**Terna Engineering College**

**Computer Engineering Department**

Program: Sem VII

[**Course: Artificial Intelligence & Soft Computing (AI&SC)**](https://github.com/Amey-Thakur/ARTIFICIAL-INTELLIGENCE-AND-SOFT-COMPUTING-AND-ARTIFICIAL-INTELLIGENCE-AND-SOFT-COMPUTING-LAB)

**Experiment No. 05**

**PART B**

**(PART B: TO BE COMPLETED BY STUDENTS)**

***(Students must submit the soft copy as per the following segments within two hours of the practical. The soft copy must be uploaded on the Blackboard or emailed to the concerned lab in charge faculties at the end of the practical in case there is no Blackboard access available)***

| Roll No. 50 | Name: AMEY THAKUR |
| --- | --- |
| Class: BE-COMPS-50 | Batch: B3 |
| Date of Experiment: 24-08-2021 | Date of Submission: 24-08-2021 |
| Grade : |  |

**Aim:** Identify the Classification problem and create a Knowledge database for that problem and apply appropriate search methods for optimization.

**B.1 Software Code written by student:**

***(Paste your Problem Statement for Classification and Data set Used as a knowledge Database for Given Classification Problem)***

**Problem Statement**

To categorise the flower dataset in order to determine if it is in stock or not, using a combination of classifiers, evaluators, and search algorithms to improve accuracy and compare findings.

