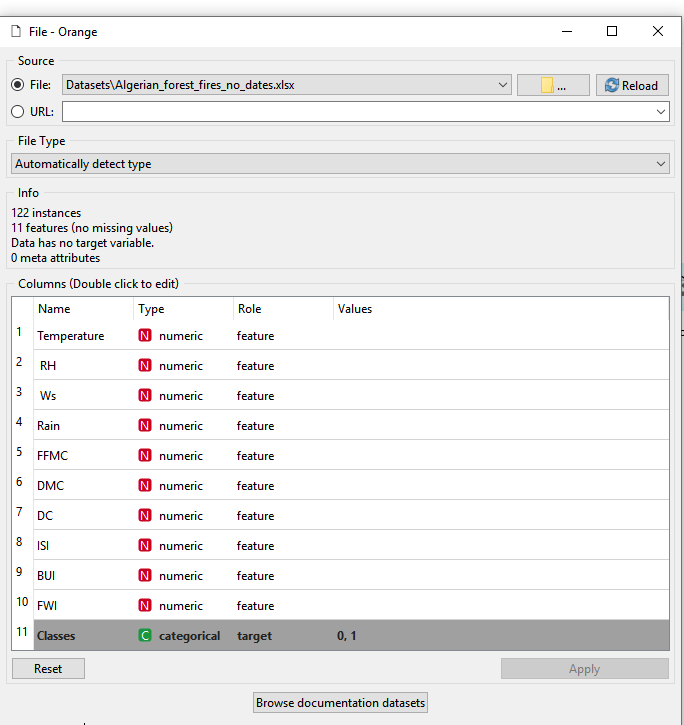
Perform comparative analysis for predicting outcome of variable using atleast 4 different types of classifiers and 3 different types dataset, make a comparative analysis using the tool orange.

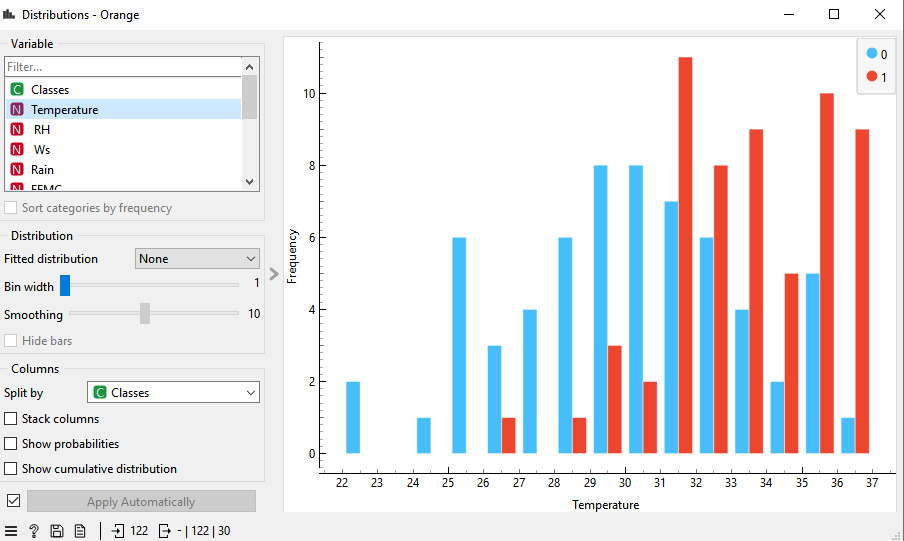
Name:Athithraja R

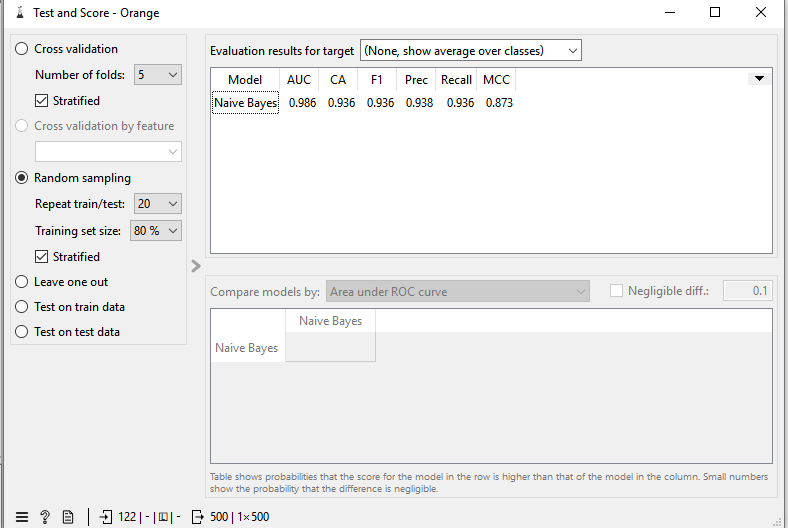
Reg.on:2022503702

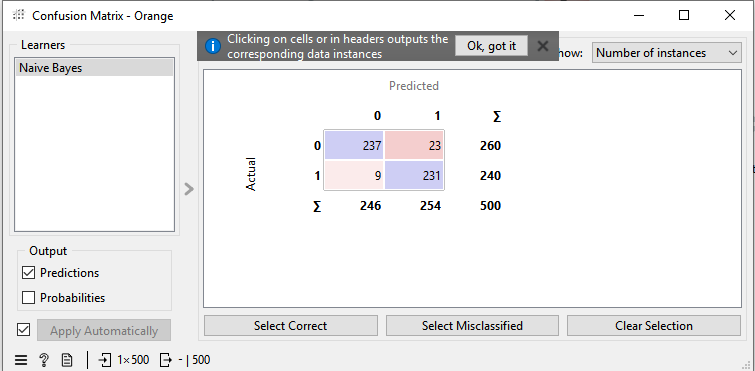
**NaivebayesClassification**

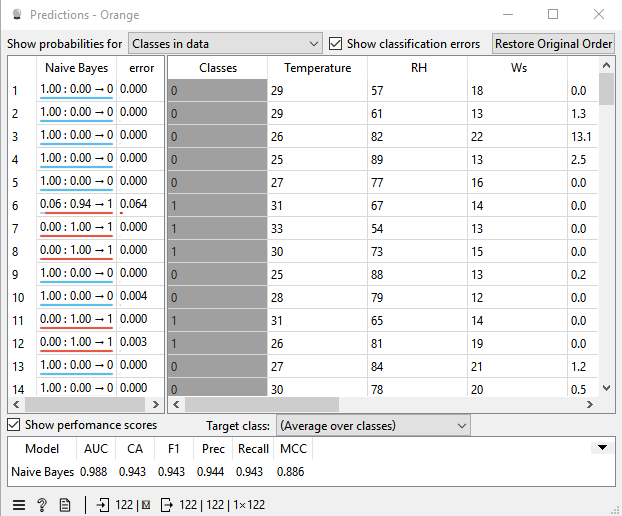
Dataset:Algerian\_forest\_fires\_no\_dates.xlsx



