Experiment-1.4

Student Name: Ashish Kumar **UID: 23MAI10008**
Branch: CSE AIML **Section/Group: 23MAI-1**
Semester: 01 **Date of Performance:**
Subject Name: Artificial Intelligence Lab **Subject Code: 23CSH-621**

Aim of the Experiment :

Aim of the Experiment is to explore the high dimensionality issues in the machine learning and Apply the three different feature selection techniques to the high dimensional cancer dataset downloaded from the UCI repository.

Objective of the Experiment :

Task to be done for this experiment is that we have to explore the high dimensionality issues in the machine learning. Use the high dimensional cancer dataset downloaded from the UCI repository and apply different feature selection techniques which are:

a) filtering method

b) Wrapper method

c) PCA(Principal Component Analysis) method

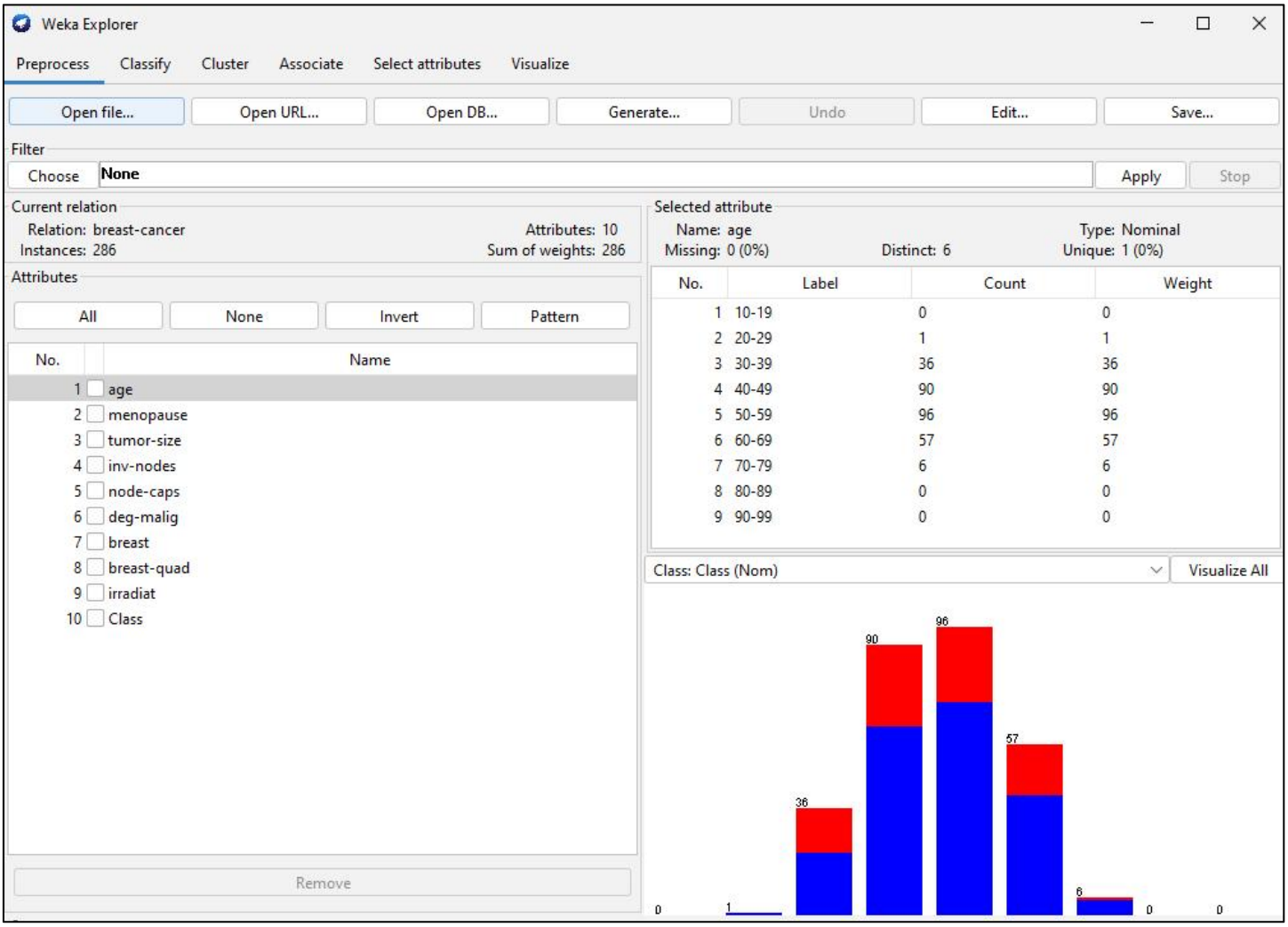
d) CorrelationAttributeEval method.