**B.2 Input and Output:**

***(Paste your screenshot of Analysis of Data, Relevant Attributes Selection by using at least Three methods)***

**Flowers.arff**

@relation Flowers

@attribute Type {Lilies, Orchids, Roses, Tulips}

@attribute Color {Red,Yellow,Blue}

@attribute Feature {Fresh,Luster,Vibrance,Strong}

@attribute InStock {Yes,No}

@data

Lilies Red Fresh Yes

Orchids Blue Luster Yes

Orchids Yellow Fresh Yes

Tulips Red Strong No

Lilies Yellow Vibrance Yes

Tulips Red Fresh No

Roses Yellow Strong No

Roses Blue Luster Yes

Lilies Blue Strong Yes

Orchids Red Vibrance Yes

Roses Yellow Fresh Yes

Tulips Red Luster No

Tulips Yellow Strong No

Roses Blue Vibrance Yes

Orchids Blue Luster Yes

Orchids Red Strong No

Lilies Blue Fresh Yes

Tulips Yellow Vibrance No

Roses Red Luster Yes

Roses Yellow Strong No

Lilies Red Vibrance Yes

Orchids Blue Fresh Yes

Lilies Red Fresh Yes

Tulips Yellow Vibrance No

Roses Blue Luster Yes

Roses Red Strong No

Lilies Red Luster Yes