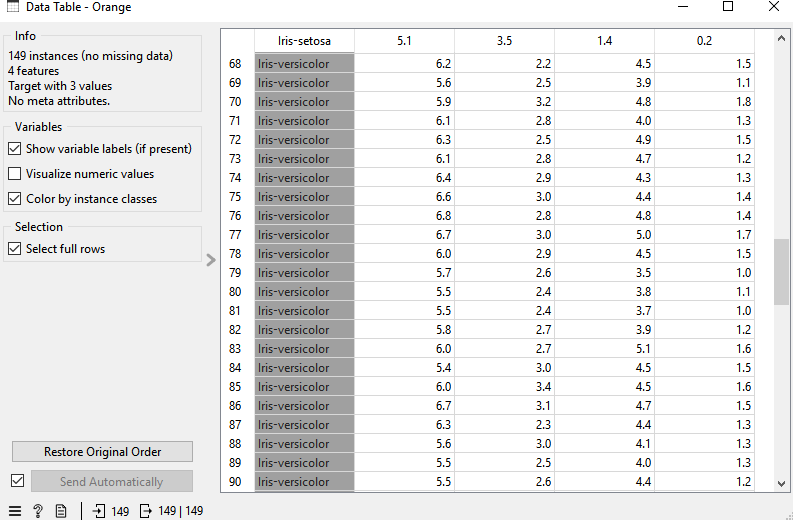


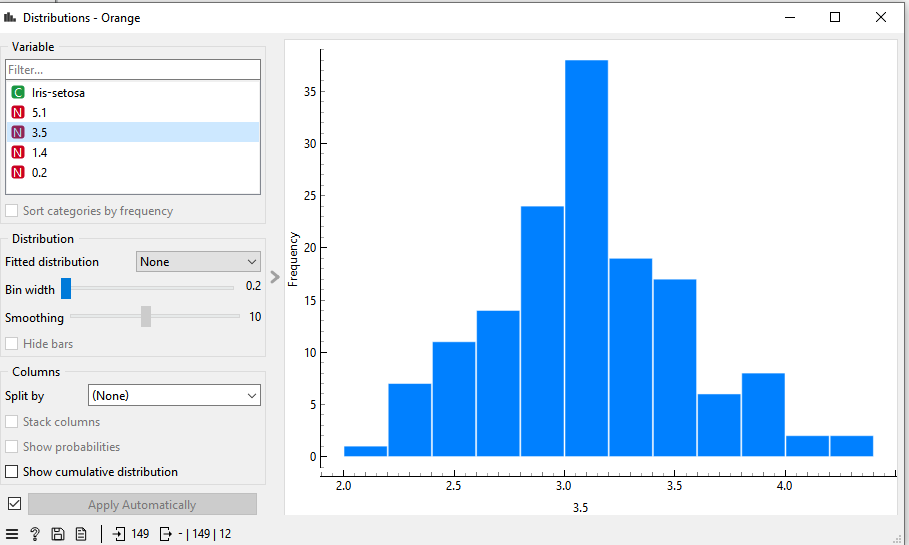


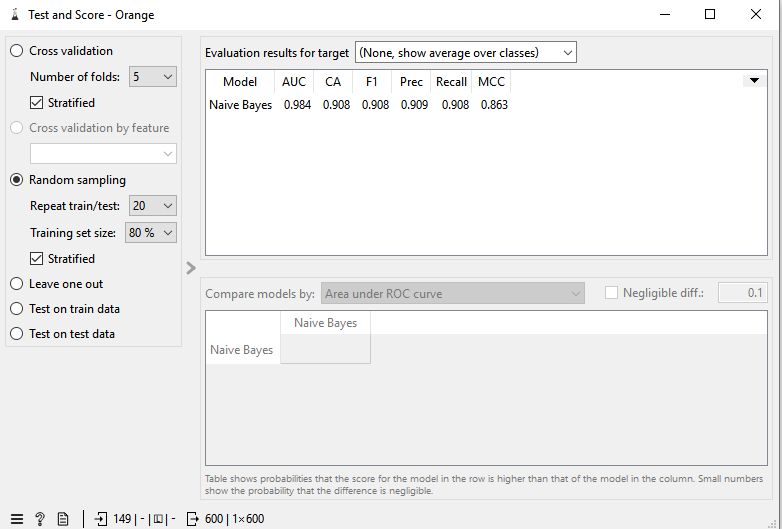


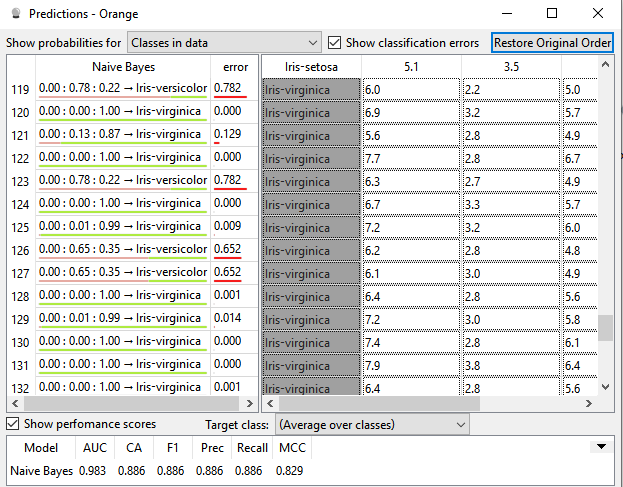
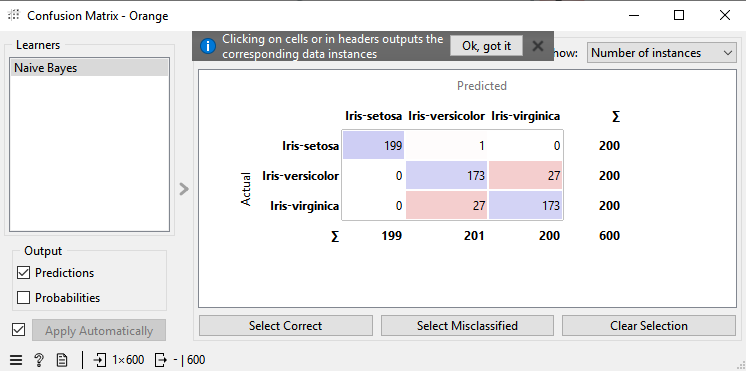


Dataset:Iris\_data.csv

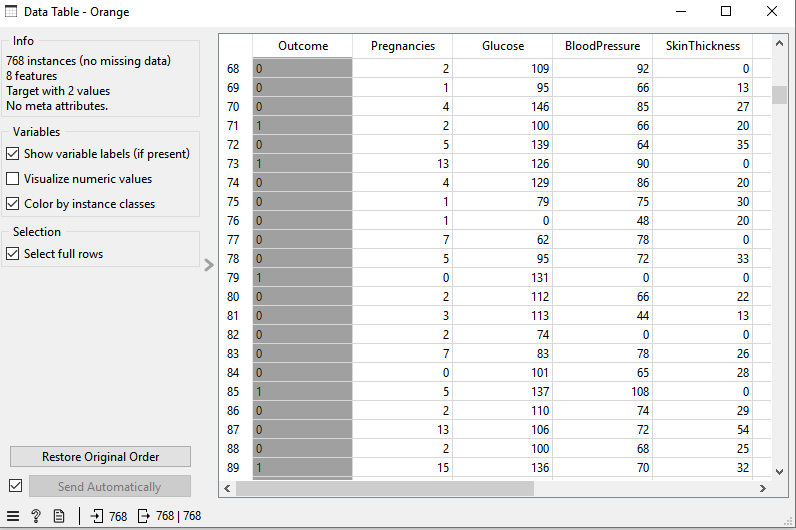


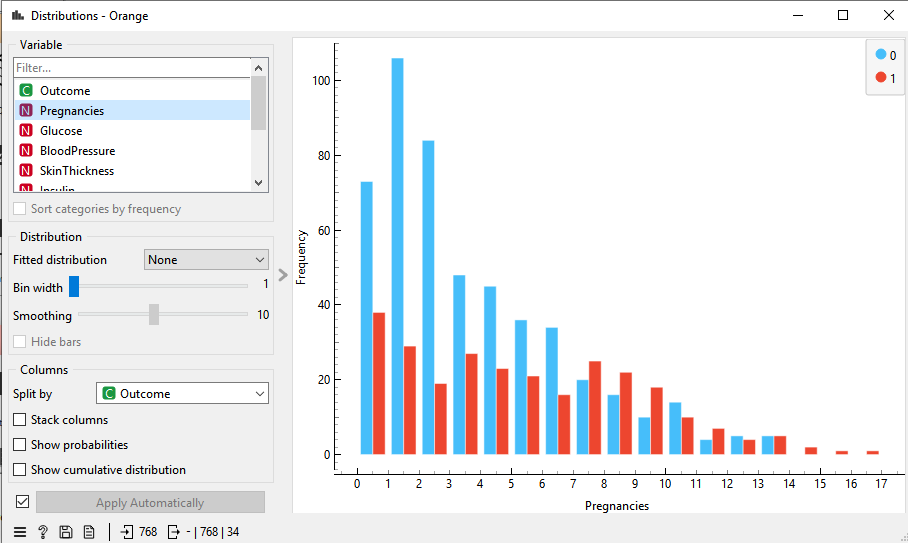


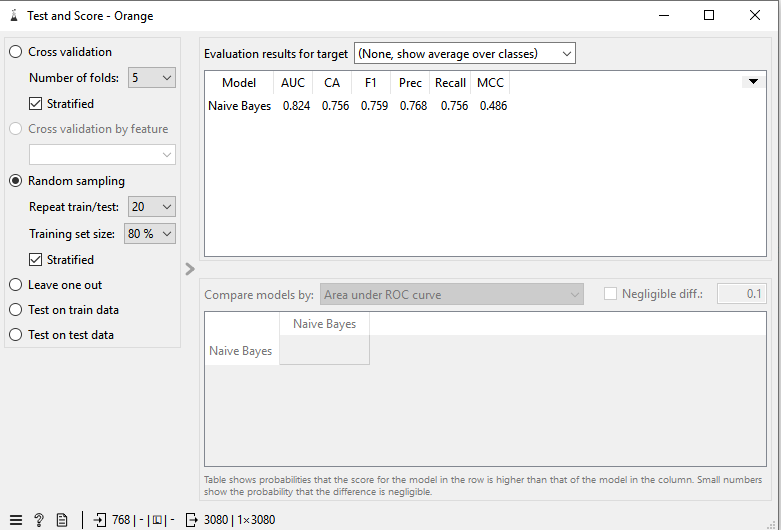


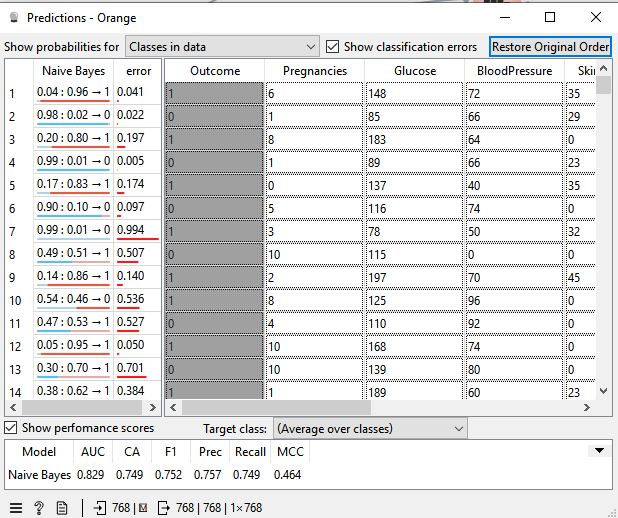
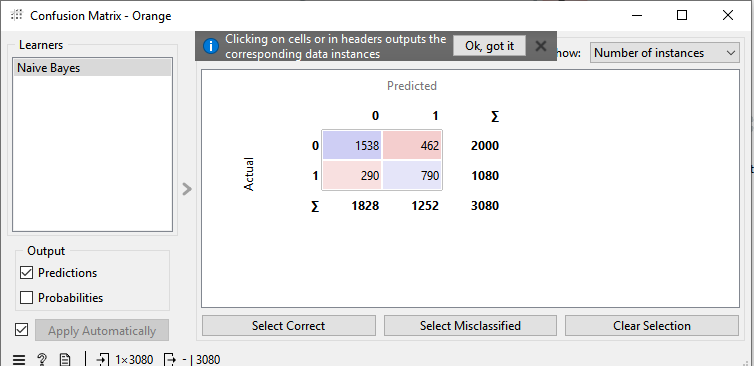


