



DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

Discover. Learn. Empower.

Experiment – 1.1

Student Name: Yuvraj Singh Tanwar

UID: 23MAI10011

Branch: CSE

Section/Group: AIML

Semester: 1

Date of Performance: 23Aug2023

Subject Name: ARTIFICIAL INTELLIGENCE LAB

Code: 23CSH-621.

Aim:

Explore various tabs of Data Mining Tool WEKA and create the below Data Table.

Name	ID	Gender	Phone No.	Height	Weight	HB	Blood Pressure	Thyroid
Rahul	1	M	9427902142	5.8	78	12	Normal	Yes
Mohan	2	M	5427904077	6.1	81	14	High	No
Priya	3	F	4427906072	5.5	67	13	Low	Yes
Reema	4	F	8887907071	5.4	55	10	Normal	No
Misha	5	F	2227909076	5.3	50	12	Low	No
Kiran	6	F	9929907077	5.4	45	13	Low	No
Amar	7	M	4427906071	5.7	60	15	Normal	No

Task to be done:

- Installation of WEKA tool
- Create a Data Table

Procedure:

To install WEKA on your machine, visit WEKA's official website and download the installation file. WEKA supports installation on Windows, Mac OS X and Linux. You just need to follow the instructions on this page to install WEKA for your OS.

The steps for installing on Mac are as follows –

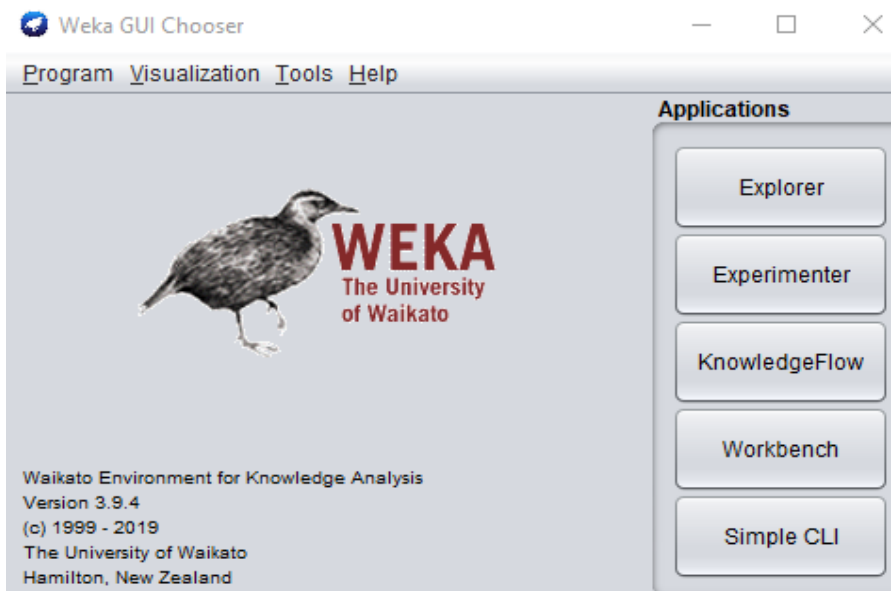
- Download the Mac installation file.
- Double click on the downloaded weka-3-8-3-corretto-jvm.dmg file.

You will see the following screen on successful installation.



- Click on the weak-3-8-3-corretto-jvm icon to start Weka.
- Optionally you may start it from the command line – `java -jar weka.jar`

The WEKA GUI Chooser application will start and you would see the following screen –