Orchids Yellow Fresh Yes

Tulips Red Luster No

Roses Blue Fresh Yes

Lilies Blue Fresh Yes

Lilies Red Strong No

Orchids Red Fresh Yes

Tulips Yellow Vibrance No

Orchids Blue Luster Yes

Tulips Yellow Strong No

Lilies Red Vibrance Yes

Tulips Red Luster No

Roses Blue Strong No

Orchids Blue Vibrance Yes

Orchids Red Vibrance Yes

Roses Yellow Fresh Yes

Tulips Blue Luster No

Lilies Yellow Strong No

Roses Red Vibrance Yes

Orchids Blue Luster Yes

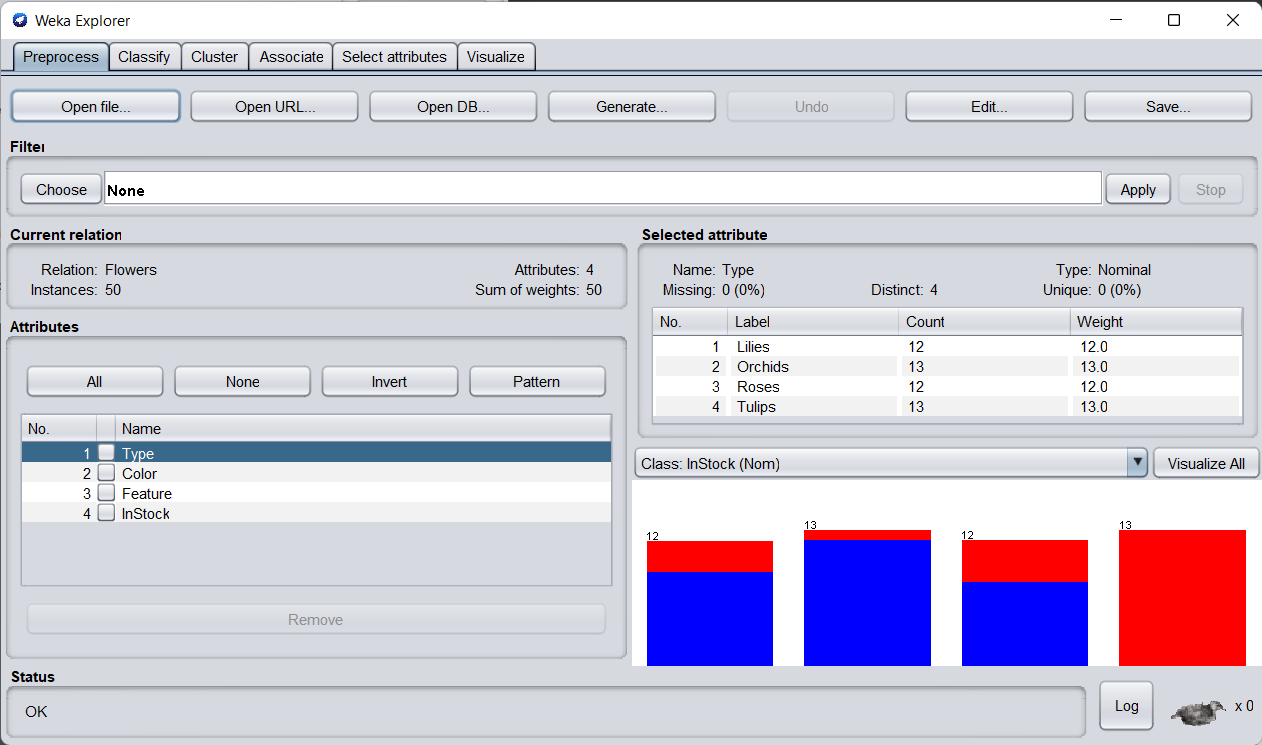
Tulips Red Fresh No

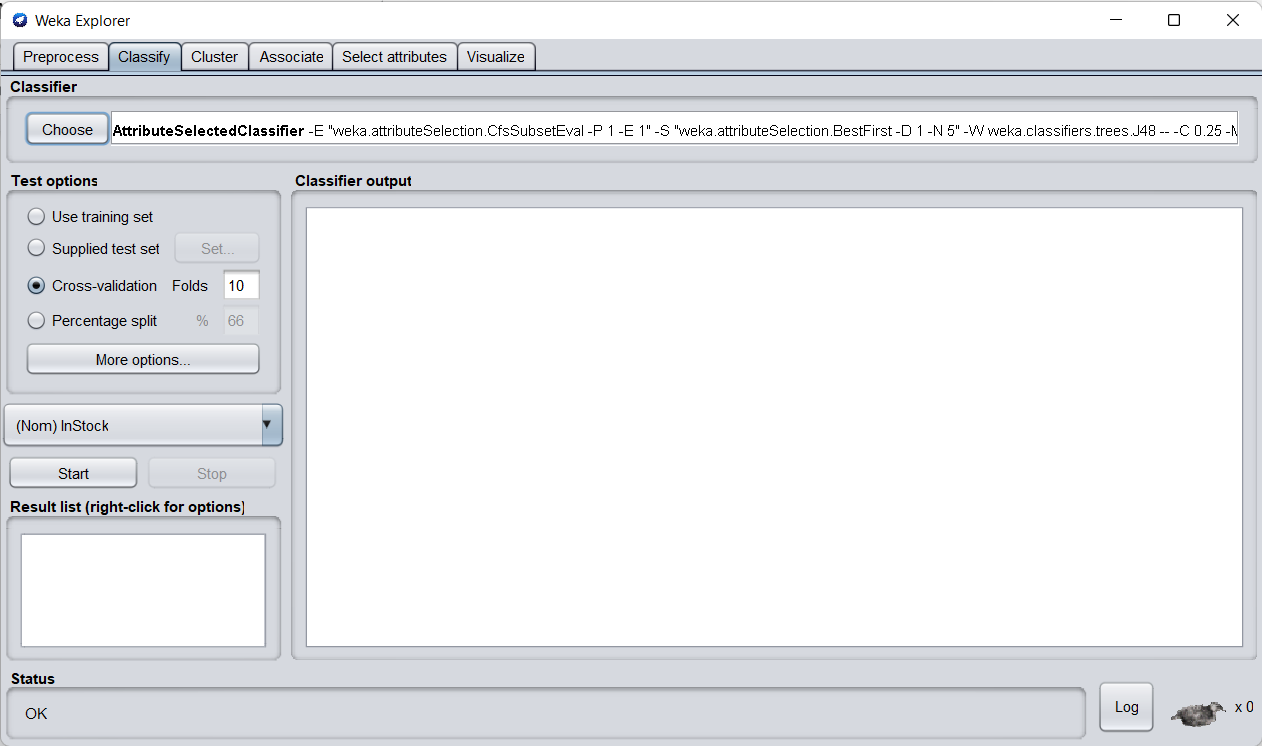
Lilies Blue Strong No

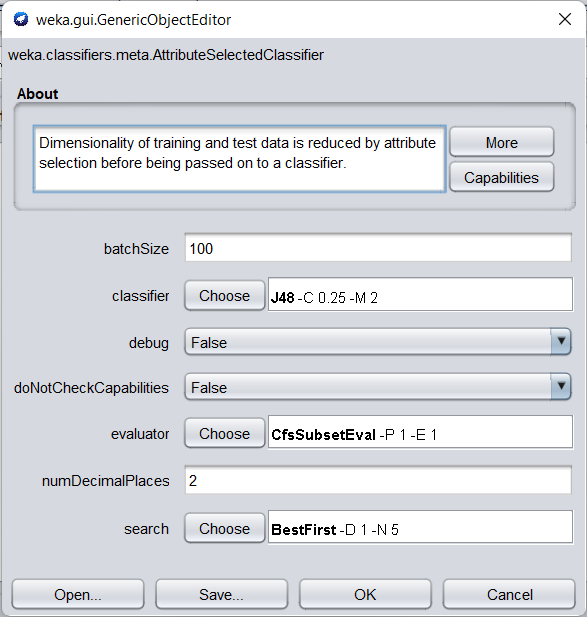
Tulips Yellow Luster No

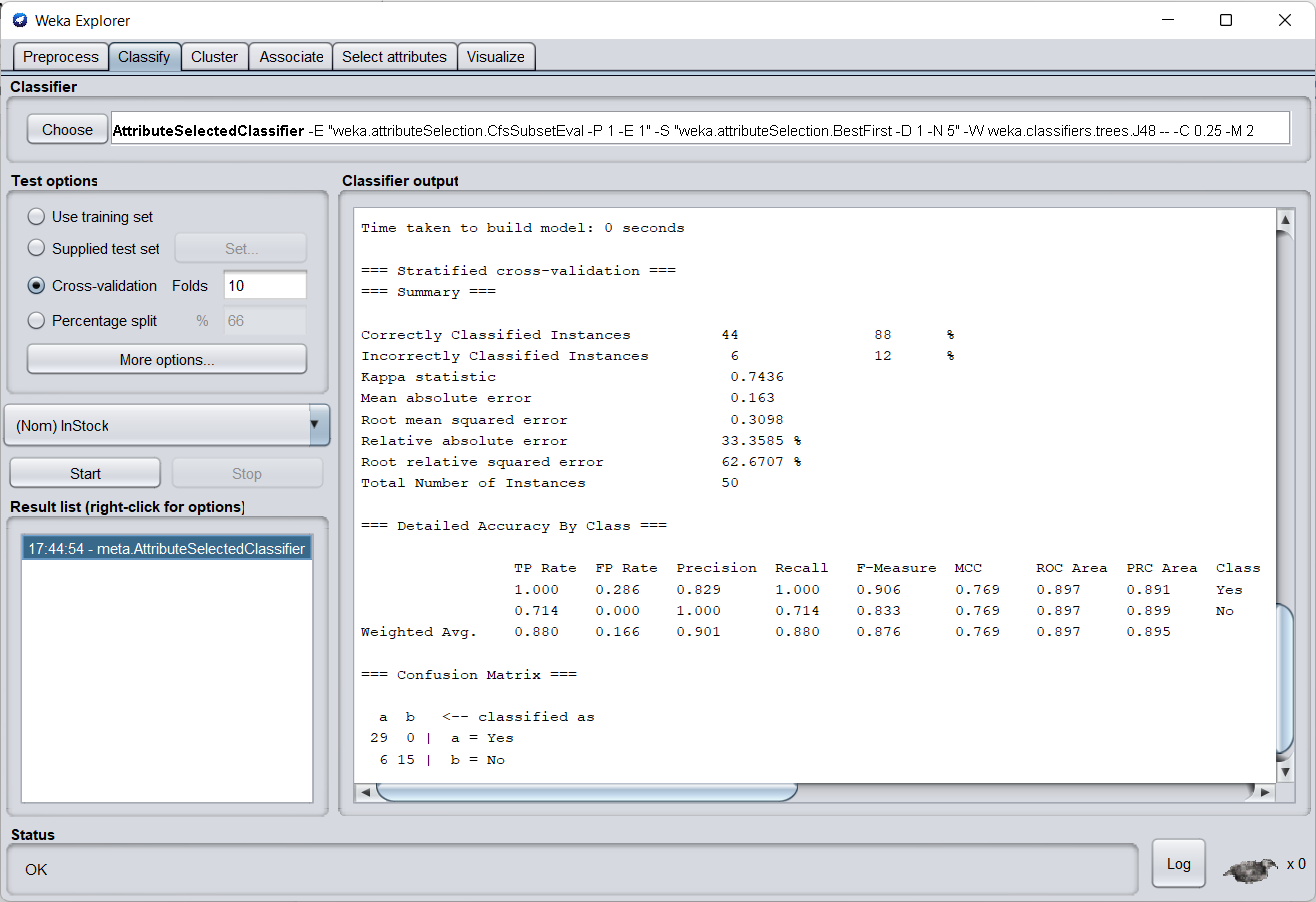
Orchids Red Luster Yes

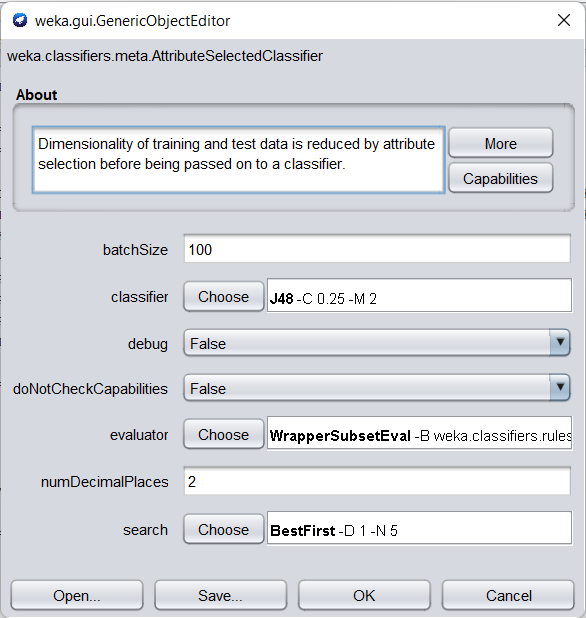
**Weka**

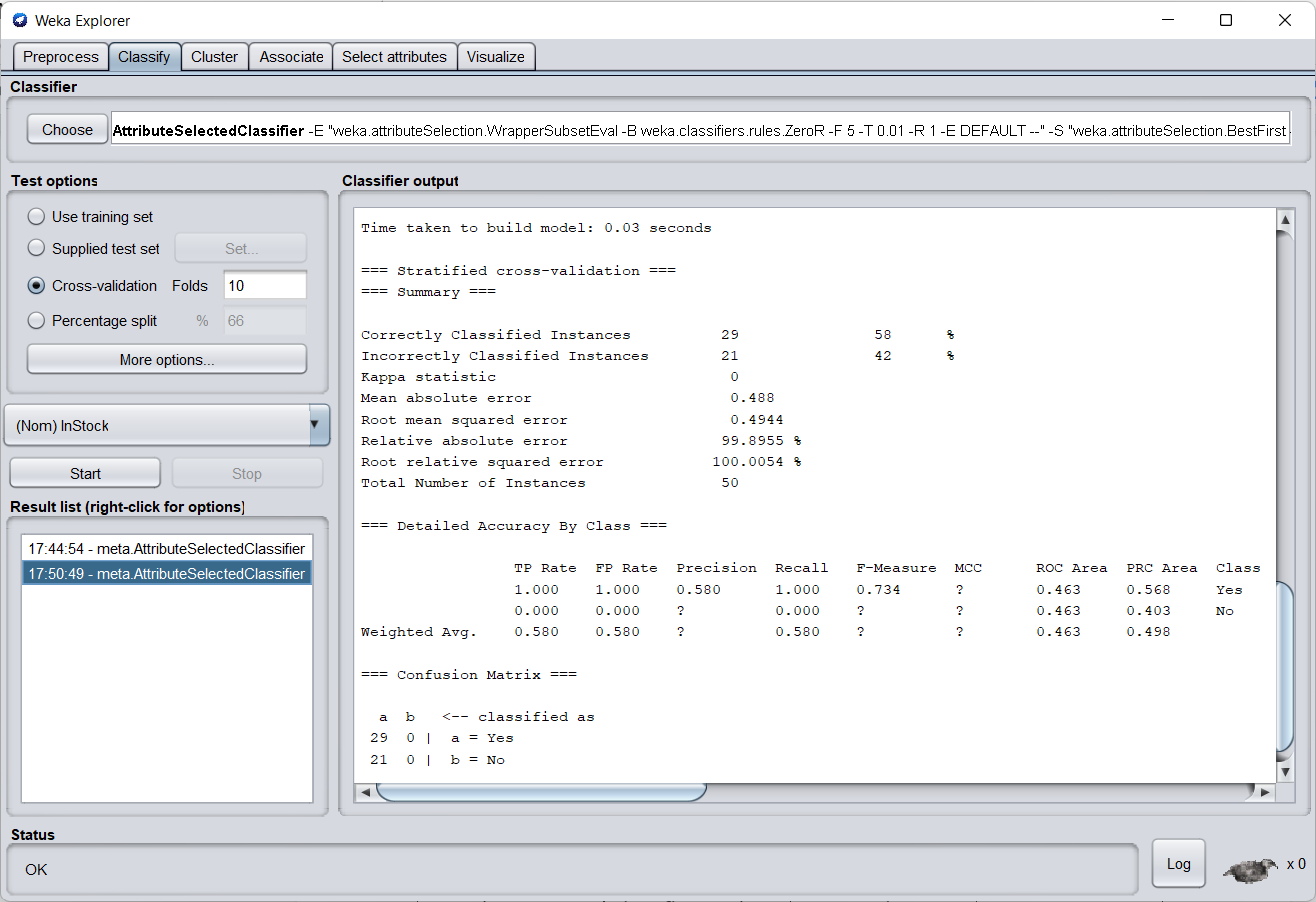
****

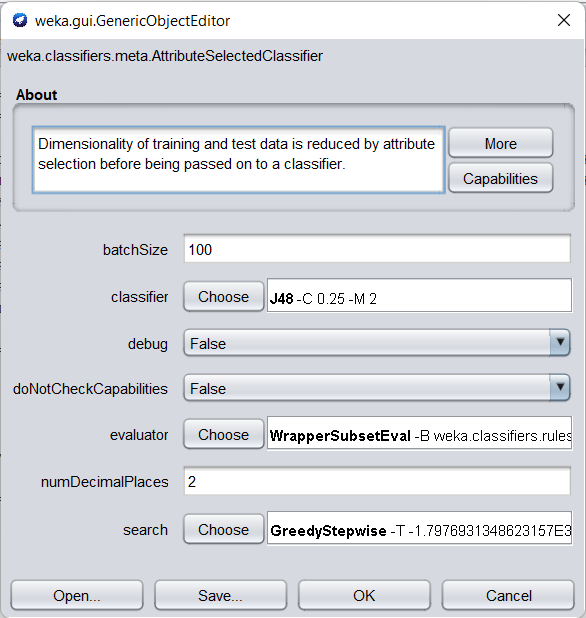
****

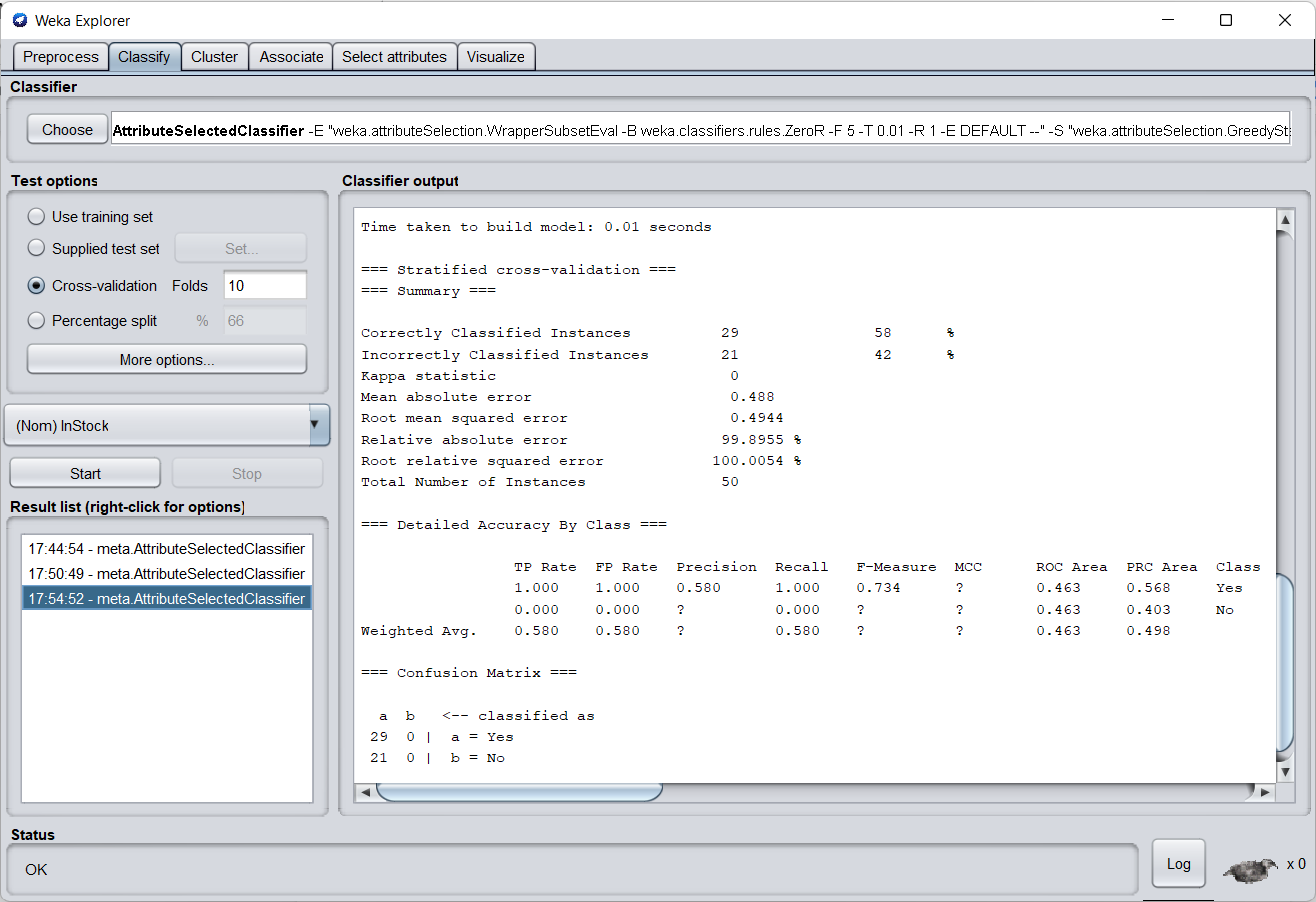
****

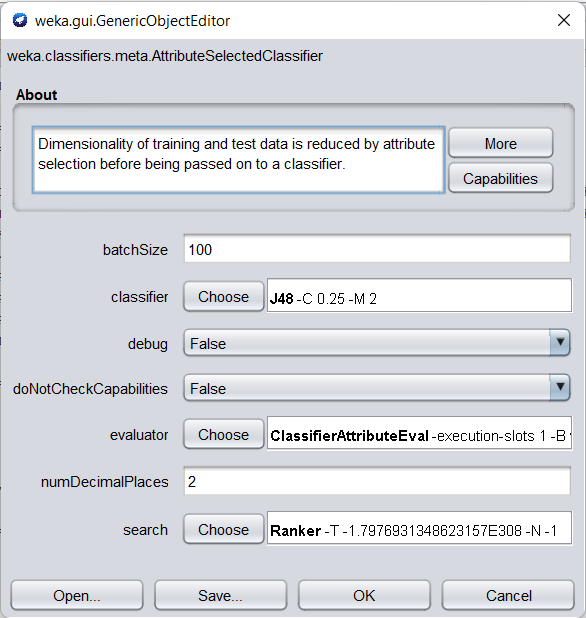
****

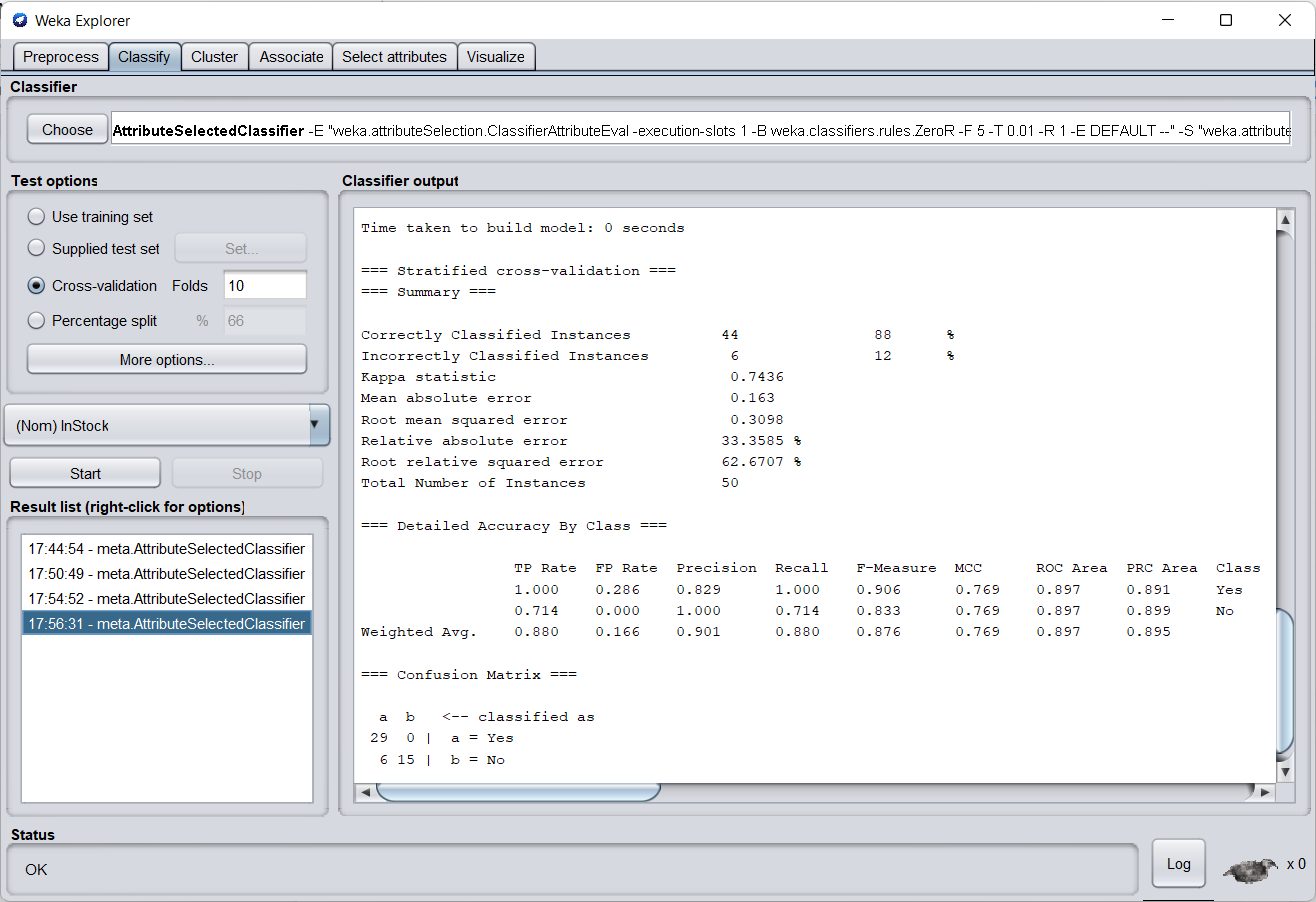