Dataset:Diabetes.csv



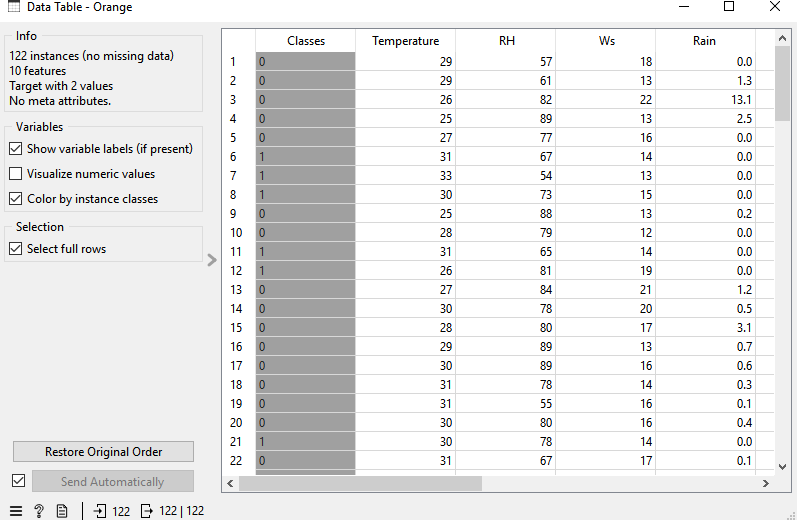


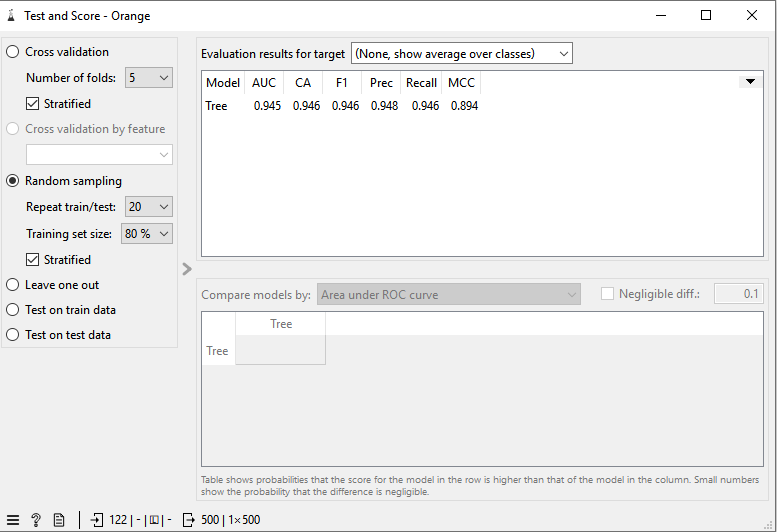


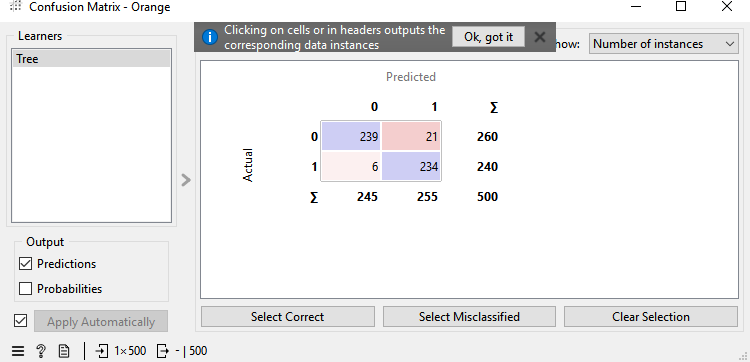


DecisionTree

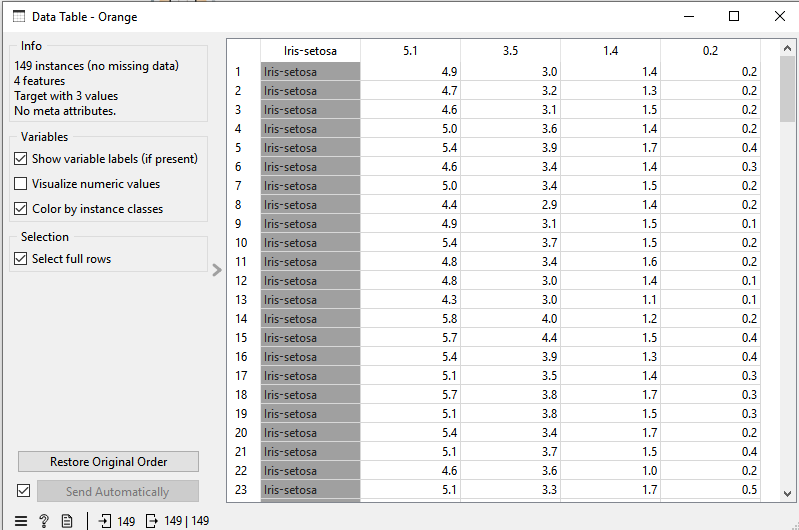
Dataset:Algerian\_forest\_fires\_no\_dates