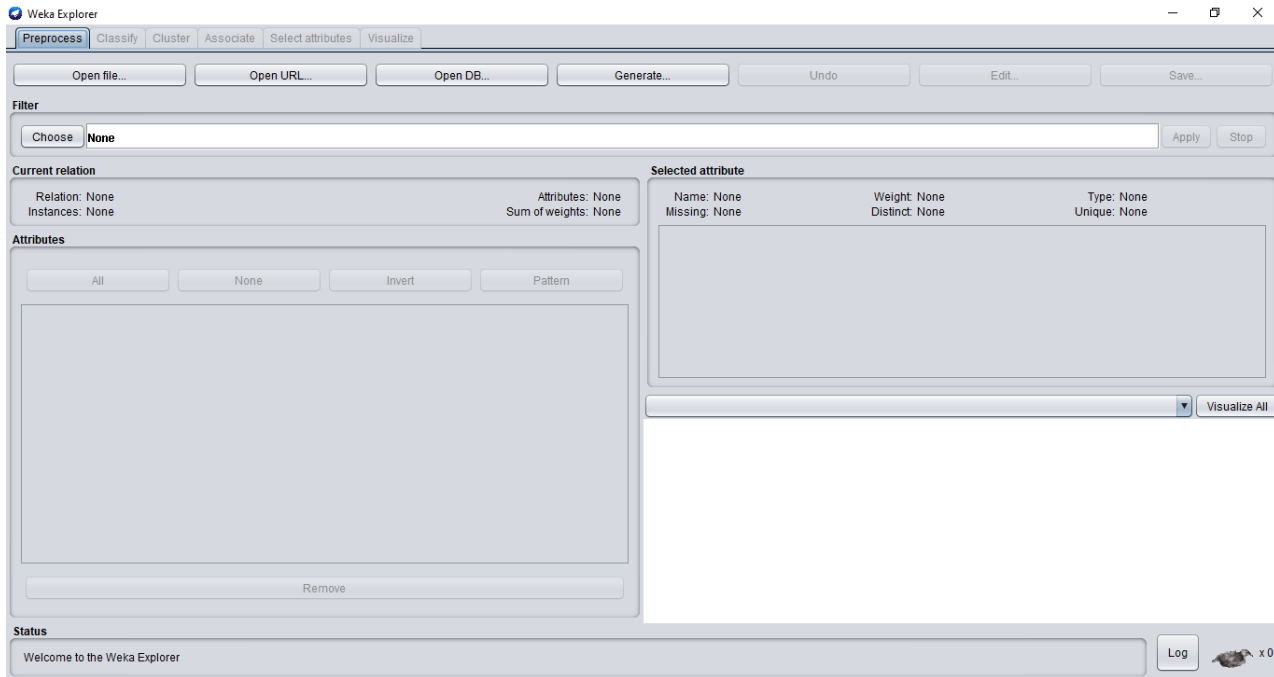


The GUI Chooser application allows you to run five different types of applications as listed here

- Explorer
- Experimenter
- Knowledge Flow
- Workbench
- Simple CLI

We will be using Explorer in this tutorial

When you click on the Explorer button in the Applications selector, it opens the following screen –



On the top, you will see several tabs as listed here –

- Pre-process
- Classify
- Cluster
- Associate
- Select Attributes
- Visualize

Under these tabs, there are several pre-implemented machine learning algorithms. Let us look into each of them in detail now.

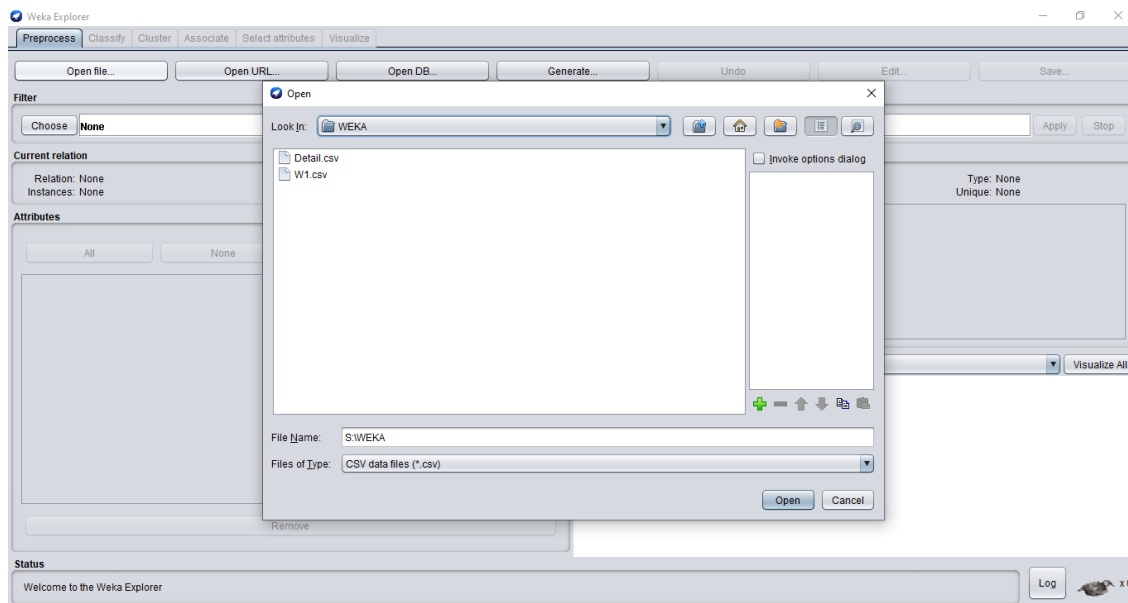
Exploring various tabs of Data Mining tool WEKA:

- First a file of .arff or .csv format is opened using the open file tab.
- The preprocess helps in processing the data, it also provides various filters for preprocessing the data.
- The classify option contains various algorithms of classification and regression.
- Cluster contains different unsupervised learning techniques.
- Association contains various association rule mining algorithms.
- Select attributes is basically feature selection, which selects the best features.
- Visualize is for the visualization of data.