****

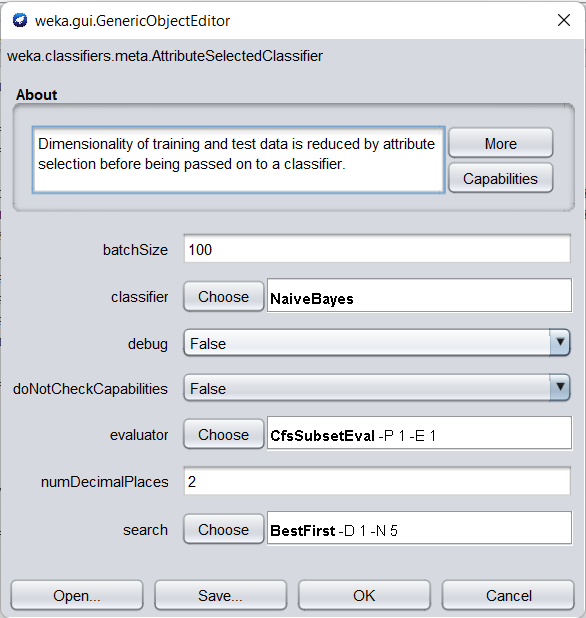
****

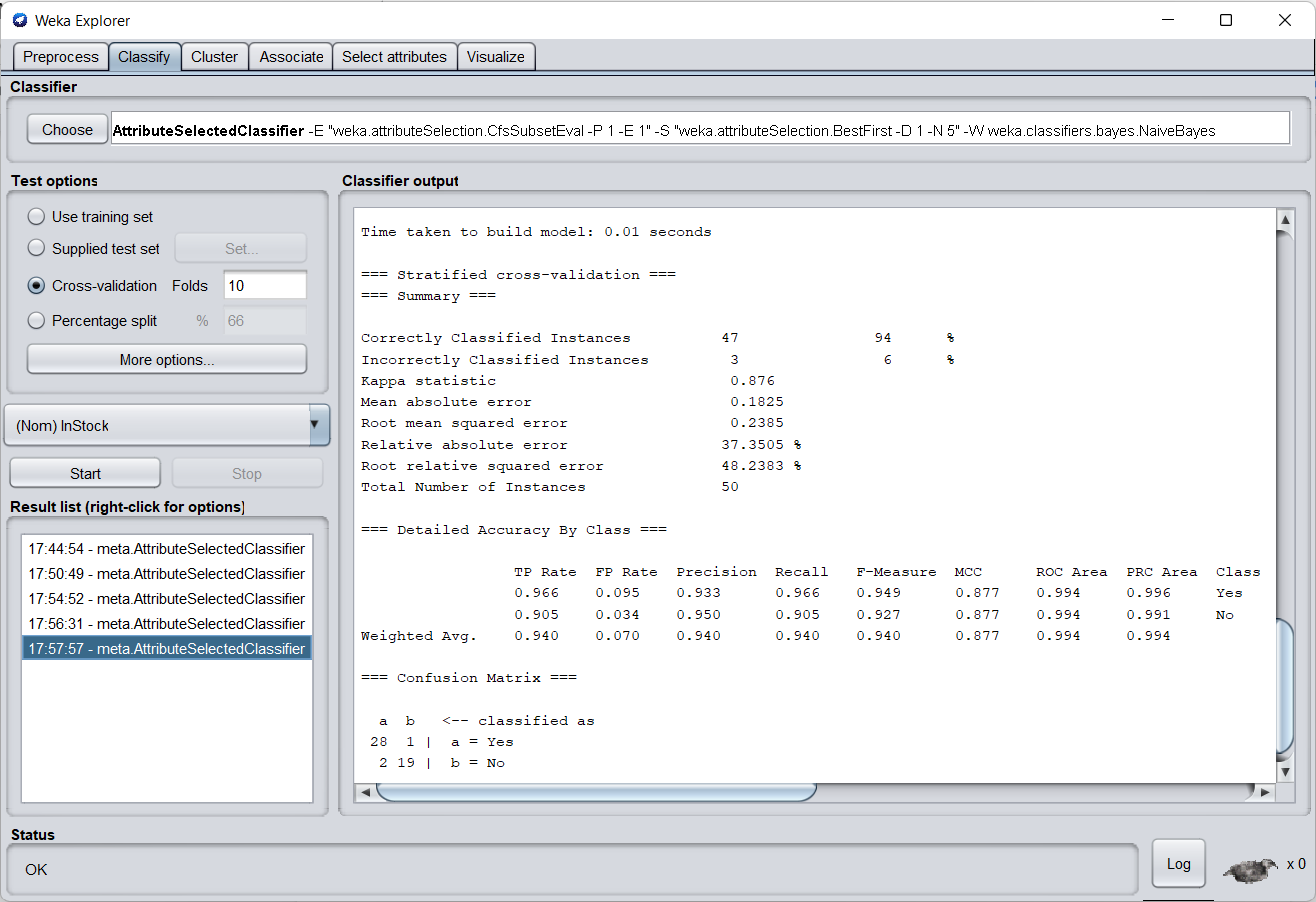
****

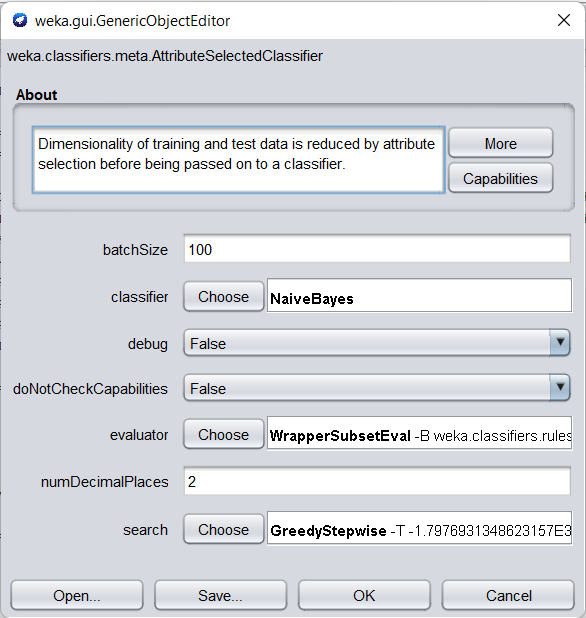
****

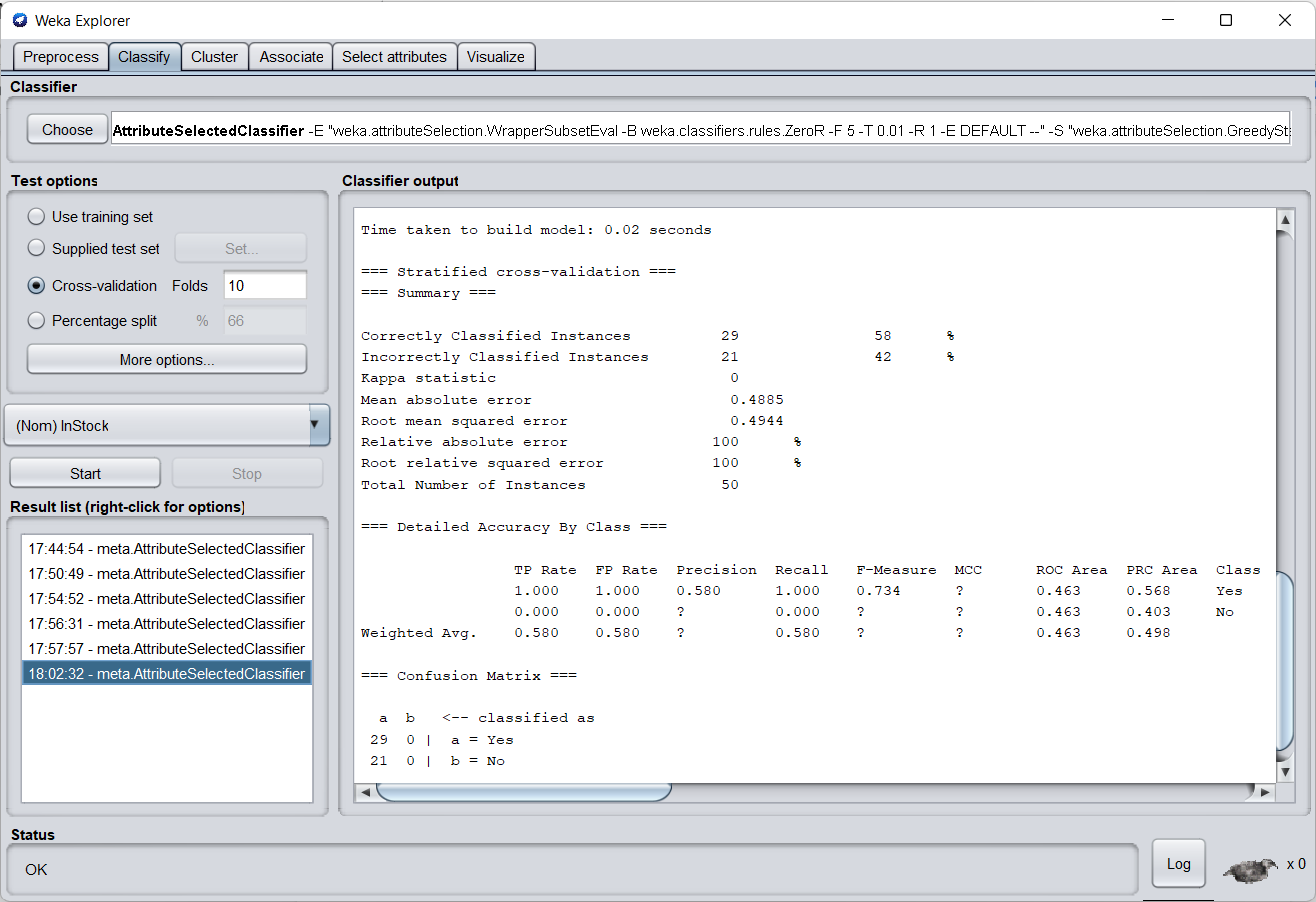
****

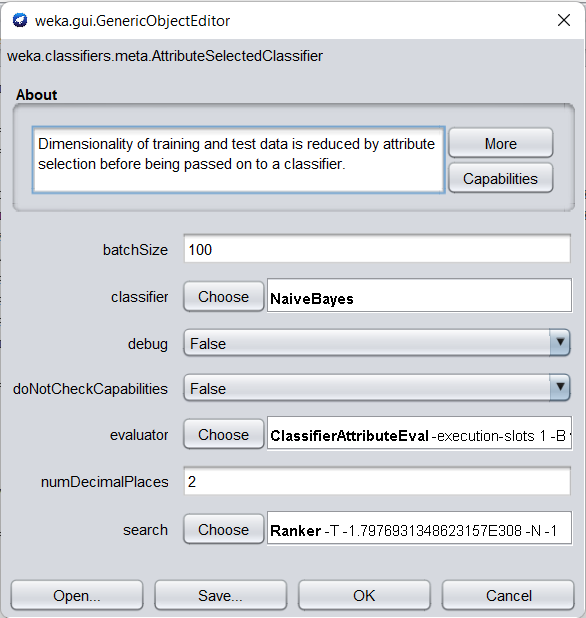
****

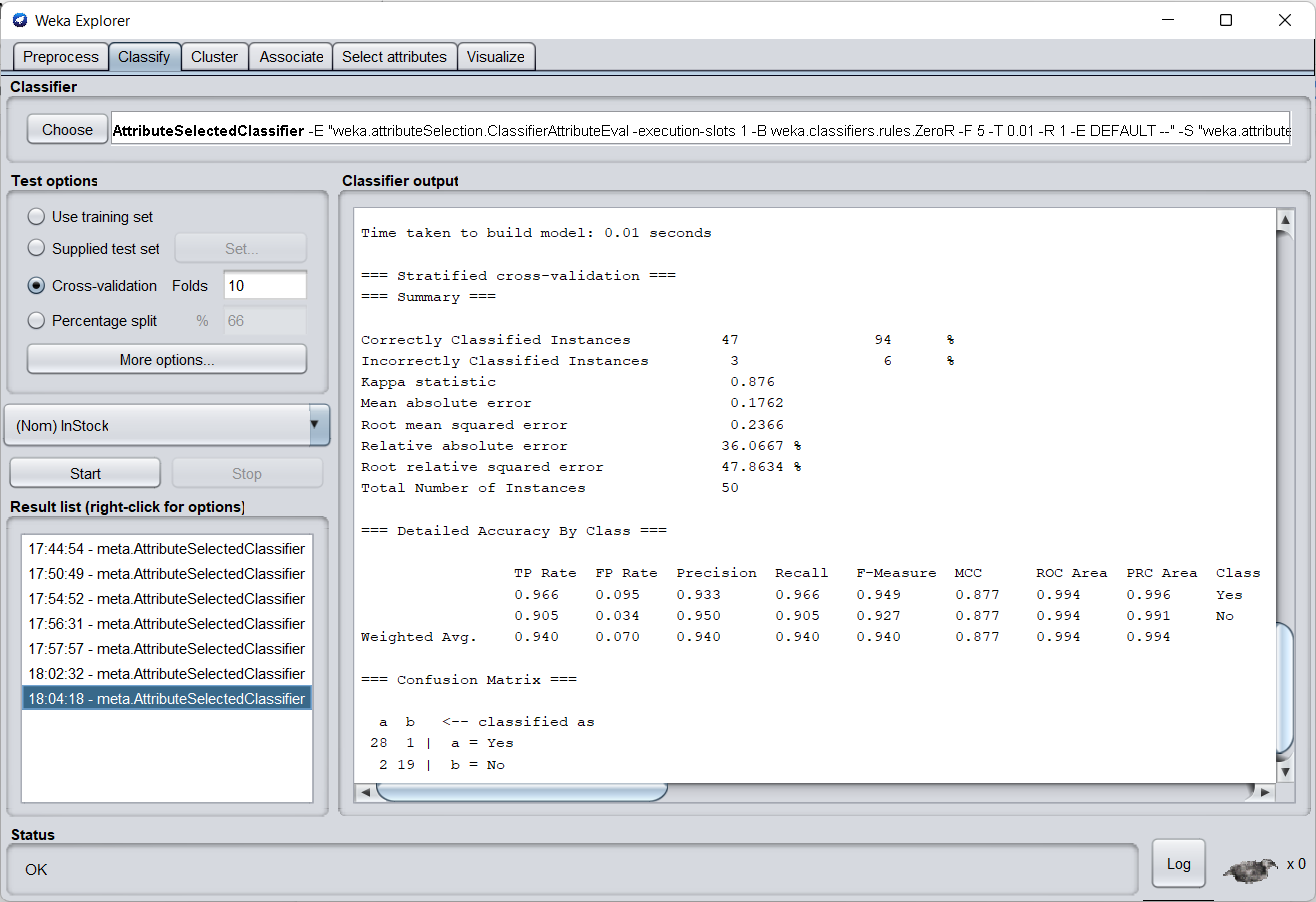
****

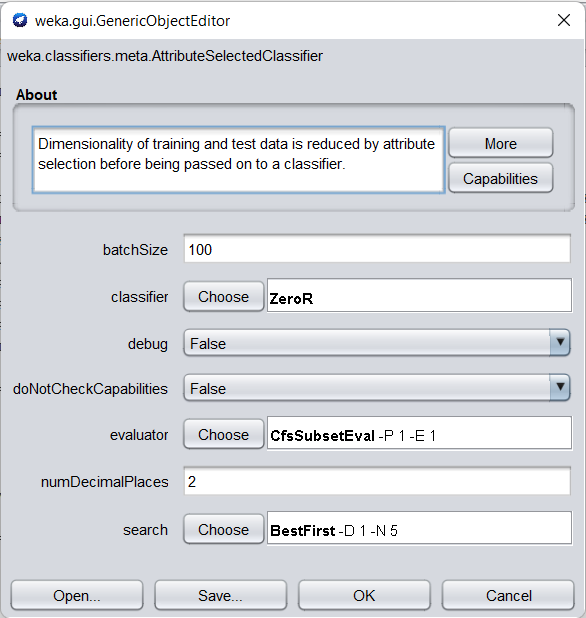
****

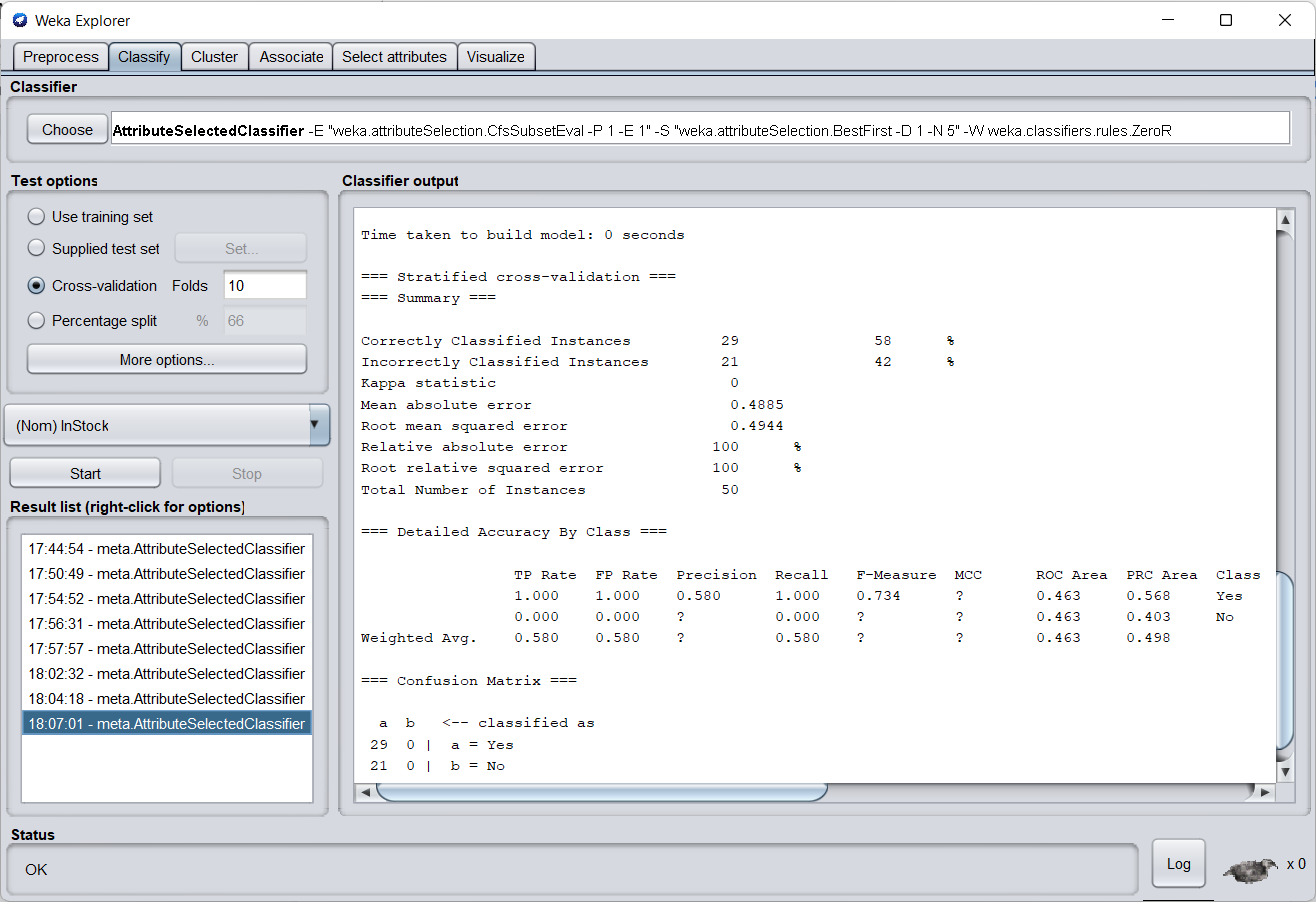
****

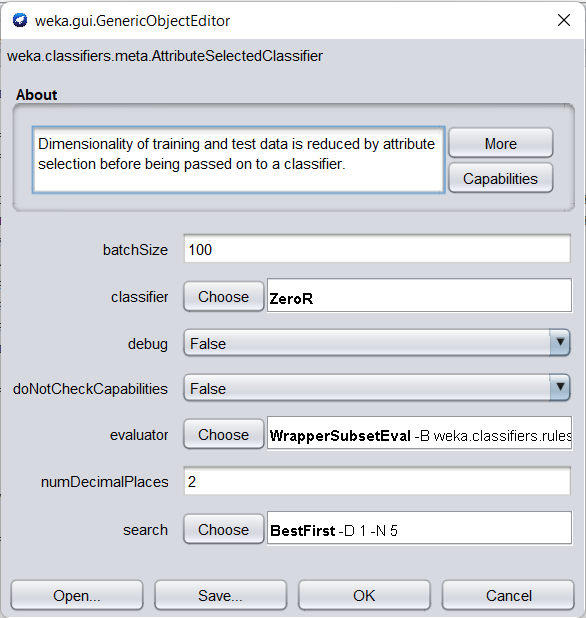
****

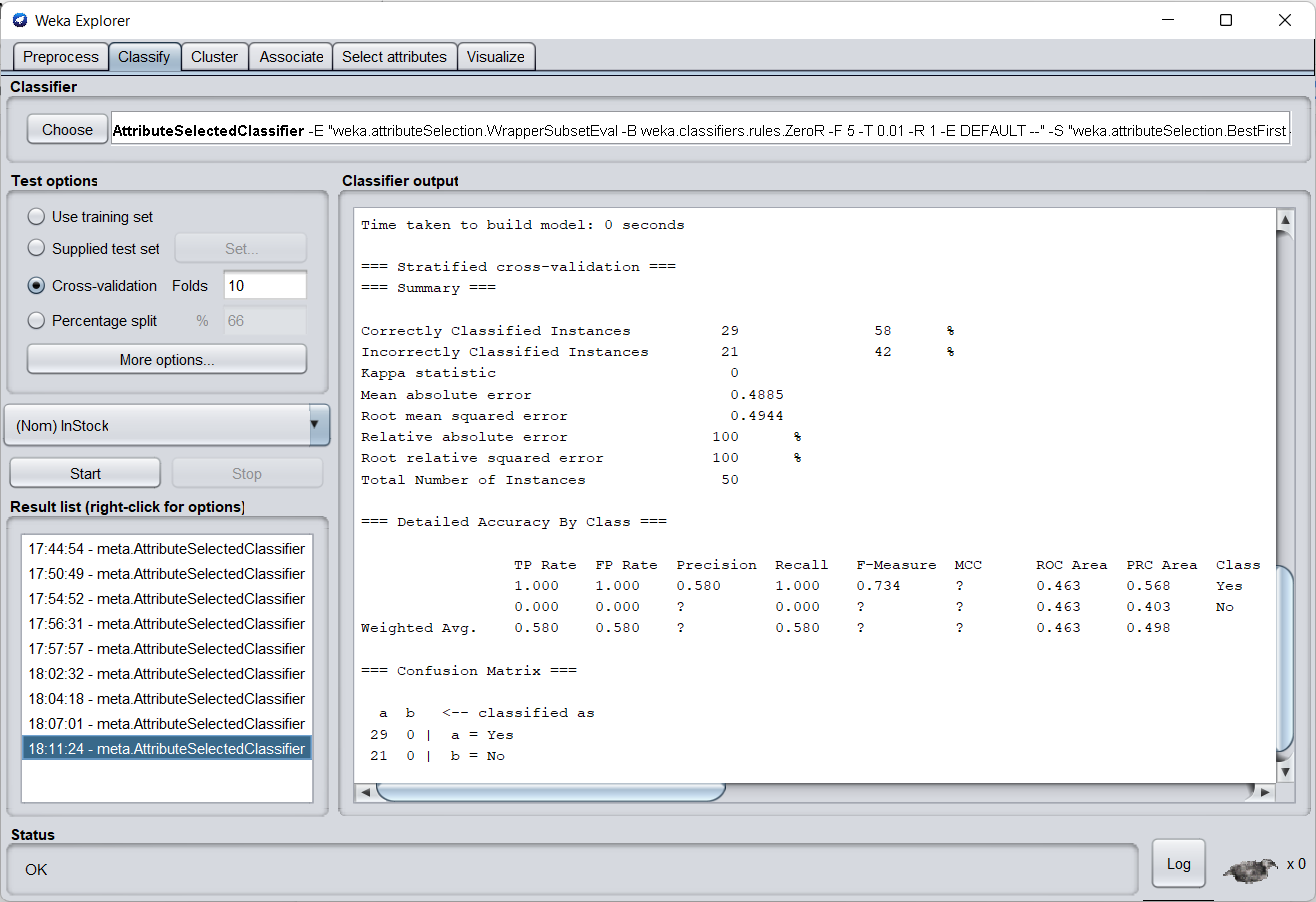
****

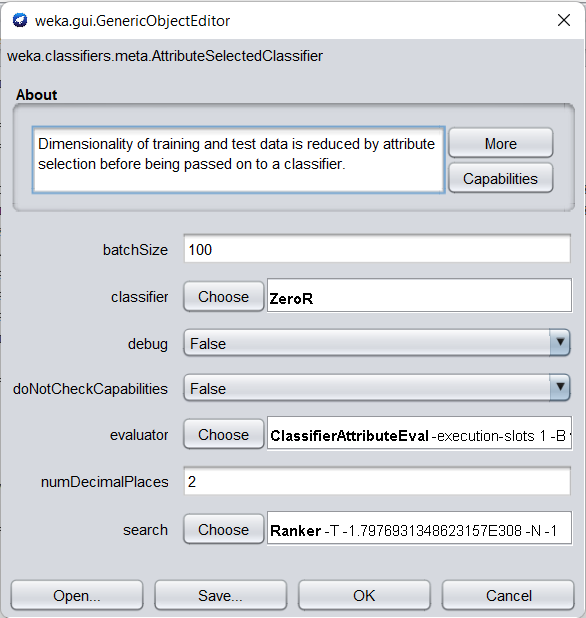
****

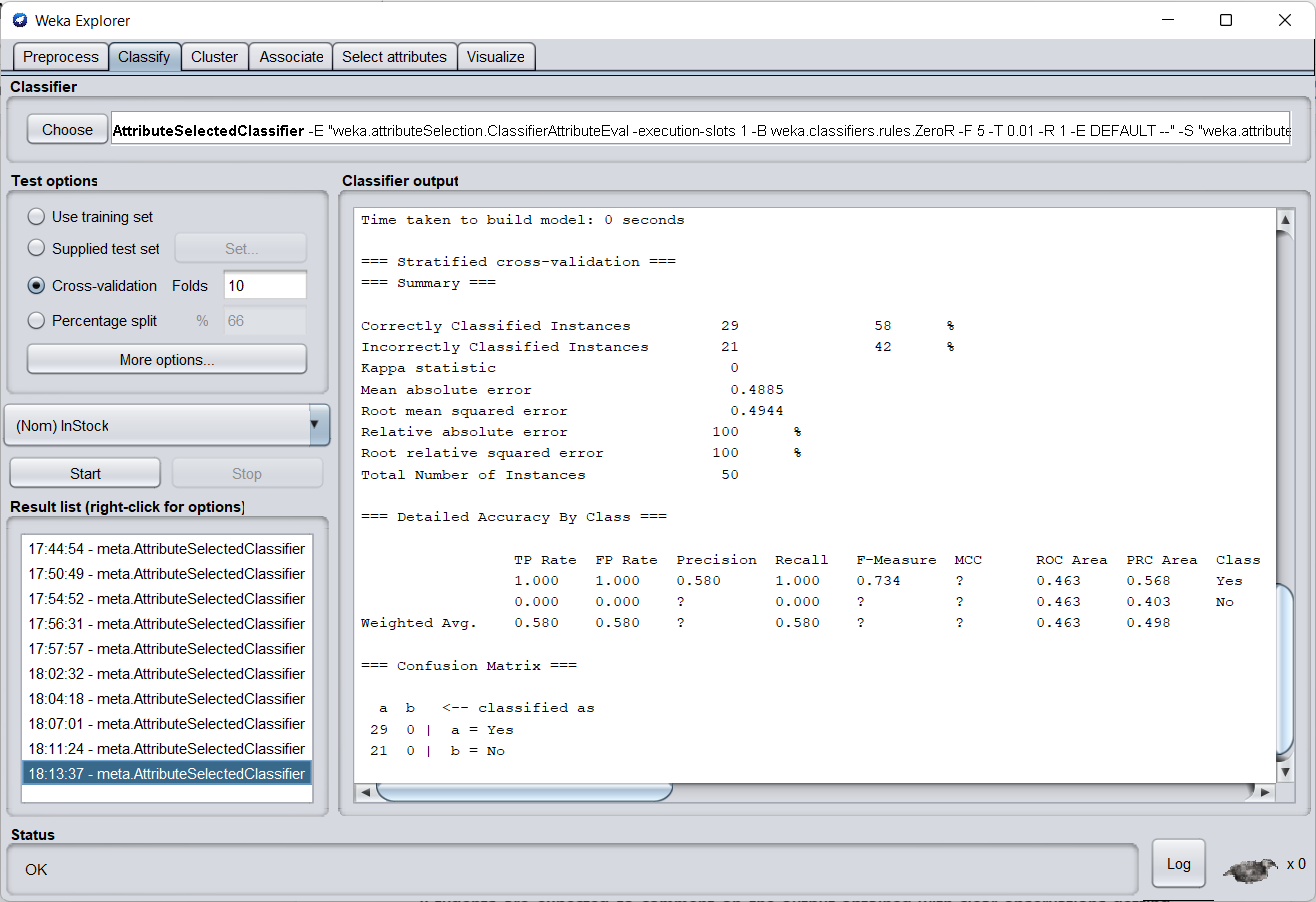
****

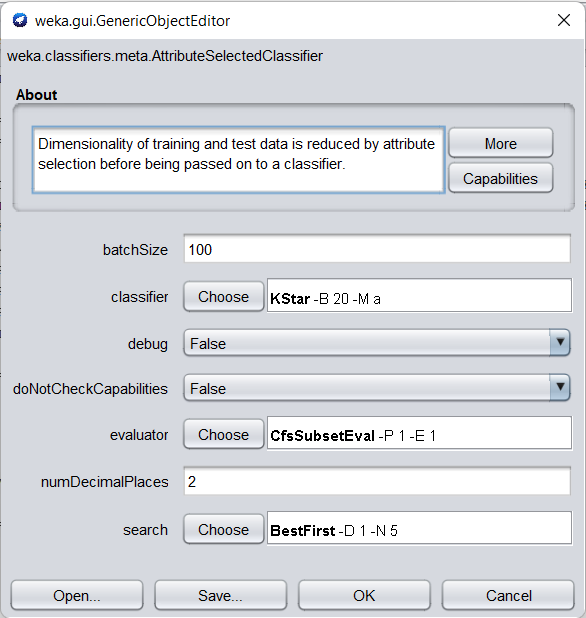
****

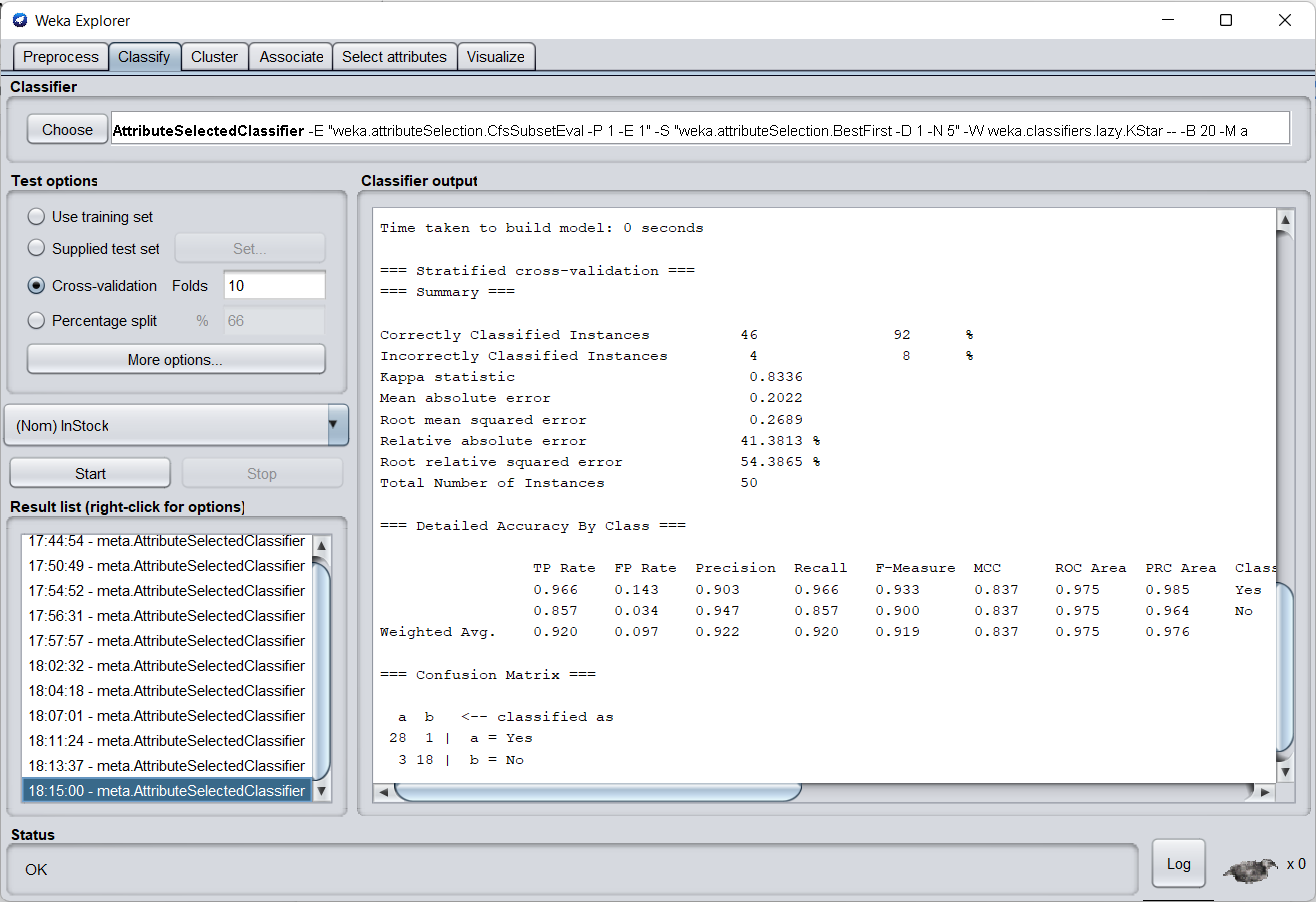
****

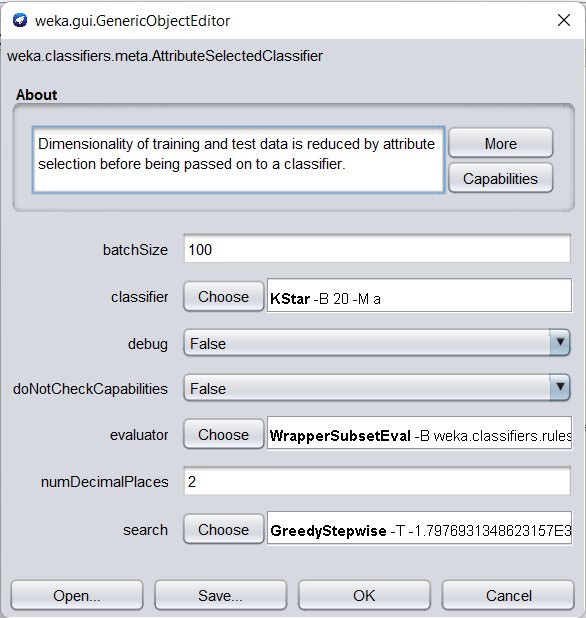
****

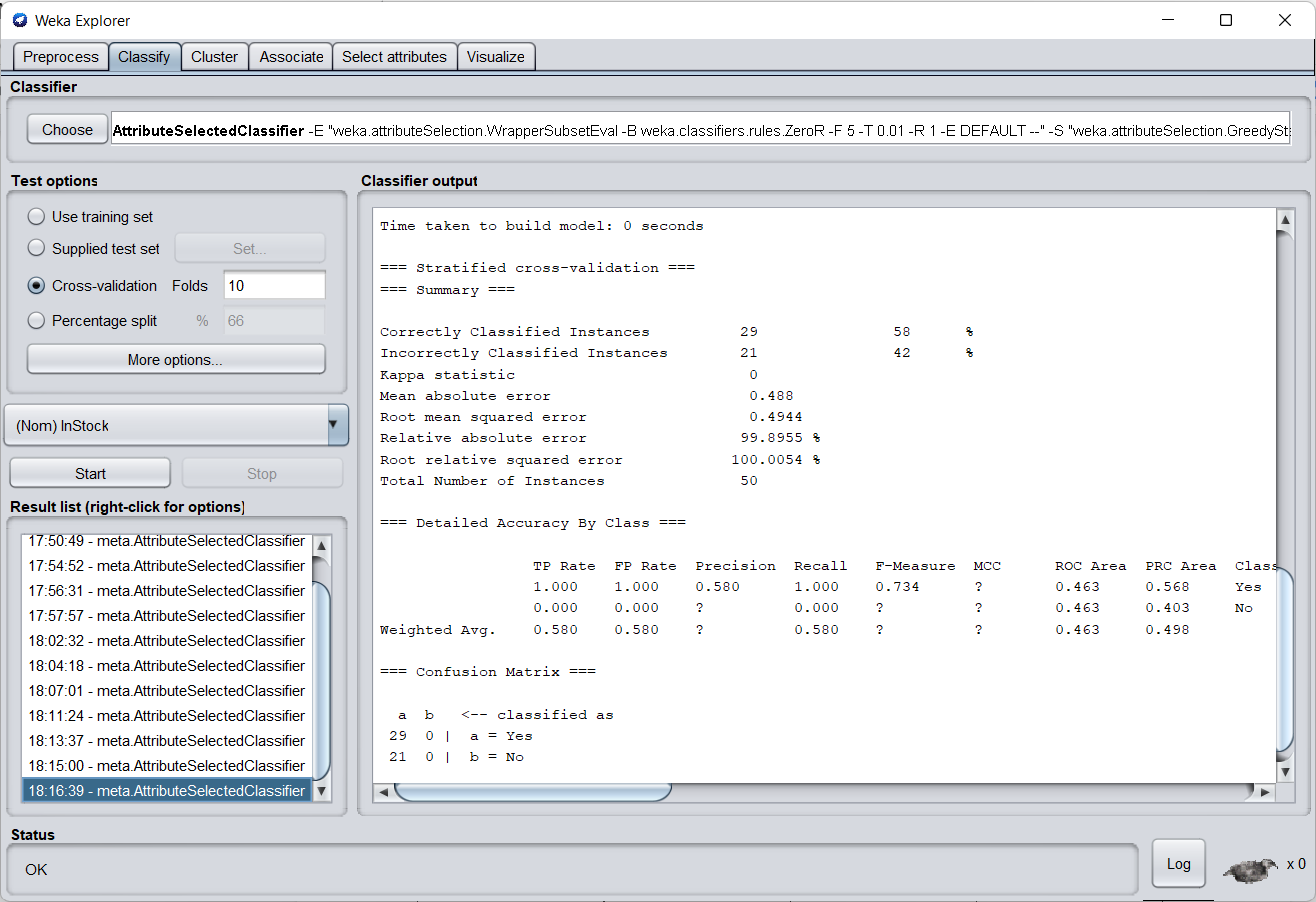
****

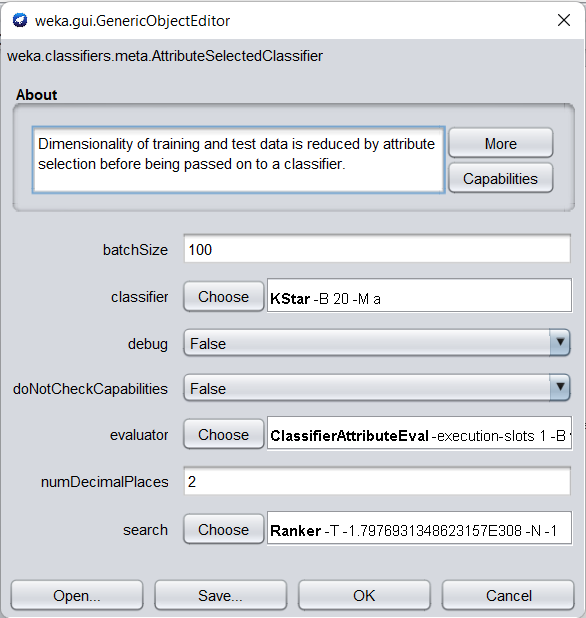
****

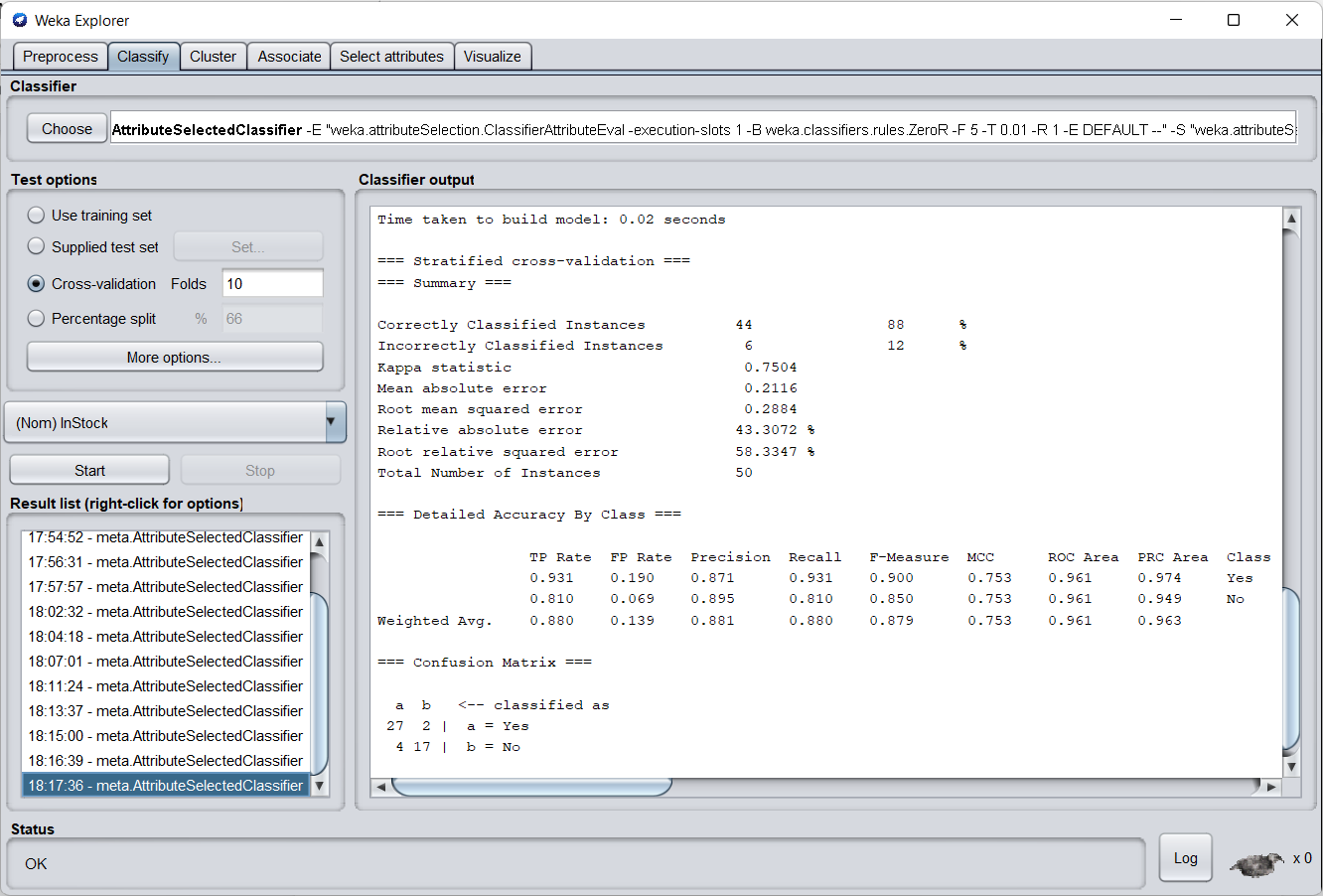
****

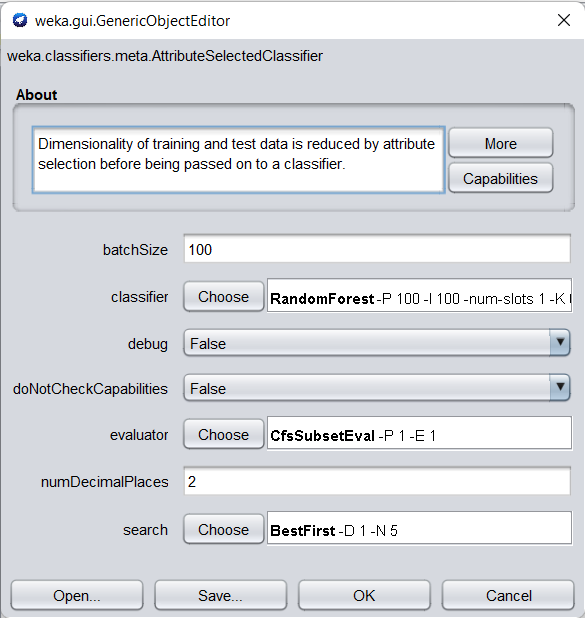
****

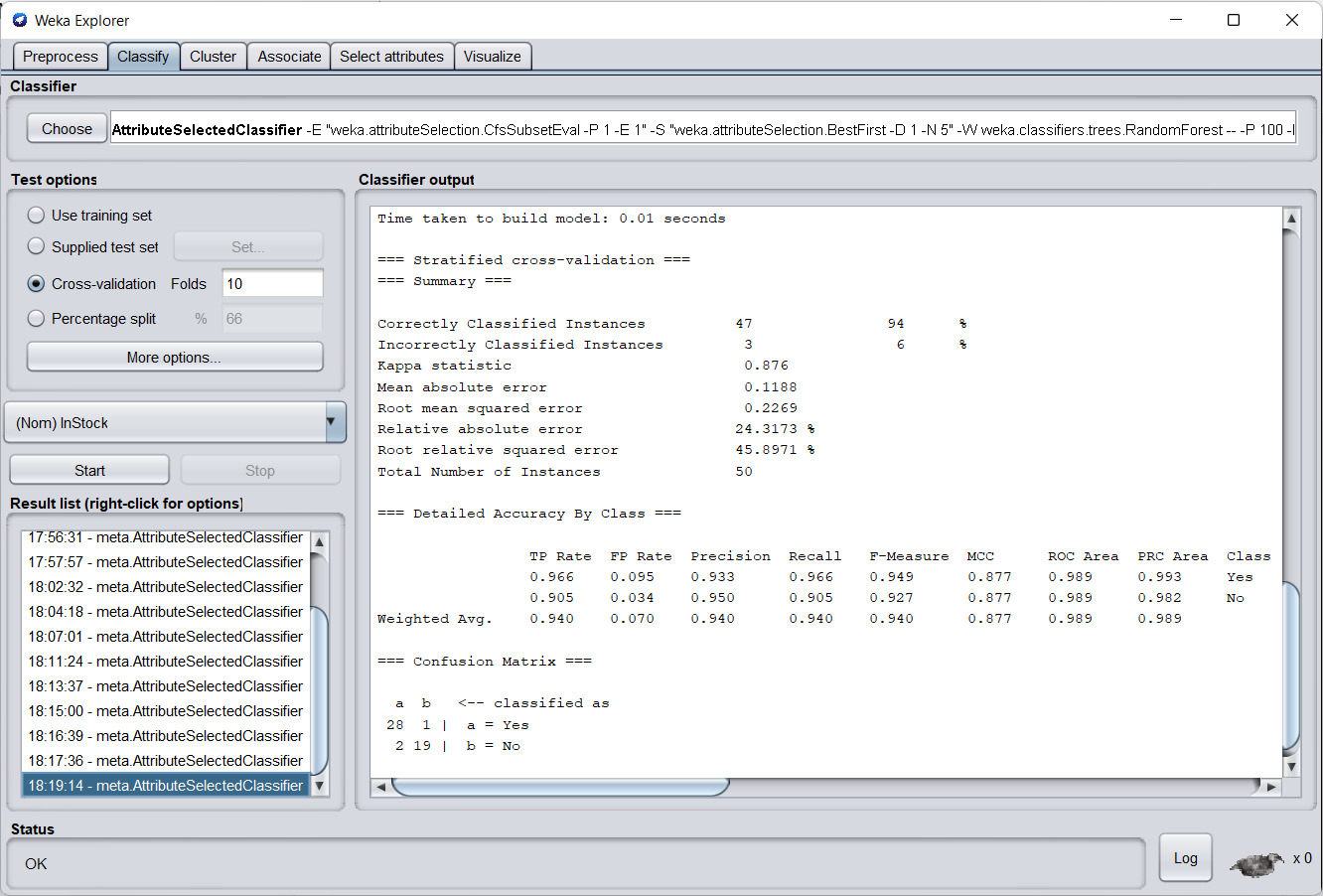
****

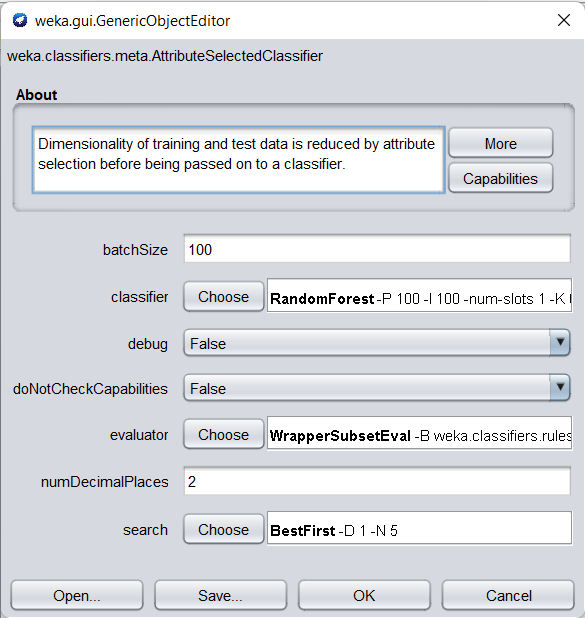