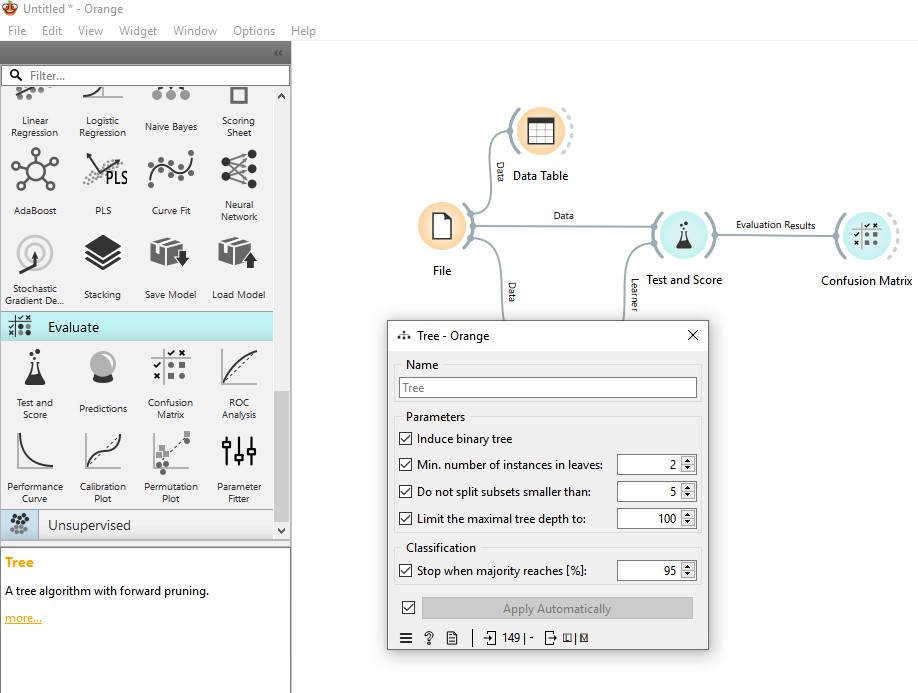


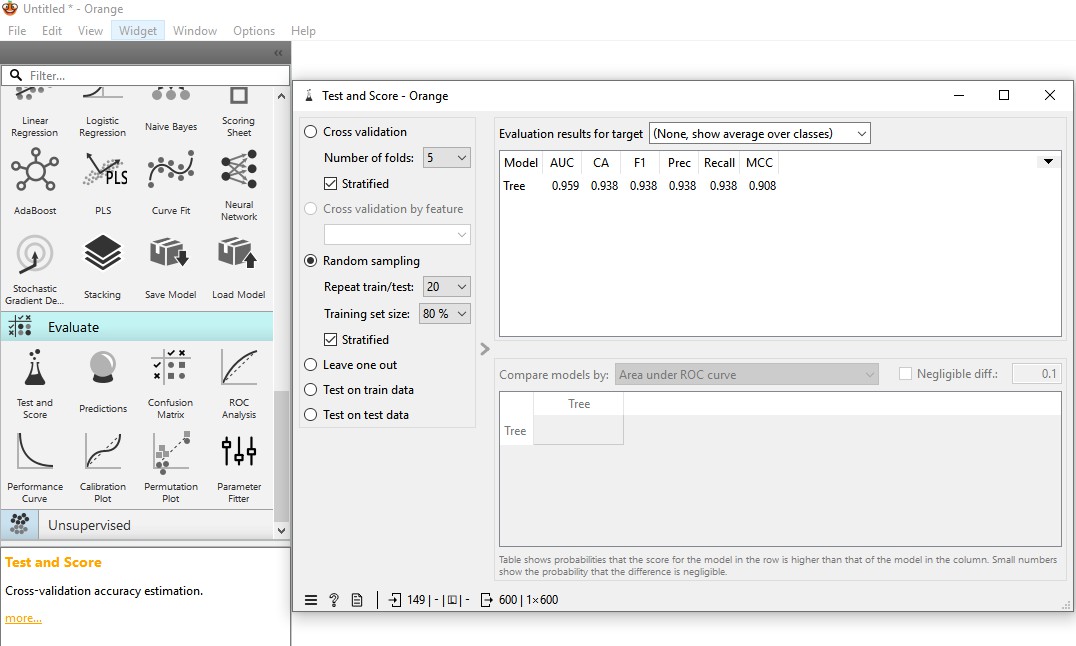


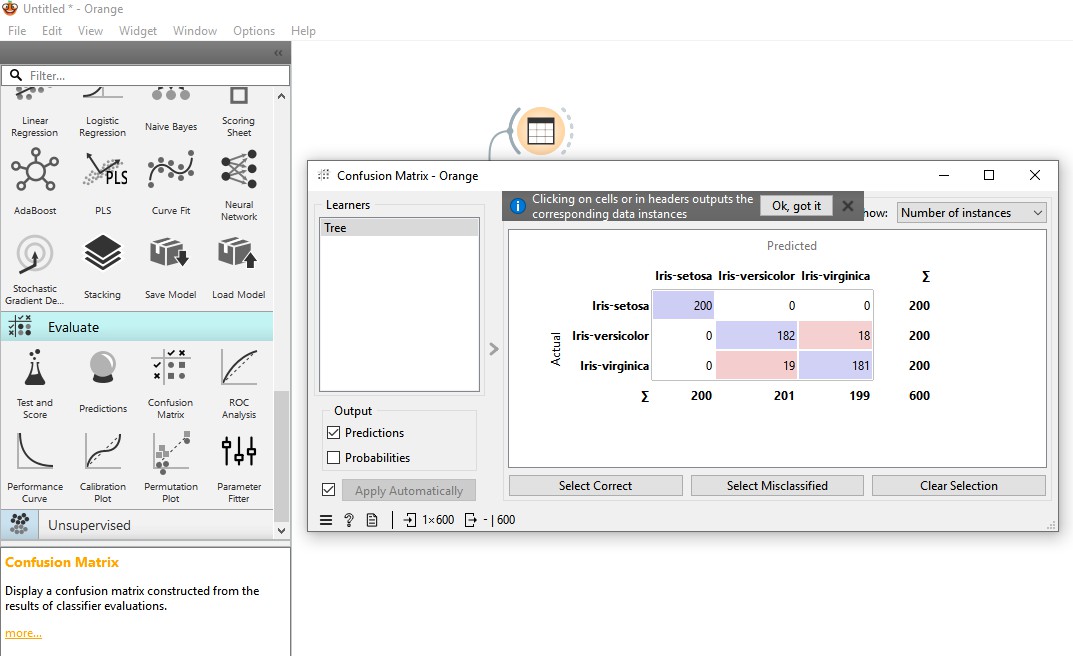


Dataset:iris\_data.csv

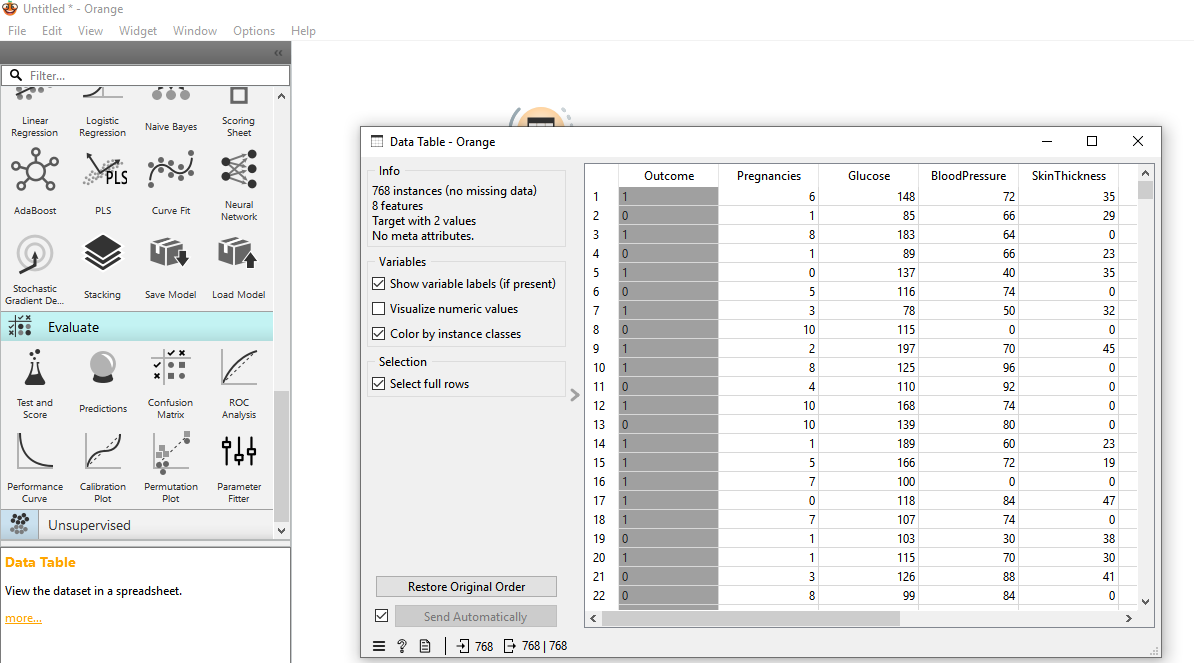


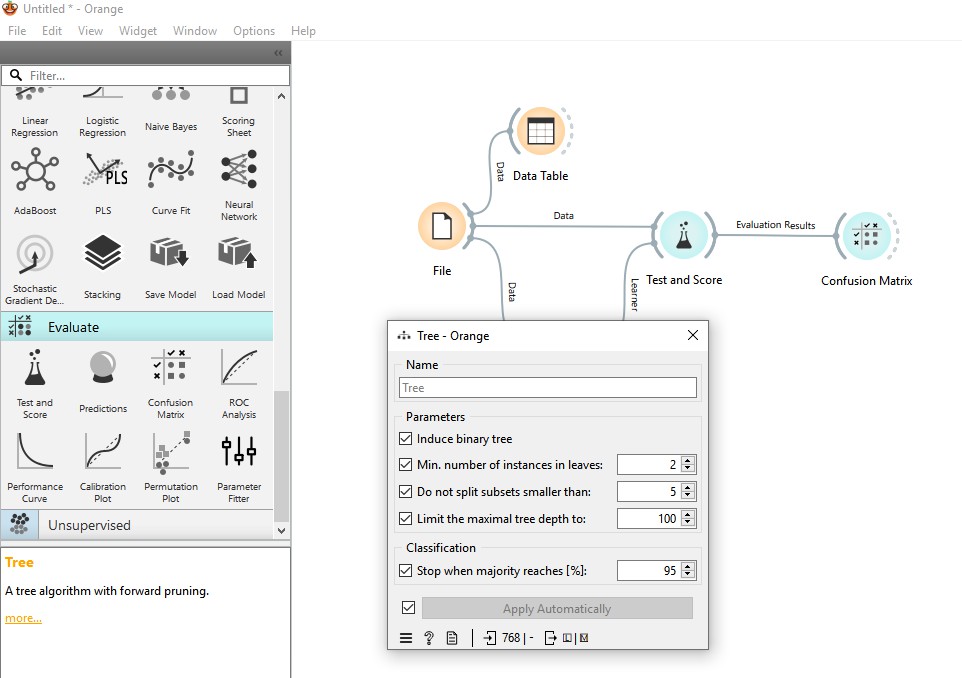


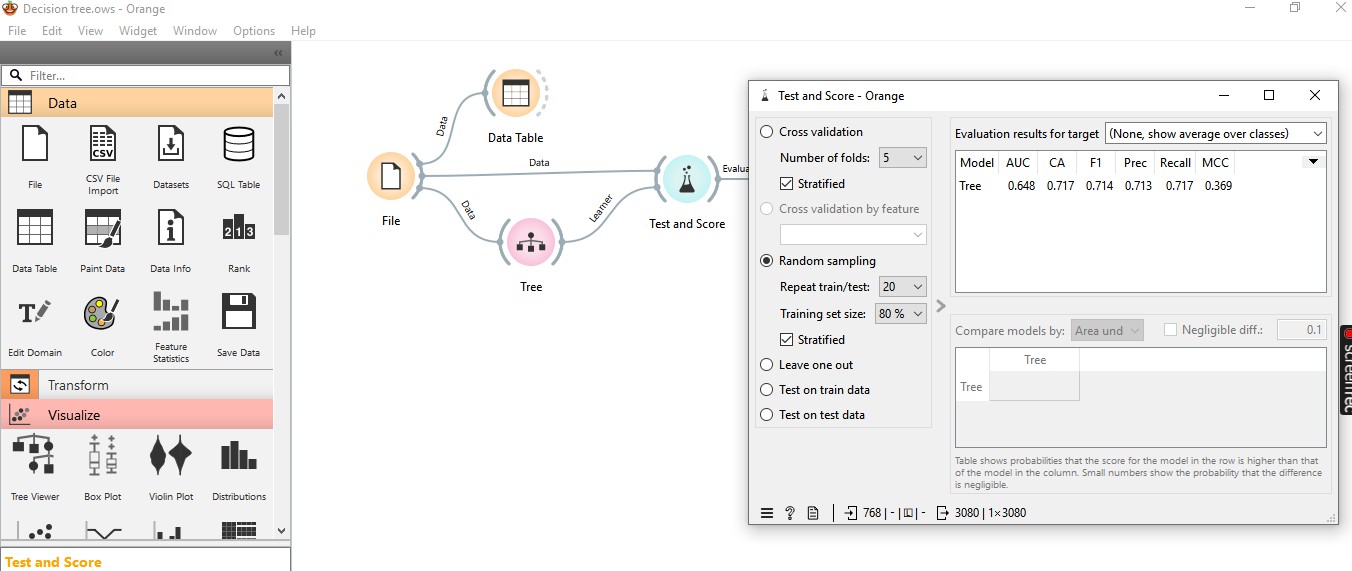


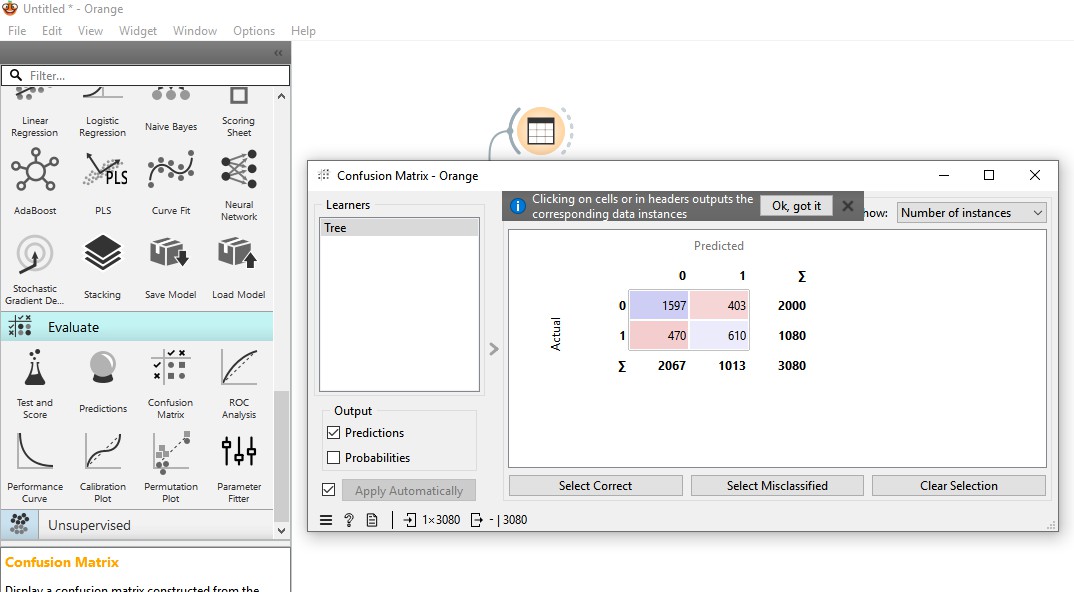