Arff file format for the given health data is shown below:

```
Name ,ID,Gender,Phone no.,Height,Weight,HB,BP,Thyrod
ABC,1,m,8787,5.8,67,12,Normal,yes
BCA,2,m,4566,5.2,64,14,High,no
DEF,3,m,4612,5.9,62,10,Normal,no
FED,4,f,4895,5,52,13,Low,no
QWE,5,f,7895,4.8,58,12,Low,yes
EWQ,6,f,6575,5.2,60,15,Normal,no
GHY,7,m,1652,5.6,85,10,Normal,no
YGH,8,f,7894,5.1,58,13,High,no
```

You can also open using the csv file



After making a arff format file, we open it in the WEKA tool using the open file option



DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

Discover. Learn. Empower.

Weka Explorer

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

Open file... Open URL... Open DB... Generate... Undo Edit... Save...

Filter: Choose **None** Apply Stop

Current relation: Relation: health Instances: 8 Attributes: 9 Sum of weights: 8

Attributes: All None Invert Pattern

No.	Name
1	<input checked="" type="checkbox"/> Name
2	<input type="checkbox"/> ID
3	<input type="checkbox"/> Gender
4	<input type="checkbox"/> Phone no.
5	<input type="checkbox"/> Height
6	<input type="checkbox"/> Weight
7	<input type="checkbox"/> HB
8	<input type="checkbox"/> BP
9	<input type="checkbox"/> Thyrod

Remove

Status: OK Log x 0

Selected attribute: Name: Name Missing: 0 (0%) Distinct: 8 Type: Nominal Unique: 8 (100%)

No.	Label	Count	Weight
3	DEF	1	1
4	FED	1	1
5	QWE	1	1
6	EWQ	1	1
7	GHY	1	1
8	YGH	1	1

Class: Thyrod (Nom) Visualize All

Learning outcomes (What I have learnt):

We know how to install the WEKA tool.

We know how to open file in WEKA tool.



DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

Discover. Learn. Empower.

Experiment - 1.2

Student Name: Yuvraj Singh Tanwar

UID: 23MAI10011

Branch: CSE

Section/Group: AIML

Semester: 1

Date of Performance: 23Aug2023

Subject Name: ARTIFICIAL INTELLIGENCE LAB

Code: 23CSH-621.

Aim:

Train and test the classification performance of Tree based Supervised Learning Algorithms for the following classifiers: a. Decision Tree b. Neural Network c. Naïve Bayes d. Random Forest Analyze the difference in the performance of the classifiers using different parameters of Confusion Matrix

Name	ID	Gender	Phone no.	Height	Weight	HB	BP	Thyrod
ABC		1 m	8787	5.8	67	12	Normal	yes
BCA		2 m	4566	5.2	64	14	High	no
DEF		3 m	4612	5.9	62	10	Normal	no
FED		4 f	4895	5	52	13	Low	no
QWE		5 f	7895	4.8	58	12	Low	yes
EWQ		6 f	6575	5.2	60	15	Normal	no
GHY		7 m	1652	5.6	85	10	Normal	no
YGH		8 f	7894	5.1	58	13	High	no

Task to be done:

- Import the data table in WEKA tool
- Train and test the classification performance of various classifiers such as decision tree, neural network, Naïve Bayes, random forest

Procedure:

First open the data file (csv format or .arff format) that you have created in the previous experiment and import that in the weka tool using open file option.

After you have imported the file go to classify section of the weka tool and perform the other classification algorithms :-

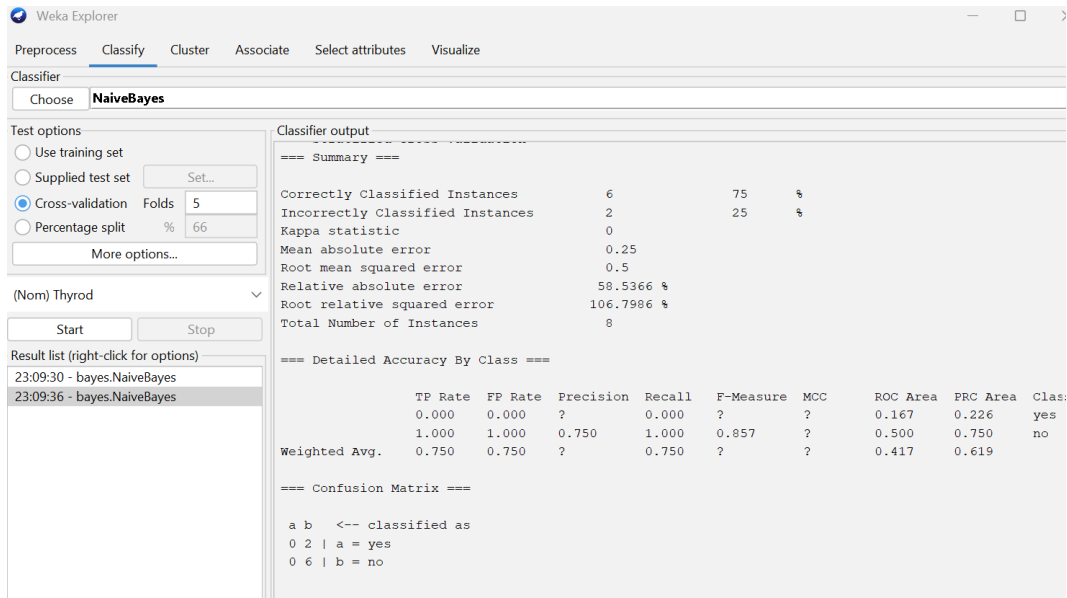
Naïve Bayes: -

- First go to classify section and select choose option over there.