Algorithm/ Steps for Experiment :

Step 1: Download the **Cancer dataset** from UCI repository.

Step 2: Convert the csv file into **arff** file using the WEKA Tool.

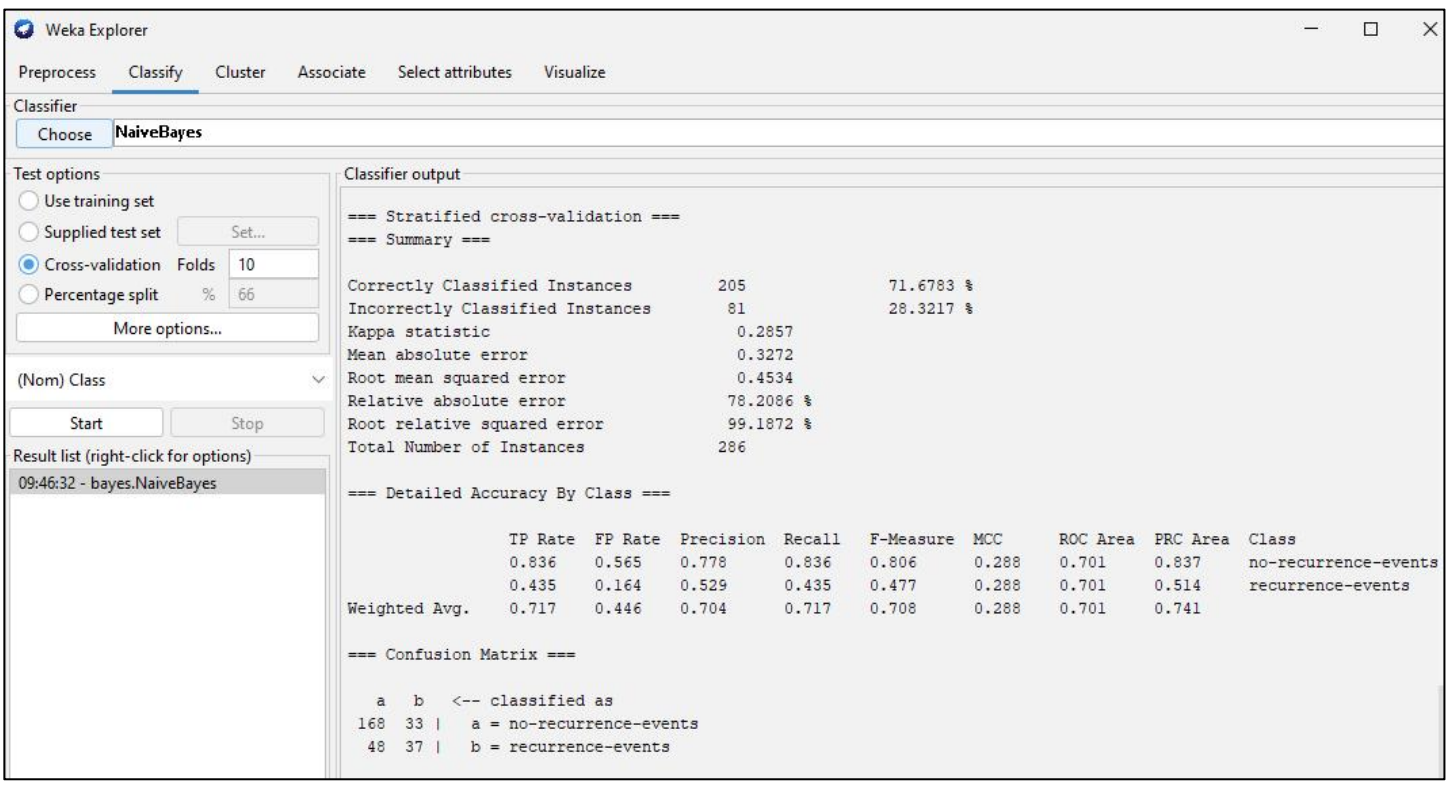
Step 3: Now open the Cancer dataset in the WEKA Tool using **'Open file'** option.