Dataset:diabetes.csv



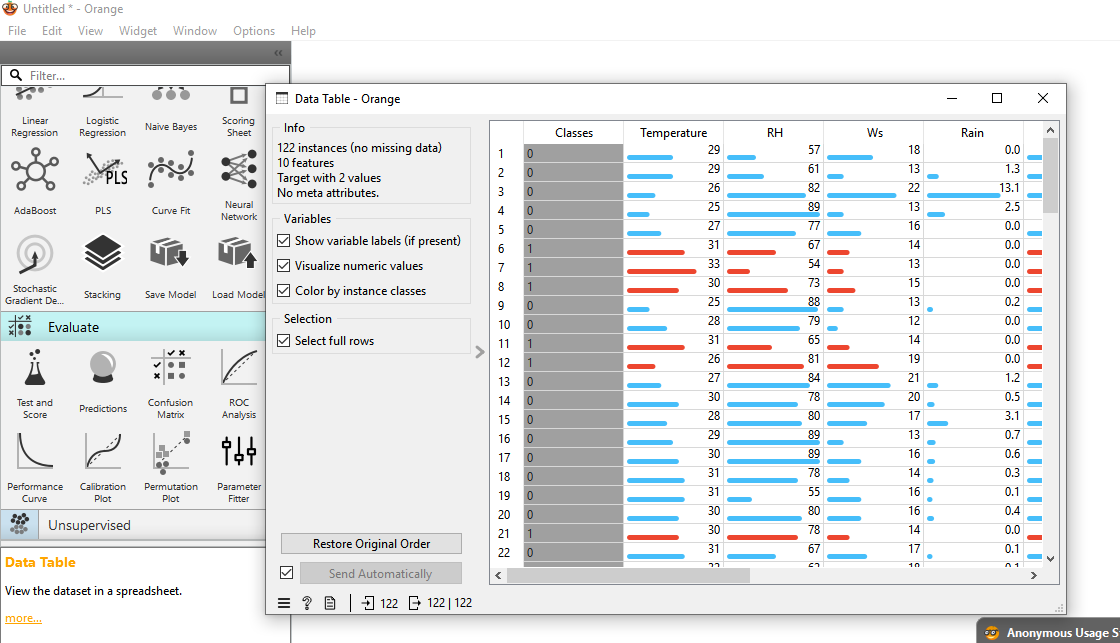


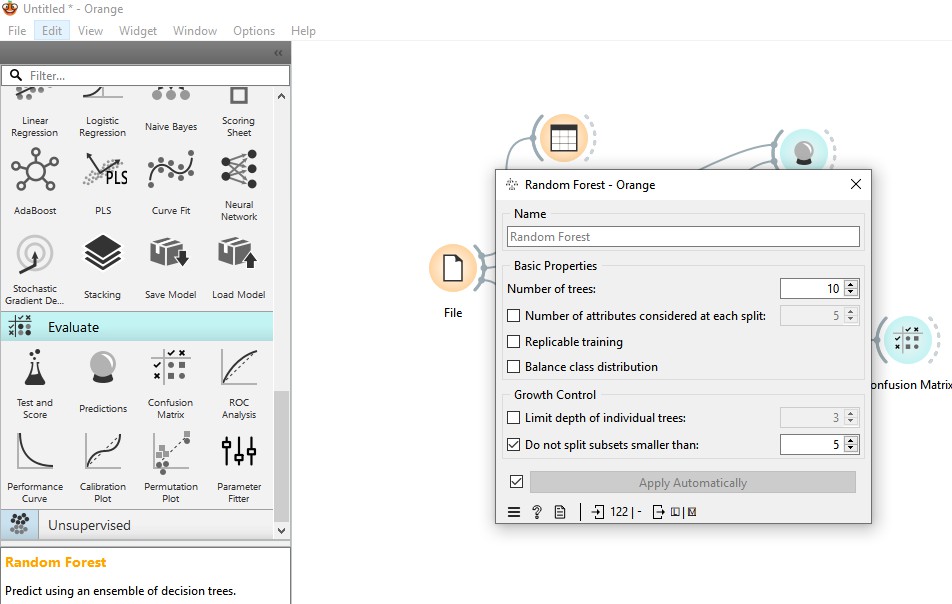


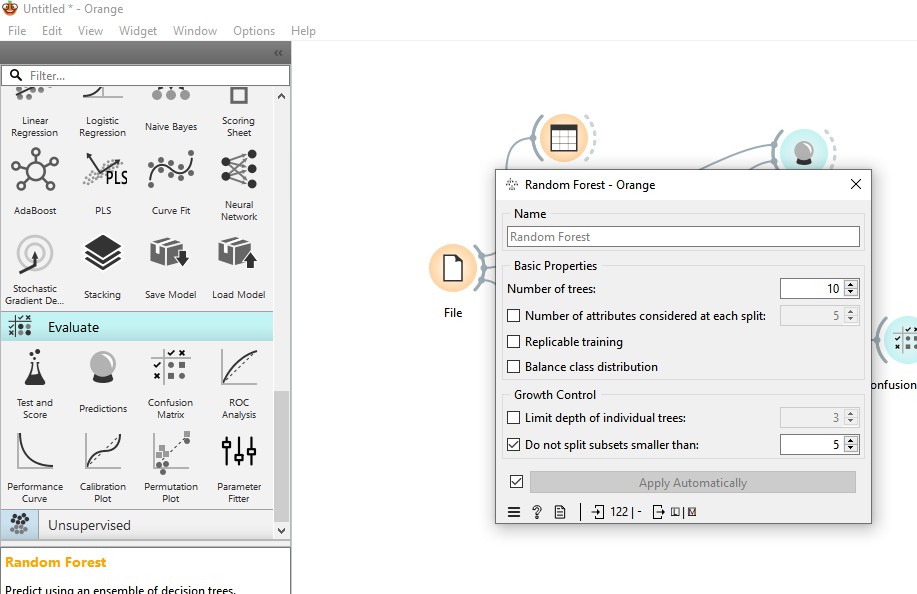


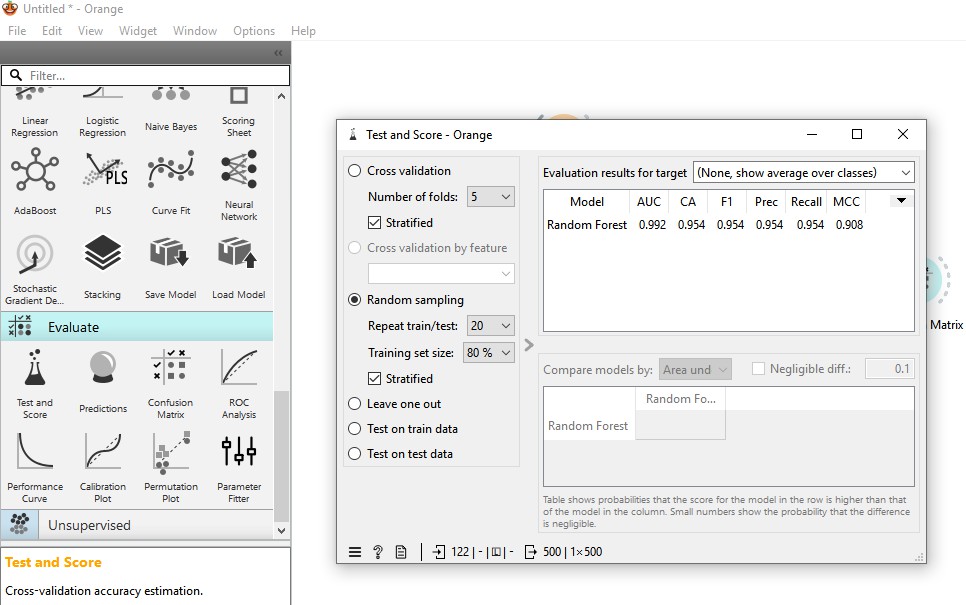
RandomForest:

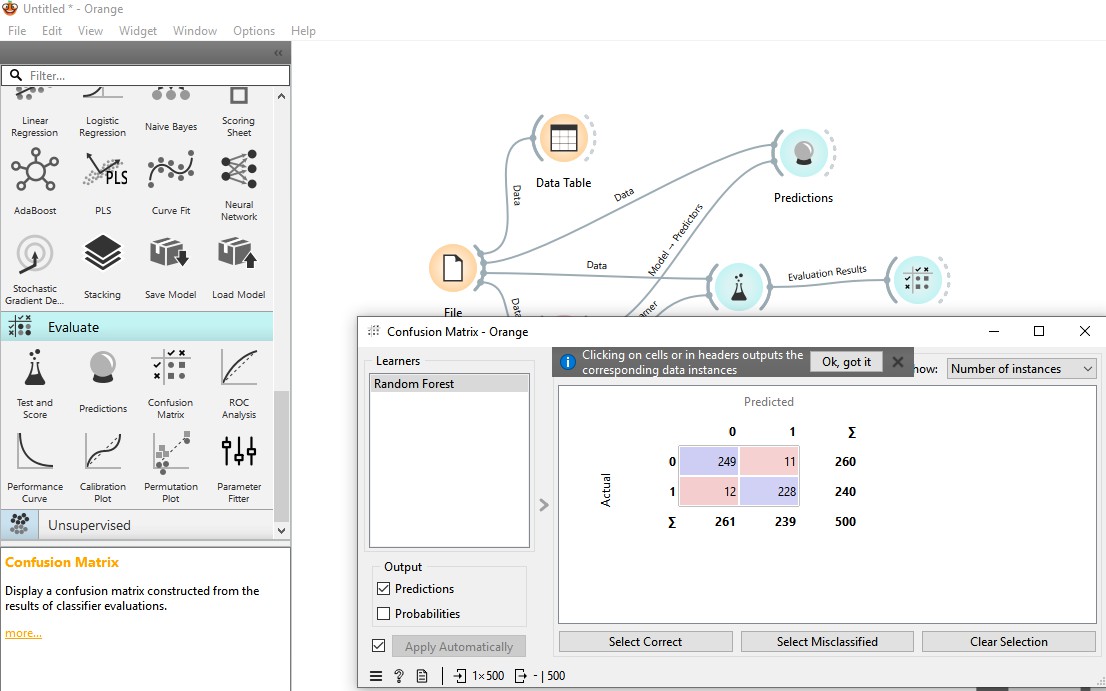
Dataset:Algerian\_forest\_fires\_no\_dates