- You will get various options of classifiers over there, under Bayes section select Naïve Bayes.
- Set k-fold as 5.
- Select start option.
- Check the confusion matrix obtained for Naïve Bayes

Decision Tree: -

- First go to classify section and select choose option over there.
- You will get various options of classifiers over there, under tree section select J48.
- Set k-fold as 5.
- Select start option
- Check the confusion matrix obtained for Decision Tree



The screenshot shows the Weka Explorer interface with the 'Classify' tab selected. The 'Classifier' dropdown is set to 'NaiveBayes'. Under 'Test options', 'Cross-validation' is selected with 'Folds' set to 5. The 'Start' button has been clicked, and the 'Classifier output' pane displays the following results:

```

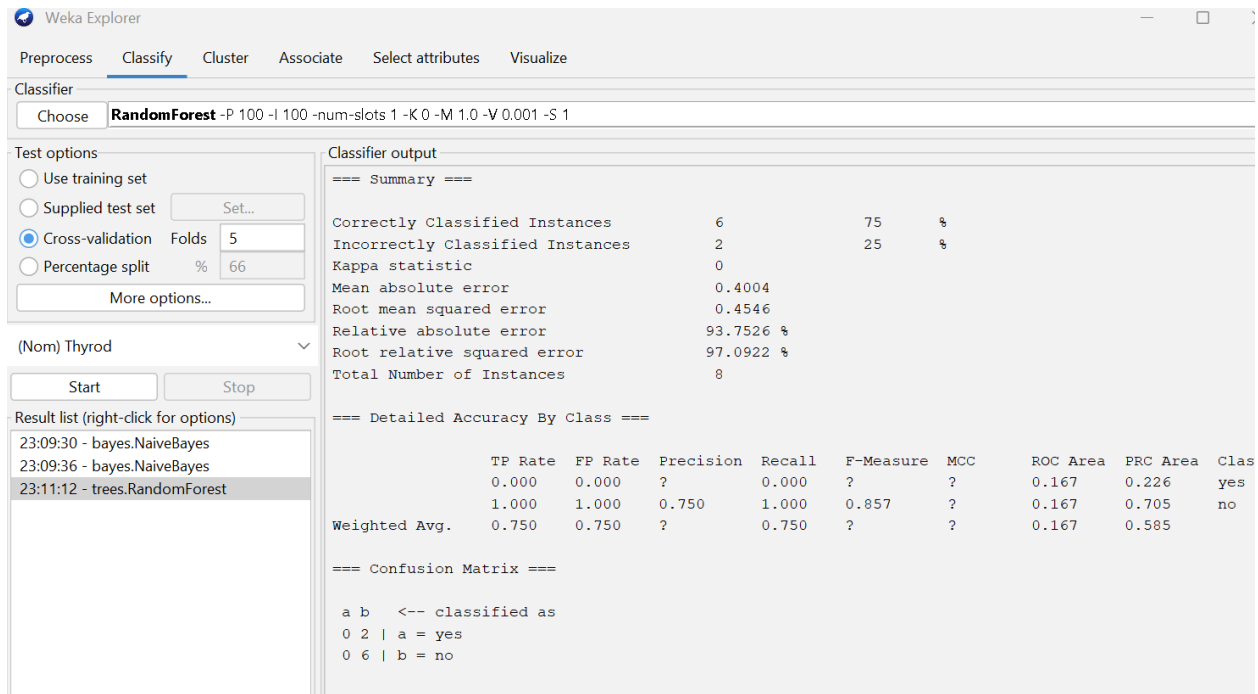
==== Summary ====
Correctly Classified Instances      6           75  %
Incorrectly Classified Instances    2           25  %
Kappa statistic                     0
Mean absolute error                 0.25
Root mean squared error             0.5
Relative absolute error             58.5366 %
Root relative squared error        106.7986 %
Total Number of Instances          8

==== Detailed Accuracy By Class ====
              TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Clas
0.000    0.000    ?      0.000    ?      ?      0.167    0.226    yes
1.000    1.000    0.750    1.000    0.857    ?      0.500    0.750    no
Weighted Avg.    0.750    0.750    ?      0.750    ?      ?      0.417    0.619

==== Confusion Matrix ====
a b  <-- classified as
0 2 | a = yes
0 6 | b = no
  