A) Wrapper Method:

1) Go to **Select Attributes tab** and in **Attribute Evaluator,** click on **'Choose'** button.
2) Select the **Classifier Subset Evaluator** from the given list.
3) Click on Classifier Subset Evaluator → A dialog box will appear.
4) In the dialog box, choose the **Naive Bayes** classifier.
5) Select the attribute **(Nom) class** and click on **'Start'** button.

Before applying attribute evaluator:



After applying attribute evaluator:

Weka Explorer

Preprocess | Classify | Cluster | Associate | Select attributes | Visualize

Attribute Evaluator

Choose: **ClassifierSubsetEval** -B weka.classifiers.bayes.NaiveBayes -T -H "Click to set hold out or test instances" -E DEFAULT

Search Method

Choose: **BestFirst** -D 1 -N 5

Attribute Selection Mode

- Use full training set
- Cross-validation Folds 10 Seed 1

(Nom) Class

Start | Stop

Result list (right-click for options)

09:57:07 - BestFirst + ClassifierSubsetEval

Attribute selection output

```
=== Attribute Selection on all input data ===

Search Method:
        Best first.
        Start set: no attributes
        Search direction: forward
        Stale search after 5 node expansions
        Total number of subsets evaluated: 58
        Merit of best subset found:    0.776

Attribute Subset Evaluator (supervised, Class (nominal): 10 Class):
        Classifier Subset Evaluator
        Learning scheme: weka.classifiers.bayes.NaiveBayes
        Scheme options:
        Hold out/test set: Training data
        Subset evaluation: classification error

Selected attributes: 1,2,4,6,7,8 : 6
                     age
                     menopause
                     inv-nodes
                     deg-malig
                     breast
                     breast-quad
```

Weka Explorer

Preprocess | Classify | Cluster | Associate | Select attributes | Visualize

Classifier

Choose: **NaiveBayes**

Test options

- Use training set
- Supplied test set Set...
- Cross-validation Folds 10
- Percentage split % 66

More options...

(Nom) Class

Start | Stop

Result list (right-click for options)