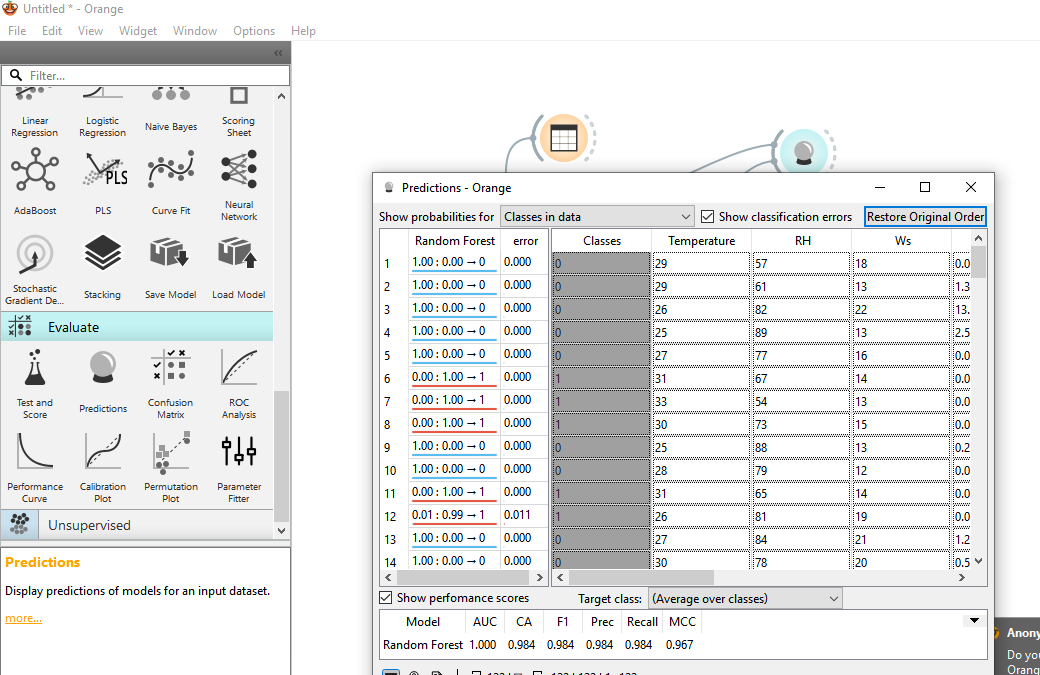




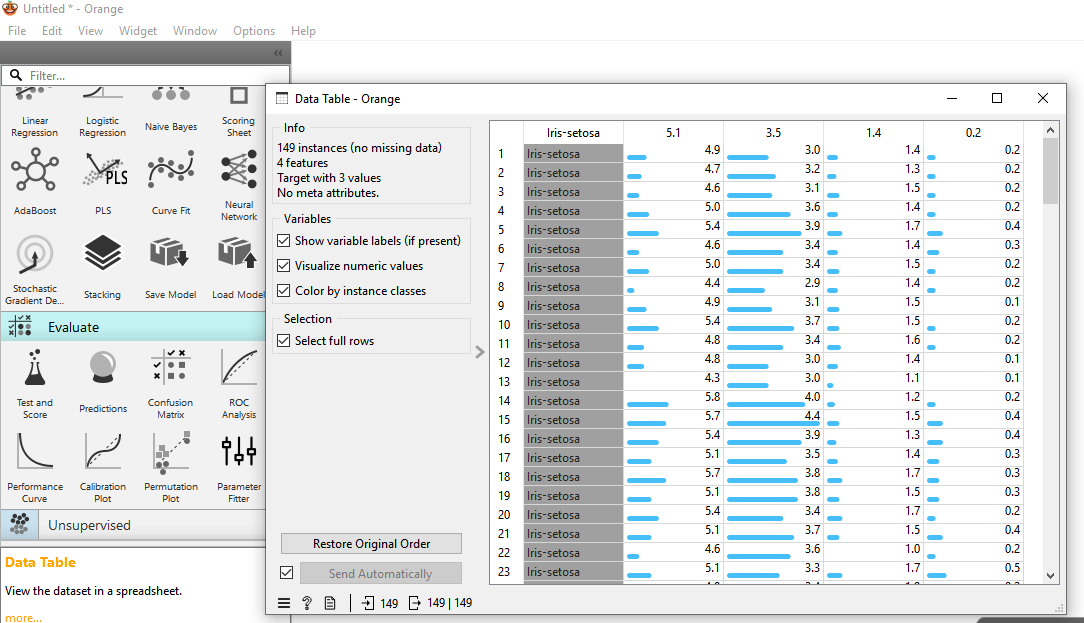


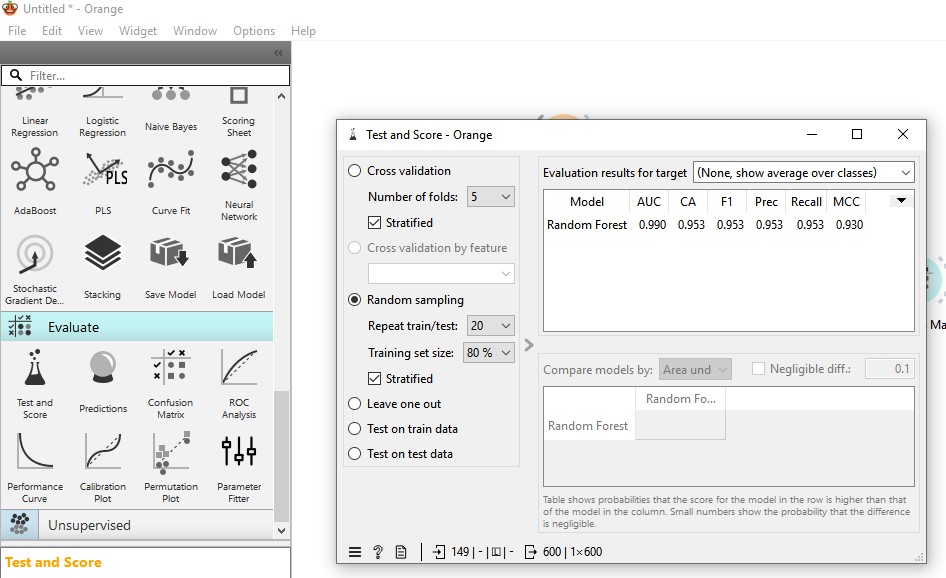


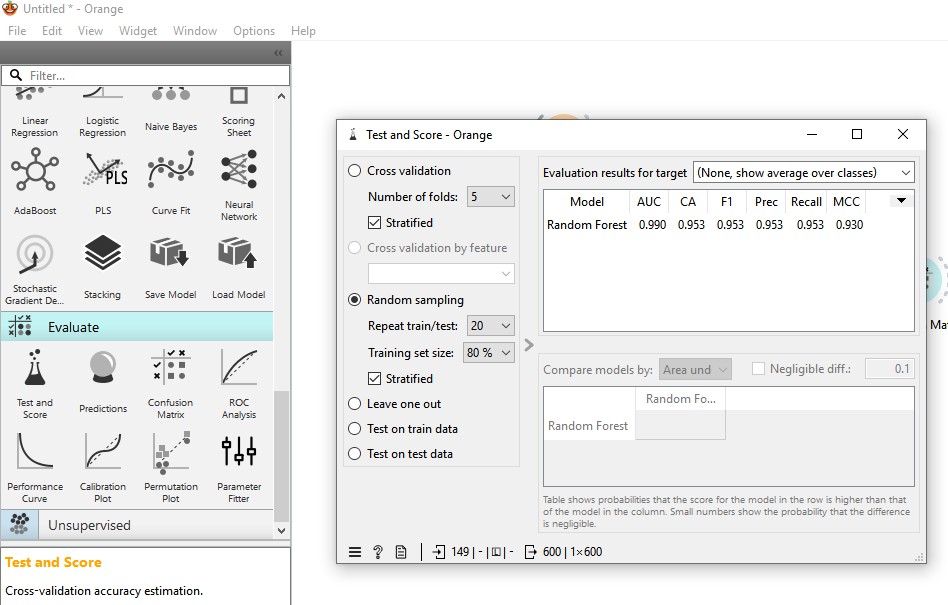
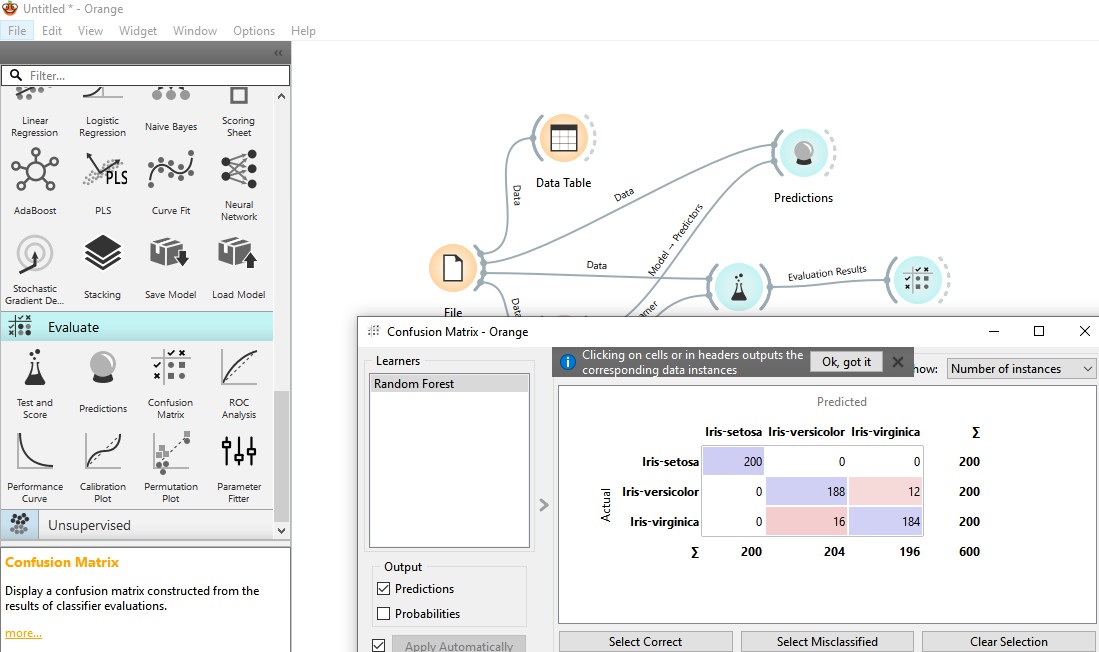