```

Random Forest: -

- First go to classify section and select choose option over there.
- You will get various options of classifiers over there, under tree section select Random Forest.
- Set k-fold as 5.
- Select start option
- Check the confusion matrix obtained for Random Forest



Weka Explorer

Preprocess **Classify** Cluster Associate Select attributes Visualize

Classifier: Choose **RandomForest** -P 100 -I 100 -num-slots 1 -K 0 -M 1.0 -V 0.001 -S 1

Test options

☐ Use training set

☐ Supplied test set

☒ Cross-validation Folds

☐ Percentage split %

(Nom) Thyrod

Result list (right-click for options)

- 23:09:30 - bayes.NaiveBayes
- 23:09:36 - bayes.NaiveBayes
- 23:11:12 - trees.RandomForest**

Classifier output

=== Summary ===

Correctly Classified Instances	6	75	%
Incorrectly Classified Instances	2	25	%
Kappa statistic	0		
Mean absolute error	0.4004		
Root mean squared error	0.4546		
Relative absolute error	93.7526 %		
Root relative squared error	97.0922 %		
Total Number of Instances	8		

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.000	0.000	?	0.000	?	?	0.167	0.226	yes
	1.000	1.000	0.750	1.000	0.857	?	0.167	0.705	no
Weighted Avg.	0.750	0.750	?	0.750	?	?	0.167	0.585	

=== Confusion Matrix ===

```

a b  <-- classified as
0 2 | a = yes
0 6 | b = no

```

Neural Networks: -

- First go to classify section and select choose option over there.
- You will get various options of classifiers over there, under function section select Multilayer perceptron.
- Set k-fold as 5.
- After selecting Multilayer perceptron just click again on it a dialog box will appear
- Set GUI as true and then click on True.
- Select start option
- A box will appear which will show the pictorial representation of neural network architecture
- Just below that image there will two option start and accept option, First click on start and then accept.
- Continue the previous step for five times as the k-fold was set to 5.
- After 5th iteration the confusion matrix will appear in the screen
- Check the confusion matrix obtained for Neural Networks



DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

Discover. Learn. Empower.

Classifier output

=== Summary ===

Correctly Classified Instances	5	62.5	%
Incorrectly Classified Instances	3	37.5	%
Kappa statistic	0.3333		
Mean absolute error	0.4923		
Root mean squared error	0.4957		
Relative absolute error	115.2724	%	
Root relative squared error	105.889	%	
Total Number of Instances	8		

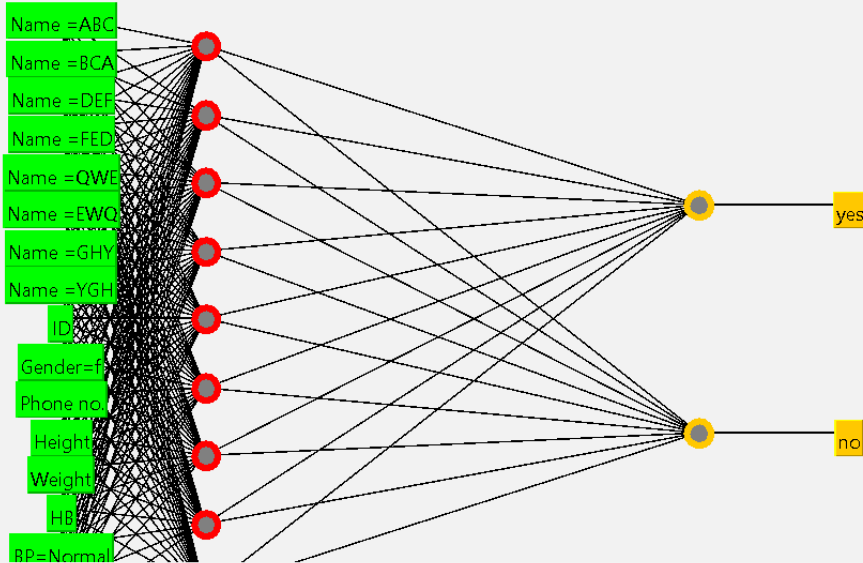
=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	1.000	0.500	0.400	1.000	0.571	0.447	0.583	0.367	yes
	0.500	0.000	1.000	0.500	0.667	0.447	0.583	0.877	no
Weighted Avg.	0.625	0.125	0.850	0.625	0.643	0.447	0.583	0.750	

=== Confusion Matrix ===

```
a b  <-- classified as
2 0 | a = yes
3 3 | b = no
```

Neural Network



Controls

Start Epoch 500
Num Of Epochs 500
Accept Error per Epoch = 0.0005794

Learning Rate = 0.3

Momentum = 0.2



DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

Discover. Learn. Empower.

Learning outcomes (What I have learnt):

1. Learnt how to import data in weka tool
2. Learnt how to perform test and classification using 4 classifiers decision tree, random forest, neural networks, naïve Bayes
3. According to report the F- measures of random forest, naïve Bayes and neural networks is coming as 72.5% whereas for decision tree it is 60% for the same data



DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

Discover. Learn. Empower.