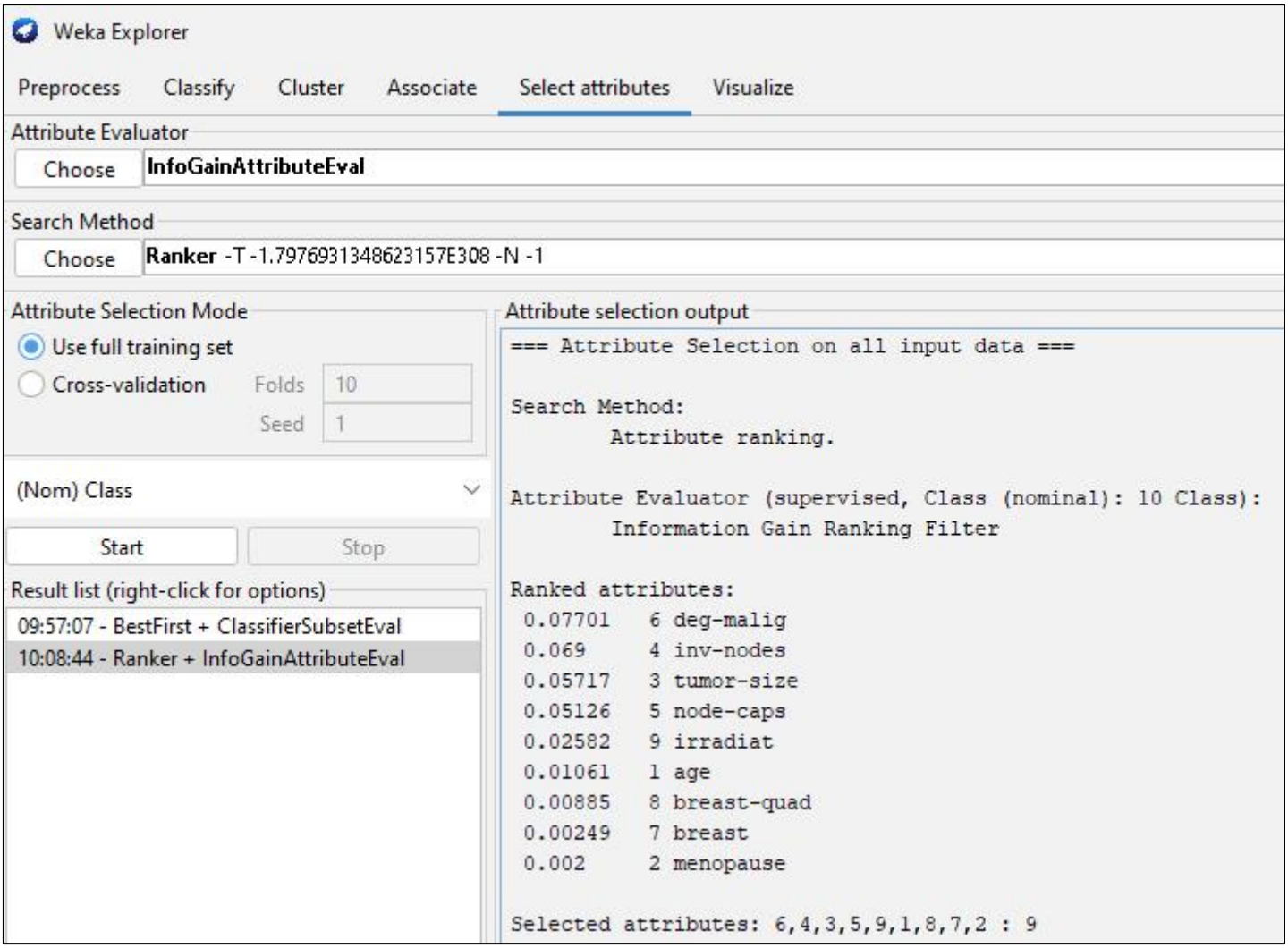
09:46:32 - bayes.NaiveBayes
10:01:50 - bayes.NaiveBayes

Classifier output

```
=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances         211               73.7762 %
Incorrectly Classified Instances        75               26.2238 %
Kappa statistic                          0.2984
Mean absolute error                      0.3505
Root mean squared error                  0.4363
Relative absolute error                 83.7621 %
Root relative squared error             95.4487 %
Total Number of Instances              286

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.891    0.624    0.772      0.891    0.827      0.312    0.686     0.797     no-recurrence-events
                 0.376    0.109    0.593      0.376    0.460      0.312    0.686     0.494     recurrence-events
Weighted Avg.    0.738    0.471    0.718      0.738    0.718      0.312    0.686     0.707

=== Confusion Matrix ===

   a   b   <-- classified as
 179  22 |   a = no-recurrence-events
  53  32 |   b = recurrence-events
```