****

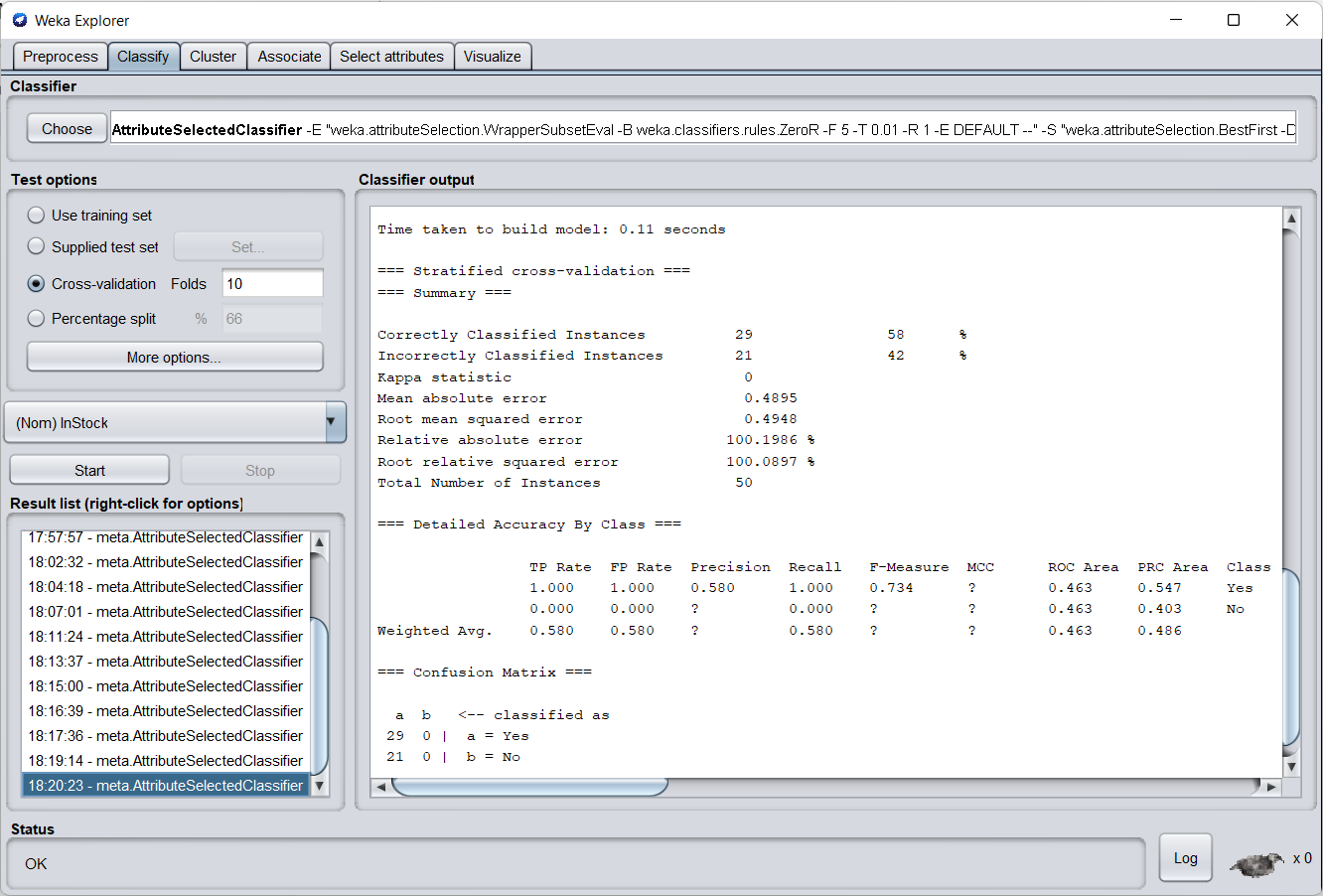
****

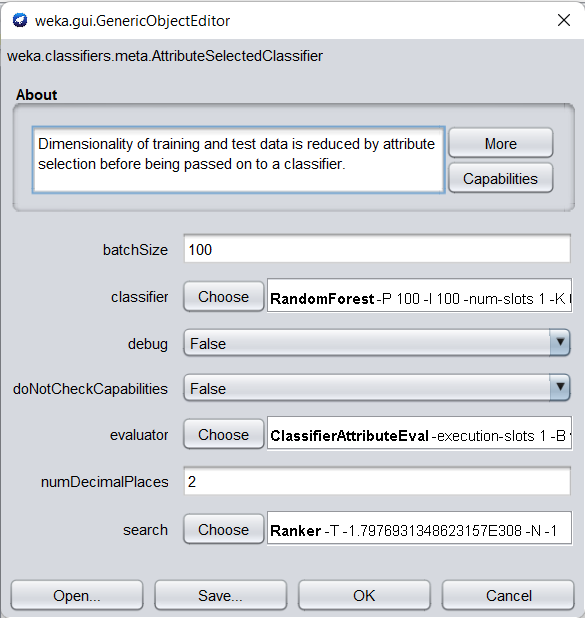
****

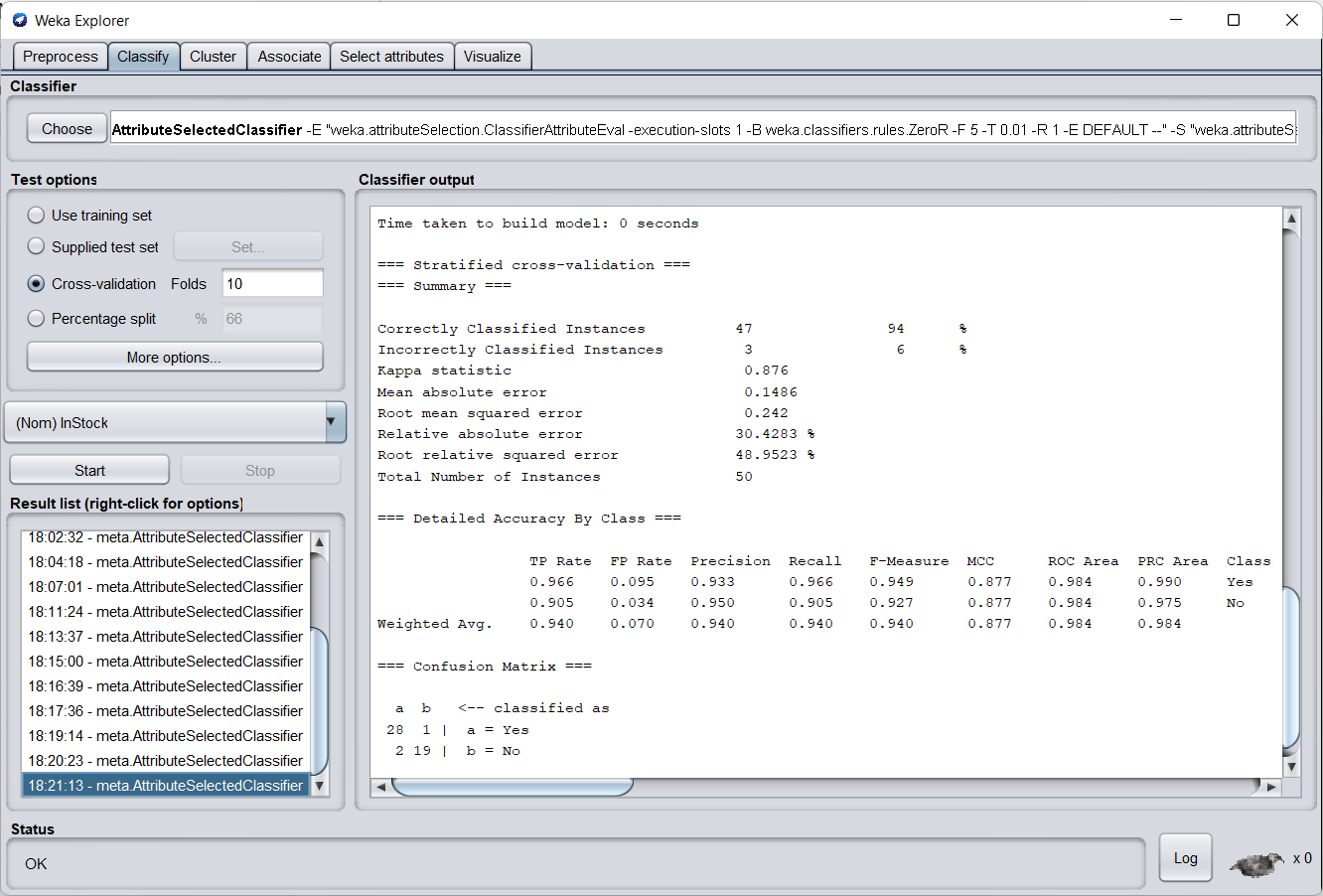
****

****

****

****

****

****

**B.3 Observations and learning: (Performance Evaluation)**

***(Students are expected to comment on the output obtained with clear observations getting from Performance Evaluation after analyzing the data and learning for each task assigned)***

| **CLASSIFIER** | **EVALUATOR** | **SEARCH METHOD** | **ACCURACY** |
| --- | --- | --- | --- |
| J48 | CFS Subset | Best First | 88% |
| J48 | Wrapper Subset | Best First | 58% |
| J48 | Wrapper Subset | Greedy Stepwise | 58% |
| J48 | Classifier Attribute | Ranker | 88% |
| Naive Bayes | CFS Subset | Best First | 94% |
| Naive Bayes | Wrapper Subset | Greedy Stepwise | 58% |
| Naive Bayes | Classifier Attribute | Ranker | 94% |
| ZeroR | CFS Subset | Best First | 58% |
| ZeroR | Wrapper Subset | Best First | 58% |
| ZeroR | Classifier Attribute | Ranker | 58% |