Experiment - 1.3

Student Name: Yuvraj Singh Tanwar

UID: 23MAI10011

Branch: CSE

Section/Group: AIML

Semester: 1

Date of Performance: 23Aug2023

Subject Name: ARTIFICIAL INTELLIGENCE LAB

Code: 23CSH-621.

Aim

To apply the following pre-processing techniques to the training data set and report/analyse the changes in the dataset: -

- a. Normalisation
- b. Class Balance
- c. Add
- d. Remove
- e. Discretization
- f. Sampling

Tasks To Be Done

- Load the dataset.
- Select a preprocessing filter and apply it on the dataset.
- Analyze the changes.

Normalization

- Open WEKA 3.8.4 and click on “*Explorer*” option in GUI.
- Click on the Preprocessing tab and load the dataset by clicking on “*Open File*” option and choosing your desired file. We are using the *breast-cancer.csv* file for implementing normalization.
- Once the dataset is loaded, we carefully choose one of our attributes as the **class attribute** in the bottom-right section. This is because the WEKA visualises the data on the basis of the chosen class attribute. For our dataset we have selected “*diagnosis*” as the class attribute.

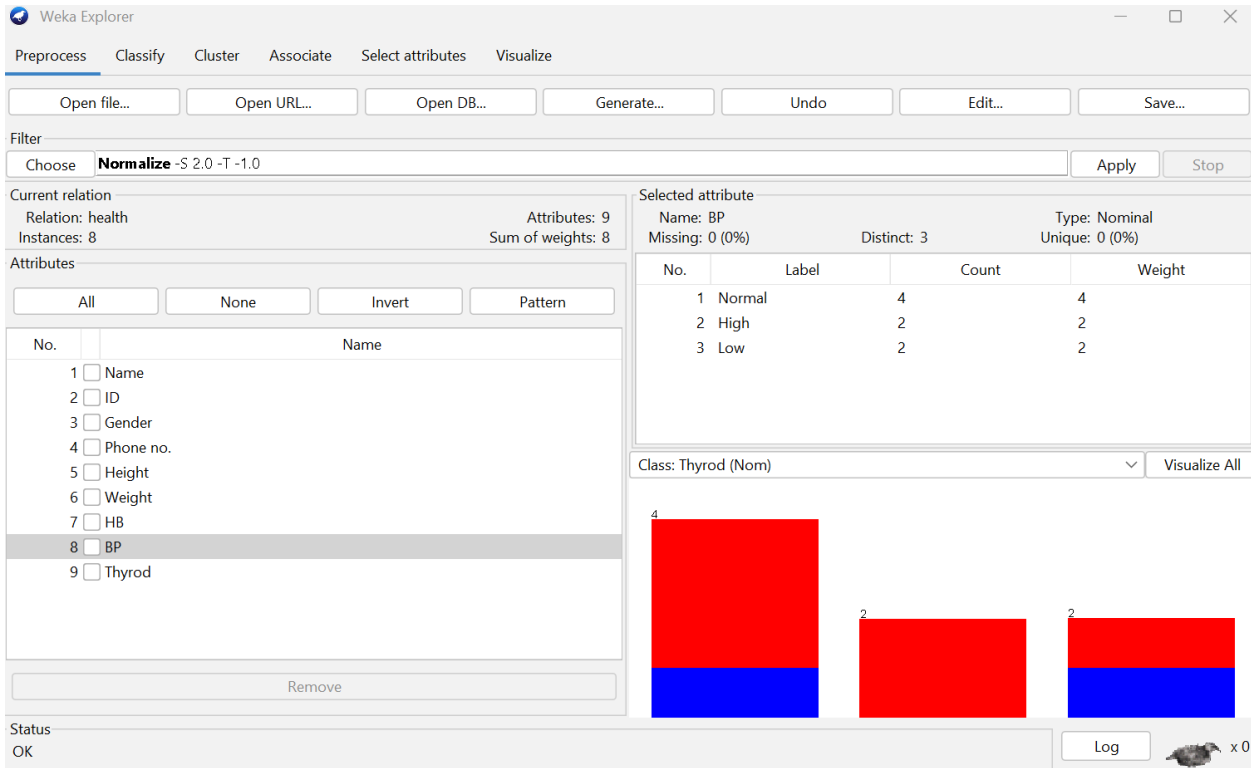


Figure a: Loading the Dataset

- Go to the Filter section and select “*Normalize*” filter from the list.
- Set scale as 2 and translation as -1.
- Click “*Apply*” option.