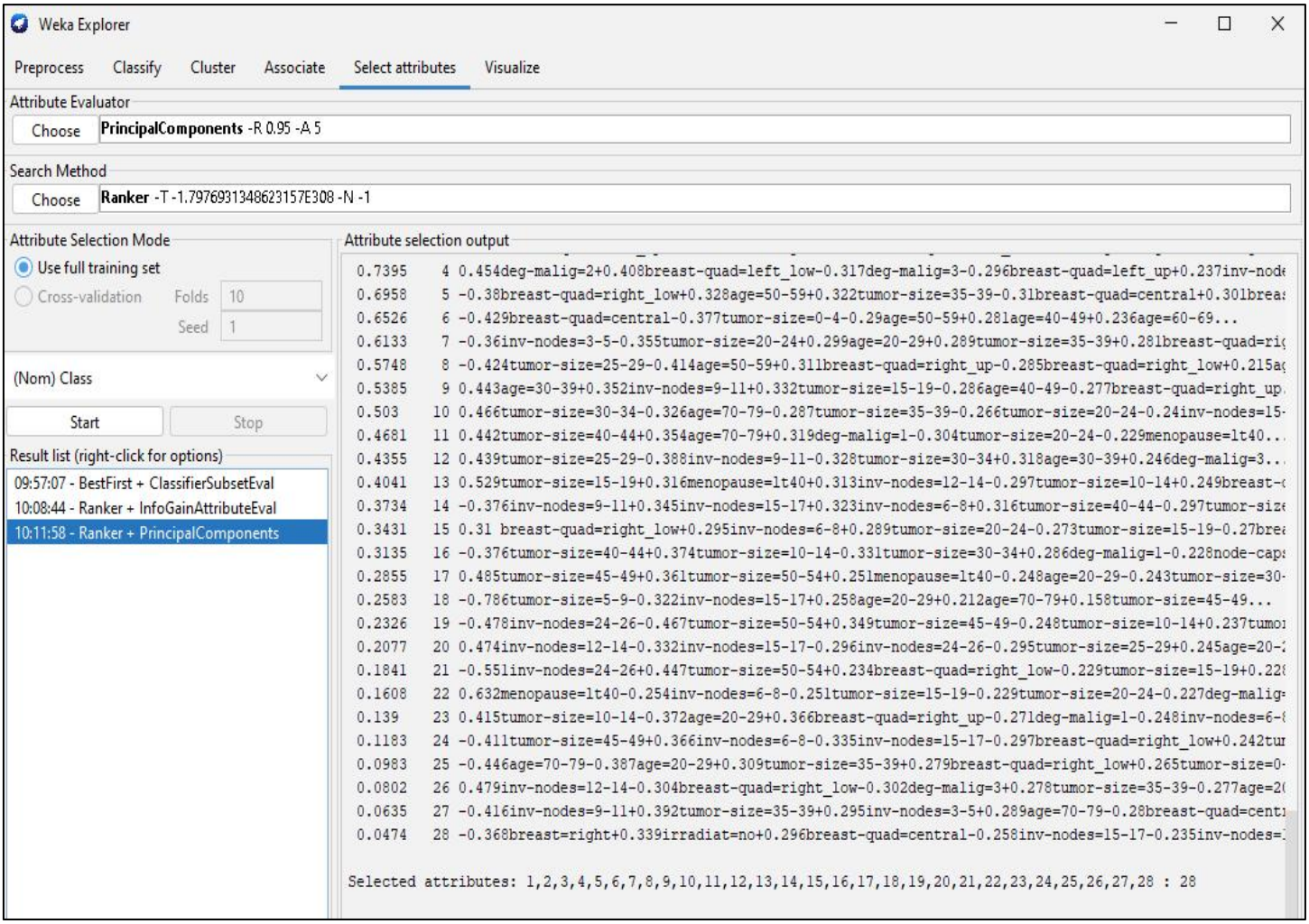
B) <u>Filtering Method:</u>

1) Go to **Select Attributes tab** and in **Attribute Evalurator**, click on **'Choose'** button.
2) Select the **InfoGainAttributeEval** from the given list.
3) Click on **Search method** → choose **Ranker** → click on Ranker → a dialog box will appear.
4) In the dialog box, keep **num to select** as -1.
5) Select the attribute **(Nom) class** and click on **'Start'** button.