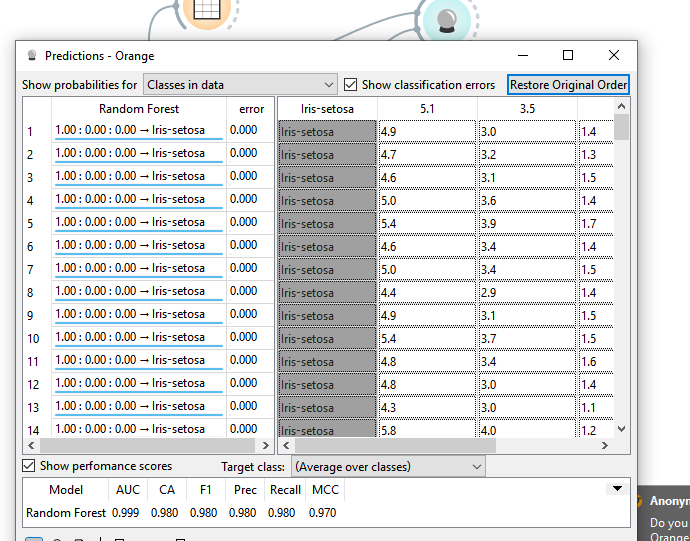


Dataset:iris\_data.csv

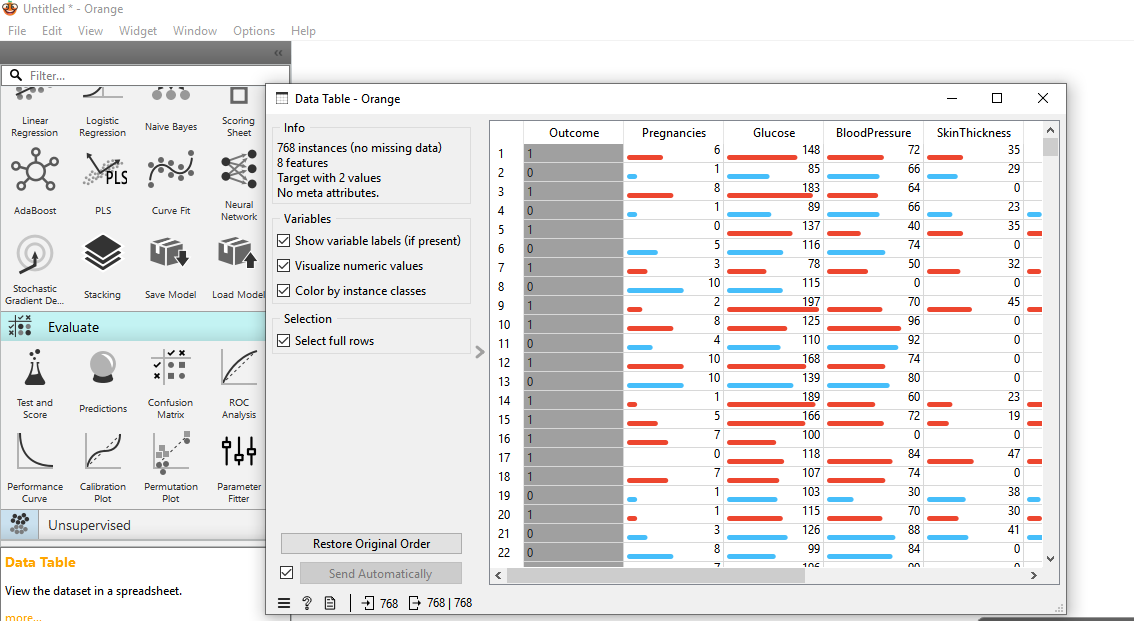


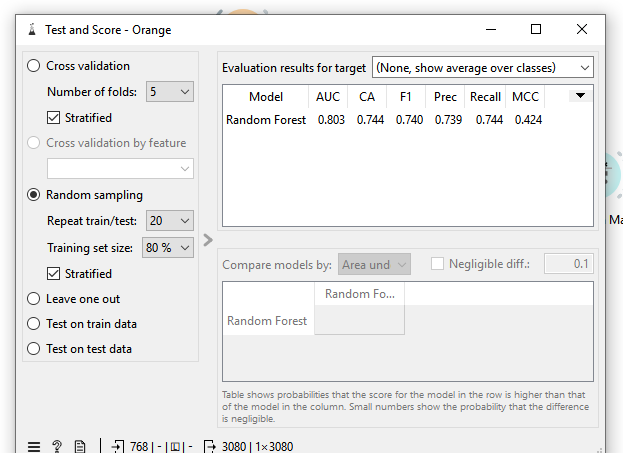


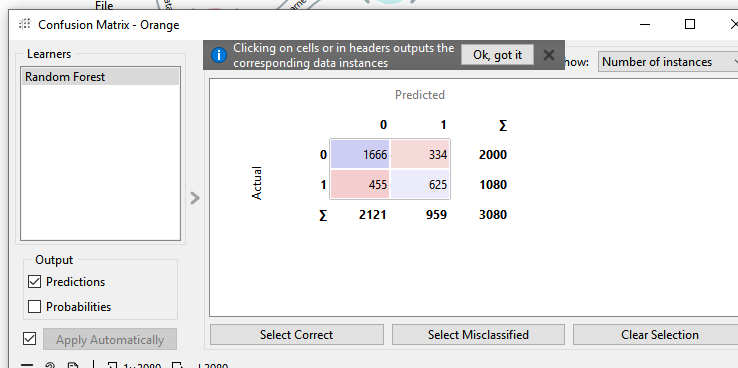


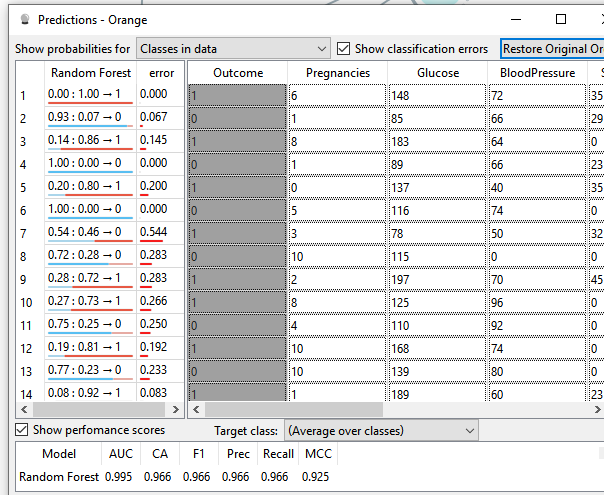


Dataset:Diabetes.csv

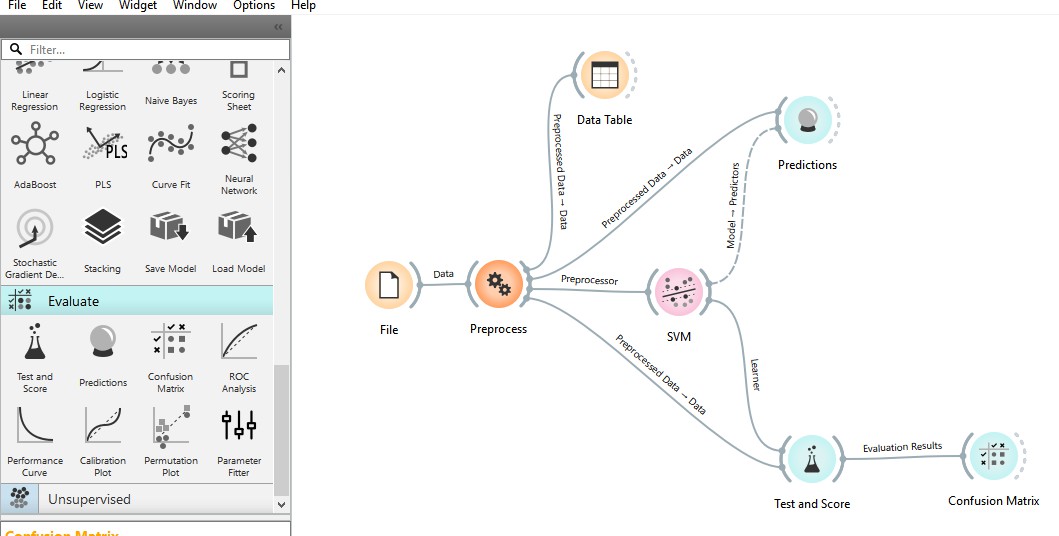




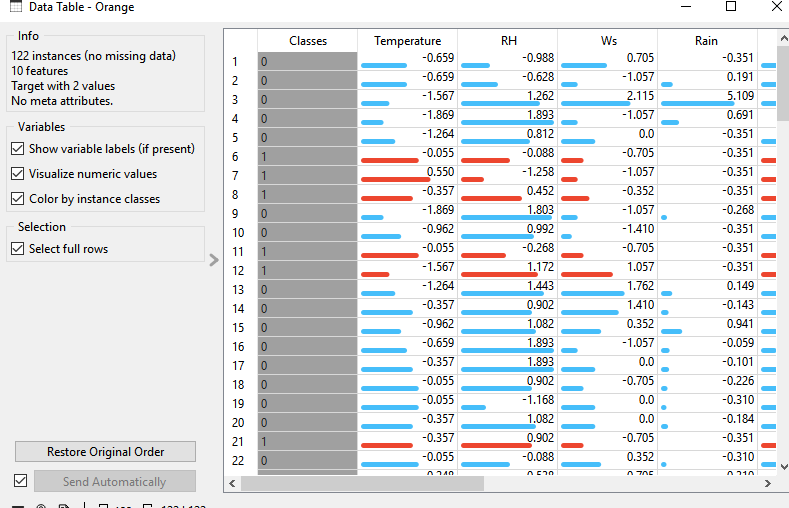
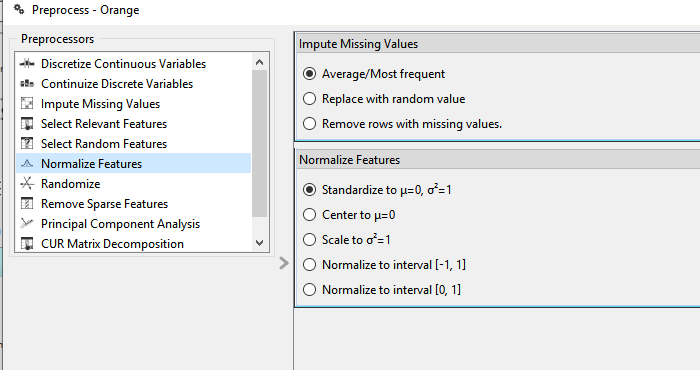


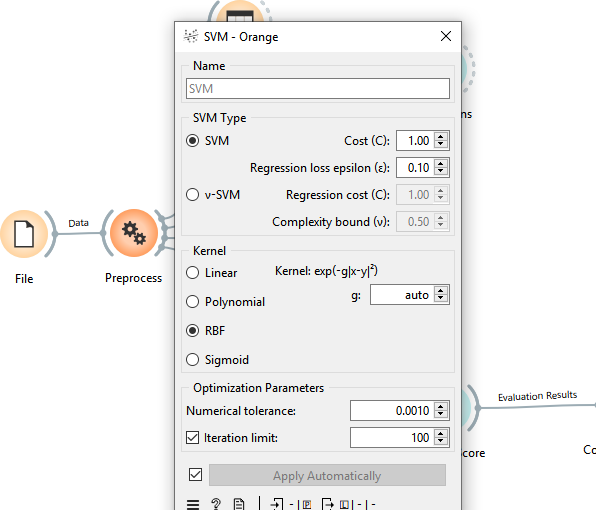


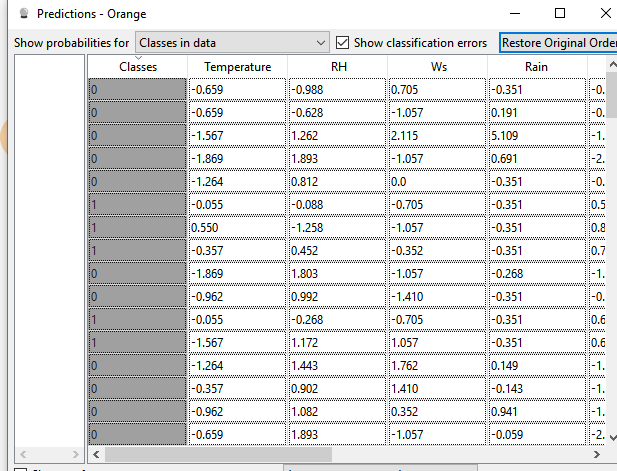
SVM

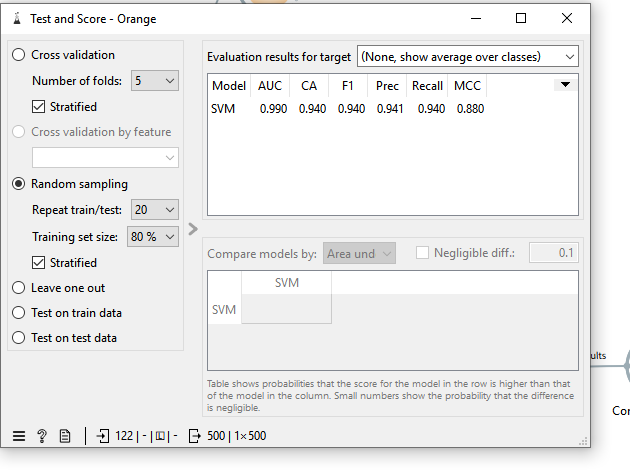


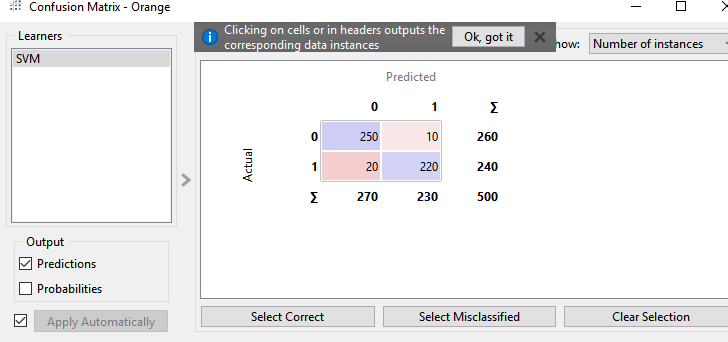
Dataset:Algerian\_forest\_fire.xlsxfromKaggle



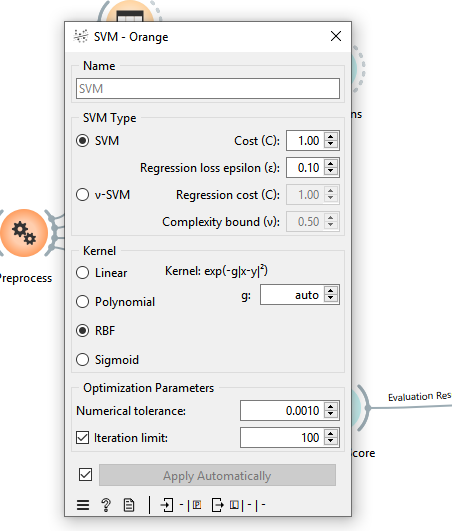
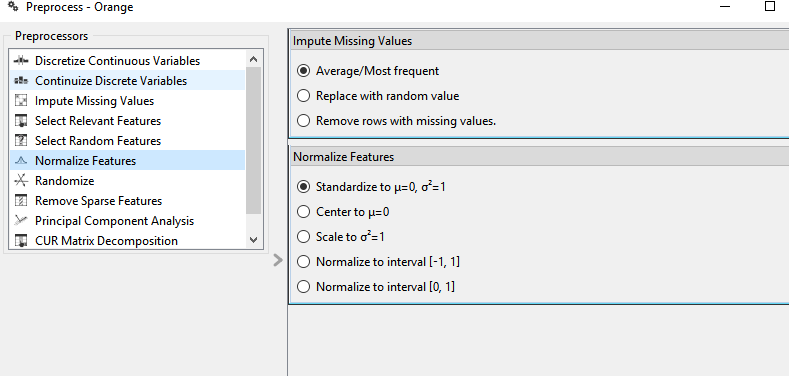