Selected attribute		
Name: Height		Type: Numeric
Missing: 0 (0%)	Distinct: 7	Unique: 6 (75%)
Statistic	Value	
Minimum	4.8	
Maximum	5.9	
Mean	5.325	
StdDev	0.396	

Figure b: Attribute Before Normalisation

Selected attribute		
Name: Height		Type: Numeric
Missing: 0 (0%)	Distinct: 7	Unique: 6 (75%)
Statistic	Value	
Minimum	-1	
Maximum	1	
Mean	-0.045	
StdDev	0.719	

Figure c: Attribute After Normalisation

Class Balance

- Open WEKA and load the dataset.
- Go to the Filter section and select “*ClassBalancer*” from the list.
- Click on “*Apply*” option.

Selected attribute			
Name: Gender		Type: Nominal	
Missing: 0 (0%)		Unique: 0 (0%)	
		Distinct: 2	
No.	Label	Count	Weight
1	m	4	4
2	f	4	4

Figure d: Attribute Before Class Balance

Selected attribute			
Name: Gender		Type: Nominal	
Missing: 0 (0%)		Unique: 0 (0%)	
		Distinct: 2	
No.	Label	Count	Weight
1	m	4	4
2	f	4	4

Figure e: Attribute After Class Balance

Add

- Open WEKA and load the dataset.
- Go to the Filter section and select “Add” from the list.
- Set **attributeName** to *My_Attribute*.
- Click on “Apply” option.
- A new attribute is added to the dataset.



DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

Discover. Learn. Empower.

Choose **Add -N new -C last -W 1.0** Apply

Current relation
Relation: health-weka.filters.supervised.instance.ClassBalancer-n...
Instances: 8
Attributes: 10
Sum of weights: 8

Attributes

No.		Name
1	<input type="checkbox"/>	Name
2	<input type="checkbox"/>	ID
3	<input type="checkbox"/>	Gender
4	<input type="checkbox"/>	Phone no.
5	<input type="checkbox"/>	Height
6	<input type="checkbox"/>	Weight
7	<input type="checkbox"/>	HB
8	<input type="checkbox"/>	BP
9	<input type="checkbox"/>	Thyrod
10	<input type="checkbox"/>	new

Selected attribute
Name: new
Missing: 8 (100%)
Distinct: 0
Type: Numeric
Unique: 0 (0%)

Statistic	Value
Minimum	NaN
Maximum	NaN
Mean (weighted)	NaN
StdDev (weighted)	NaN

Class: new (Num) Visu

Remove

1. Open WEKA and load the dataset.
2. Go to the Filter section and select “*Remove*” from the list.
3. Set **attributeIndices** to the desired index value (32, in this case).
4. Click on “*Apply*” option.
5. The attribute is removed from the dataset.

Choose **Add -N new -C last -W 1.0** Apply

Current relation
Relation: health-weka.filters.supervised.instance.ClassBalancer-n...
Instances: 8
Attributes: 10
Sum of weights: 8

Attributes

No.		Name
1	<input type="checkbox"/>	Name
2	<input type="checkbox"/>	ID
3	<input type="checkbox"/>	Gender
4	<input type="checkbox"/>	Phone no.
5	<input type="checkbox"/>	Height
6	<input type="checkbox"/>	Weight
7	<input type="checkbox"/>	HB
8	<input type="checkbox"/>	BP
9	<input type="checkbox"/>	Thyrod
10	<input type="checkbox"/>	new

Selected attribute
Name: new
Missing: 8 (100%)
Distinct: 0
Type: Numeric
Unique: 0 (0%)

Statistic	Value
Minimum	NaN
Maximum	NaN
Mean (weighted)	NaN
StdDev (weighted)	NaN

Class: new (Num) Visu

Figure f: Before Removing Attribute

Attributes

All None Invert Pattern

No.		Name
1	<input type="checkbox"/>	Name
2	<input type="checkbox"/>	ID
3	<input type="checkbox"/>	Gender
4	<input type="checkbox"/>	Phone no.
5	<input type="checkbox"/>	Height
6	<input type="checkbox"/>	Weight
7	<input type="checkbox"/>	HB
8	<input type="checkbox"/>	BP
9	<input type="checkbox"/>	Thyrod

Figure h: After Removing Attribute

Discretization

1. Open WEKA and load the dataset.
2. Go to the Filter section and select “Discretize” filter from the list.
3. Click on “Apply” option.
4. Data values are converted into discrete values.