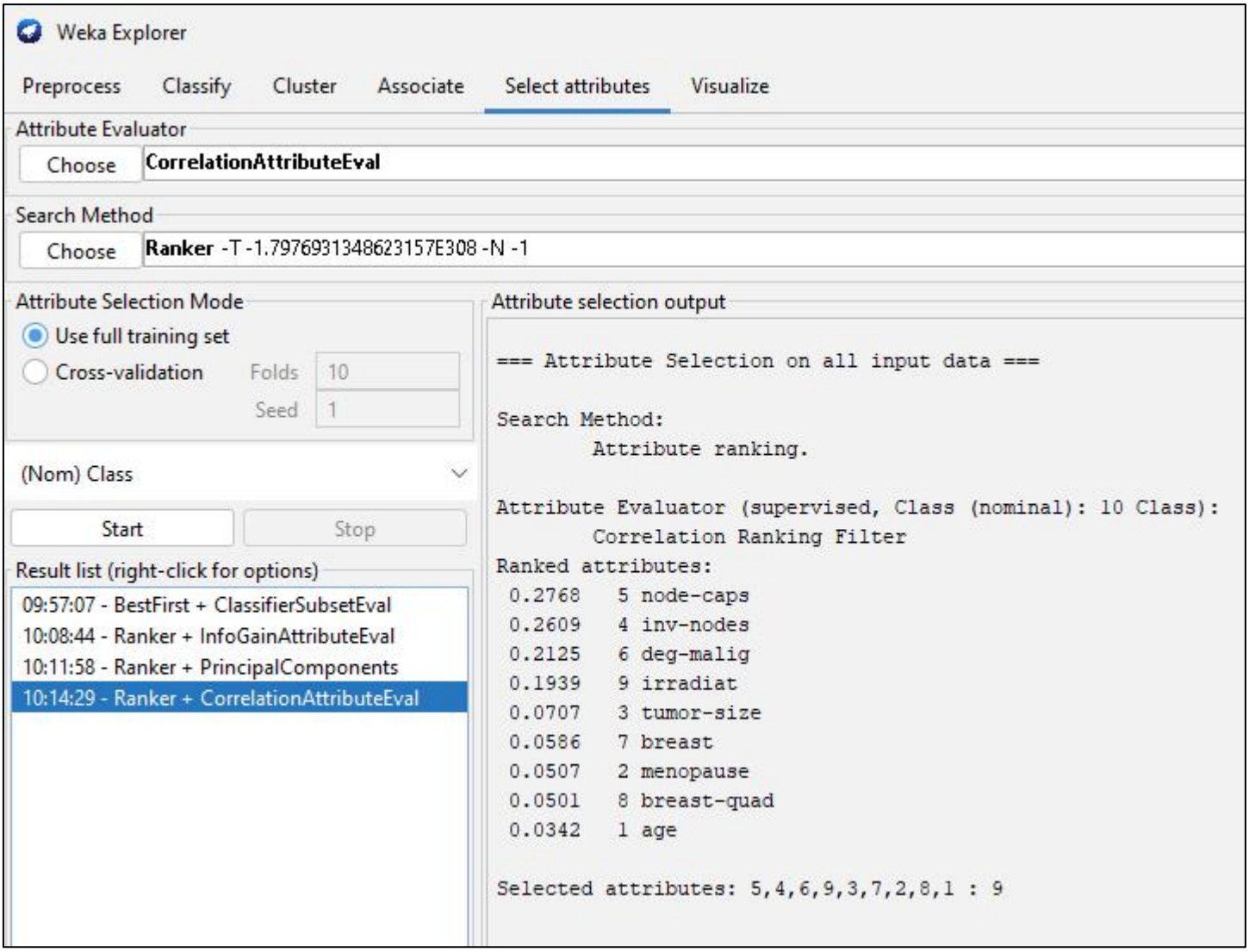
C) PCA (Principal Component Analysis):

1) Go to **Select Attributes tab** and in **Attribute Evalurator,** click on **'Choose'** button.
2) Select the **Principal Components** from the given list.
3) Click on **Search method** → choose **Ranker** → click on Ranker → a dialog box will appear.
4) In the dialog box, keep **num to select** as -1.
5) Select the attribute **(Nom) class** and click on **'Start'** button.



D) <u>Correlation Attribute Evaluation:</u>

1) Go to **Select Attributes tab** and in **Attribute Evalurator,** click on **'Choose'** button.
2) Select the **CorrelationAttributeEval** from the given list.
3) Click on **Search method** → choose **Ranker** → click on Ranker → a dialog box will appear.
4) In the dialog box, keep **num to select** as -1.
5) Select the attribute **(Nom) class** and click on **'Start'** button.



Learning outcomes (What I have learnt):

1. I learnt about the WEKA Tool and its applications.
2. I learnt about how to create dataset in .arff format.
3. I learnt about different feature selection techniques in WEKA Tool.
4. I learnt about filtering method and wrapper method in WEKA.
5. I learnt about principal component analysis and correlation attribute in WEKA.