| KStar | CFS Subset | Best First | 92% |
| KStar | Wrapper Subset | Greedy Stepwise | 58% |
| KStar | Classifier Attribute | Ranker | 88% |
| Random Forest | CFS Subset | Best First | 94% |
| Random Forest | Wrapper Subset | Best First | 58% |
| Random Forest | Classifier Attribute | Ranker | 94% |

We can deduce the following from the given comparison:

* The least accurate classifier is the ZeroR, which is useful for defining a baseline performance as a standard for other classification algorithms.
* Greedy Stepwise Search has a lower accuracy than Best First and Ranker Search Algorithms.
* The Classifier Attribute and the CFS Subset Evaluator decide the most accuracy, while the Wrapper Subset Evaluator is accountable for the least accuracy.

**B.4 Conclusion:**

*(****Students must write the conclusion as per the attainment of individual outcome listed above and learning/observation noted in section B.3)***

As a result, we were able to successfully incorporate a variety of problem-solving techniques while also optimising the accuracy for the provided dataset.

**B.5 Question of Curiosity**

***(To be answered by student based on the practical performed and learning/observations)***

**Q1)** What are the different methods for Relevant Attribute Selection?

**Ans:**

In the data mining process, Relevant Attribute Selection is a strategy for data minimization. Data reduction decreases the size of data so that it can be used more efficiently for analysis.

Attribute Subset Selection Methods

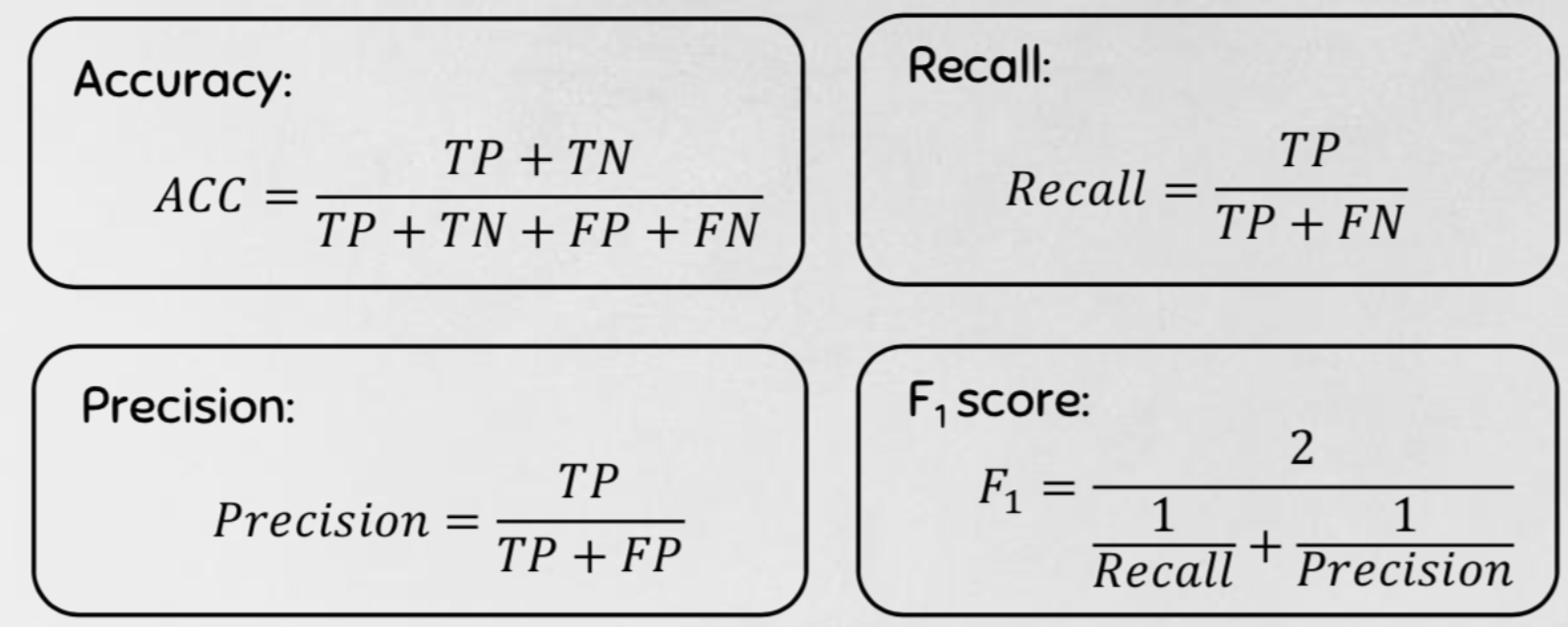
1. Stepwise Forward Selection
2. Stepwise Backward Elimination
3. Combination of Forward Selection and Backward Elimination
4. Decision Tree Induction

**Q2)** Explain Performance Evaluation Parameters for Classification Problem.

**Ans:**

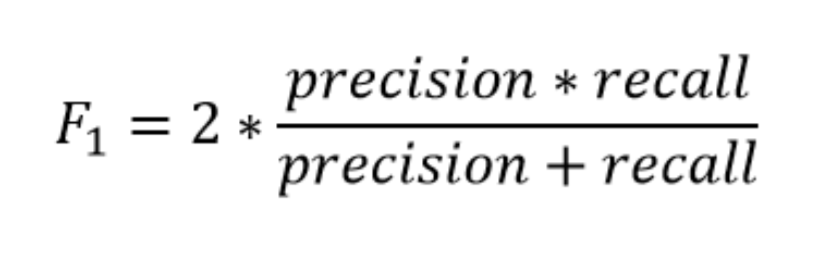
1. Confusion Matrix

The counts of test records successfully and erroneously predicted by the model are used to evaluate the performance of a classification model. The confusion matrix offers a more detailed view of not just a predictive model's performance, but also which classes are being forecasted correctly and erroneously, as well as the kind of errors that are being produced.



1. Combining Precision and Recall — F1 Score

We aim to maximise either recall or accuracy at the expense of the other measure in the three examples above. For example, we would like to reduce FN to improve recall in the case of a good or bad loan categorization. However, in instances when we wish to discover the best balance of accuracy and recall, we may use the F1 score to combine the two measures.



1. Decision Threshold & Receiver Operating Characteristic (ROC) curve

The ROC plot is a popular method of displaying the performance of a classification model. It describes the trade-off between the true positive rate (tpr) and the false positive rate (fpr) for different probability thresholds in a prediction model.