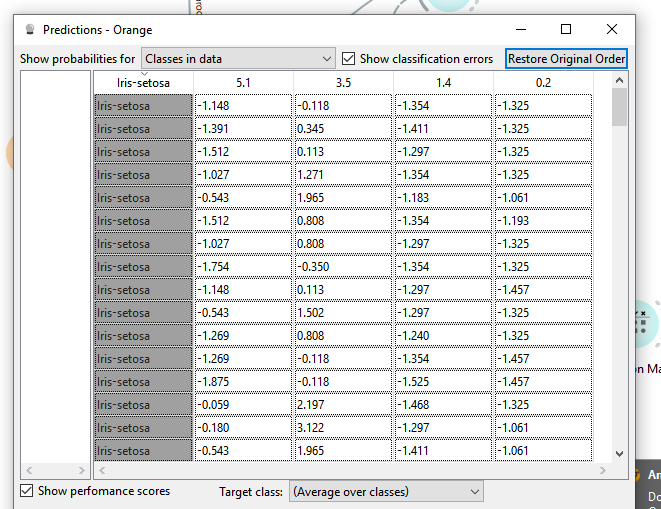


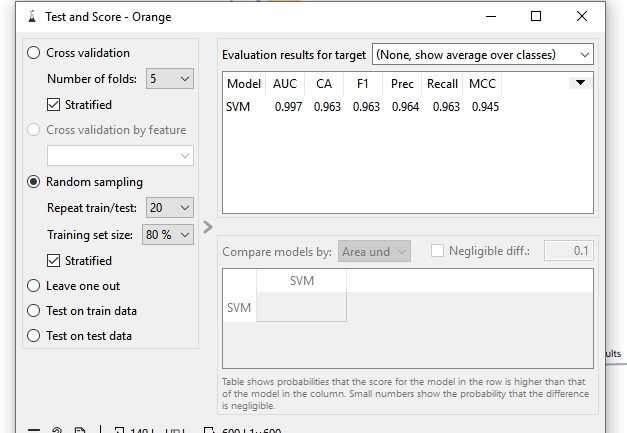


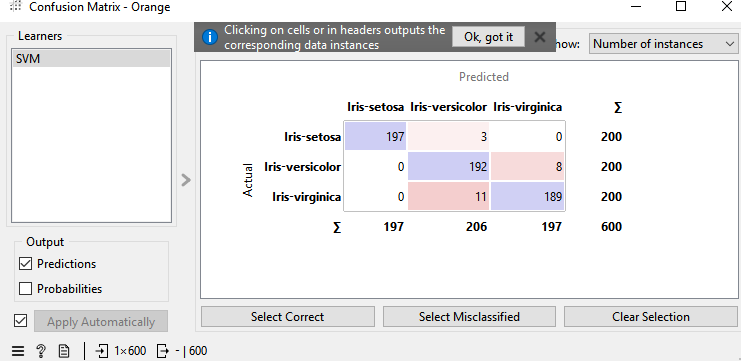


Dataset:Iris\_data.csv

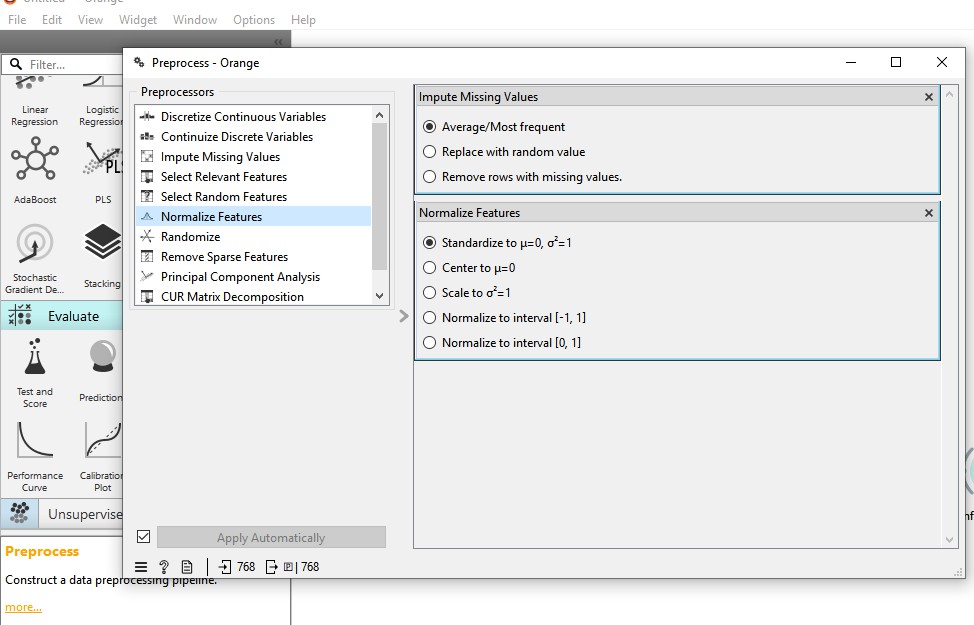


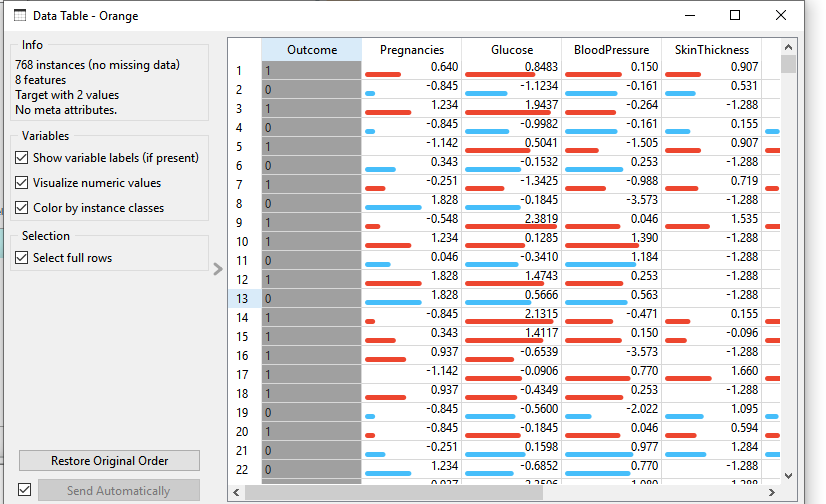


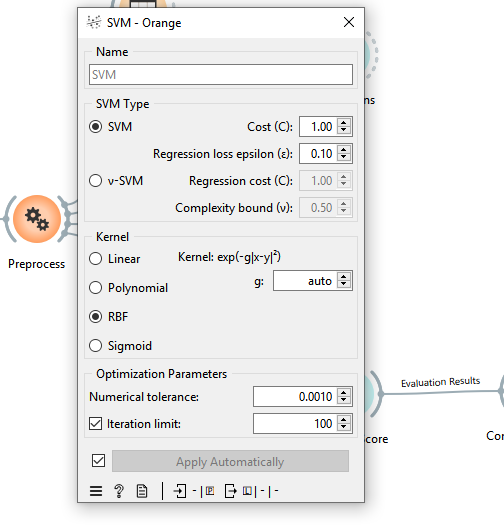


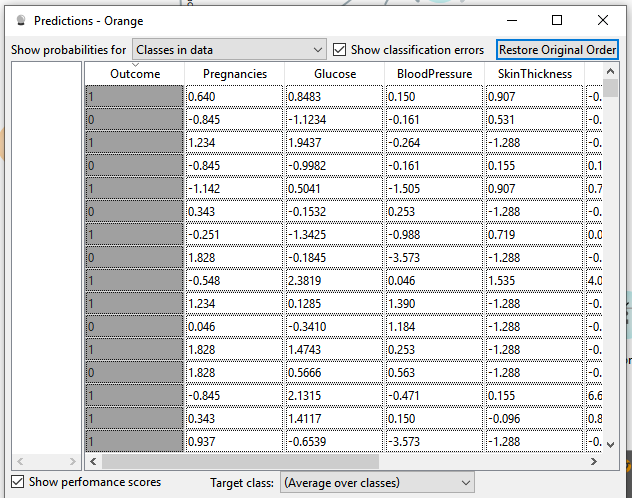


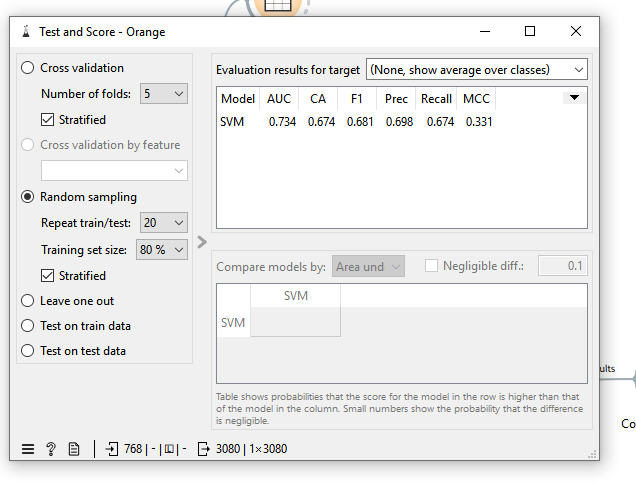
Dataset:Diabete’s.xlsx

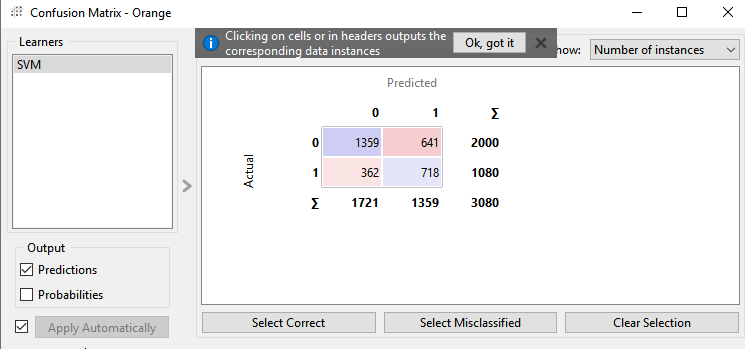












ComparativeanalysisusingthetoolOrange

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Classification | Dataset | Accuracy | Sensitivity | Specificity | Precision | F1-  Score | AUC- ROC |
| Naïve Bayes | Algerian Forest Firesfrom Kaggle | 85.6% | 83.2% | 86.9% | 84.1% | 83.6% | 87.2% |
| Iris.csv  from kaggle | 92.4% | 91.8% | 93.0% | 92.1% | 91.9% | 94.2% |
| Diabetes | 78.2% | 75.6% | 80.1% | 76.9% | 76.2% | 81.0% |
| DecisionTree | Algerian Forest Fires | 89.1% | 87.5% | 90.2% | 88.3% | 87.8% | 91.5% |
| Iris | 95.3% | 94.7% | 96.0% | 94.9% | 94.8% | 96.8% |
| Diabetes | 81.9% | 79.2% | 83.6% | 80.4% | 79.8% | 84.5% |
| RandomForest | Algerian  Forest Fires | 91.5% | 90.0% | 92.3% | 90.7% | 90.3% | 93.0% |
| Iris | 96.8% | 96.3% | 97.2% | 96.5% | 96.4% | 98.1% |
| Diabetes | 84.7% | 82.1% | 86.5% | 83.4% | 82.8% | 87.2% |
| SVM | Algerian  Forest Fires | 88.2% | 86.0% | 89.4% | 87.1% | 86.5% | 90.0% |
| Iris | 94.6% | 94.0% | 95.1% | 94.3% | 94.1% | 96.0% |
| Diabetes | 80.5% | 78.0% | 82.3% | 79.2% | 78.6% | 83.1% |

**Conclusion**

DecisionTreewasstrongbutslightlyoverfittedcomparedtoRandomForest.

RandomForestgavethebestresultsoverall.DecisionTreealsoperformedwell.NaïveBayesworkedbest for simple data like Iris but struggled with complex datasets. SVM performed well but was slightly less accurate than Random Forest. For real-world use, Random Forest is the best choice.

* Forhighaccuracyandgeneralization,RandomForestisthebestchoice.
* Forsimpledatasetswithindependentfeatures,NaiveBayesworkswell.
* ForstructureddatasetslikeIris,SVMandDecisionTreearegoodoptions.
* Forreal-worlddatasetslikeDiabetesandForestFires,DecisionTreeandRandomForestperform best.