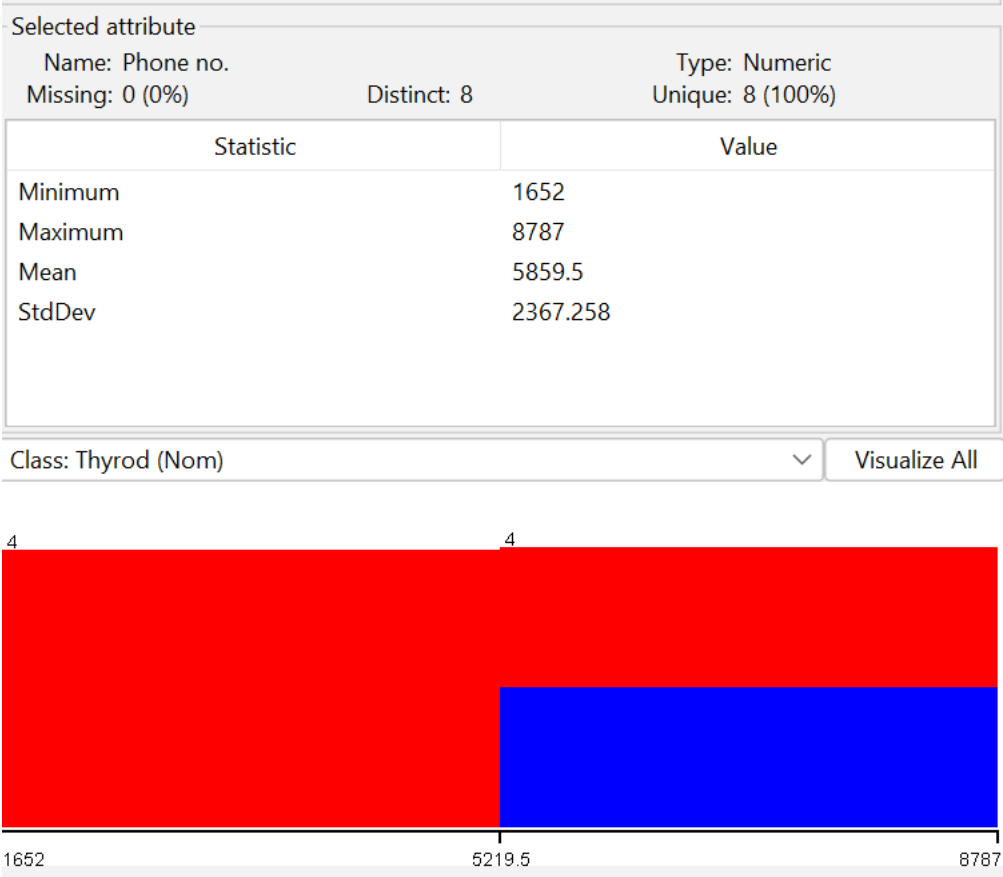


Figure g: Before Discretization



Figure j: after Discretization

Sampling

1. Open WEKA and load the dataset.
2. Go to the Filter section and select the **supervised** version “*Resample*” filter from the list. If the data set does not contain a nominal class attribute, then use the **unsupervised** version.
3. Set **biasToUniformClass** as 1.
4. Click on “*Apply*” option.



Selected attribute			
Name: BP		Type: Nominal	
Missing: 0 (0%)		Distinct: 3	
		Unique: 0 (0%)	
No.	Label	Count	Weight
1	Normal	4	4
2	High	2	2
3	Low	2	2

Class: Thyrod (Nom) Visualize All

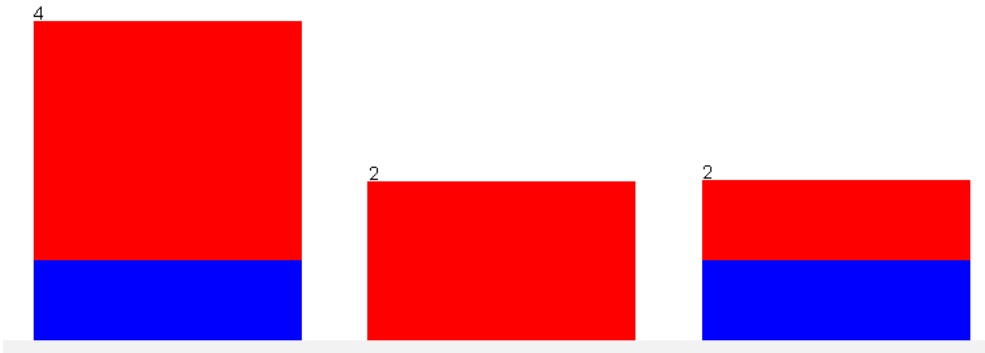


Figure h: Attribute Before Sampling



Figure i: Attribute After Sampling

Learning Outcomes

1. We know how to load dataset.
2. We know how to apply normalization, class balance, add, remove, discretization and sampling to our dataset.



DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

Discover. Learn. Empower.

Experiment - 1.4

Student Name: Yuvraj Singh Tanwar

UID: 23MAI10011

Branch: CSE

Section/Group: AIML

Semester: 1

Date of Performance: 23Aug2023

Subject Name: ARTIFICIAL INTELLIGENCE LAB

Code: 23CSH-621.

Aim: Explore high dimensionality issues in the machine learning. Apply 3 different feature selection techniques to the high dimensional cancer dataset from the UCI repository.

1. Task to be done:

- To apply feature selection in the dataset using filtering method
- To apply feature selection in the dataset using Wrapper method
- To apply feature selection in the dataset using PCA (principal component analysis) method
- To apply feature selection in the dataset using CorrelationAttributeEval

2. Procedure:

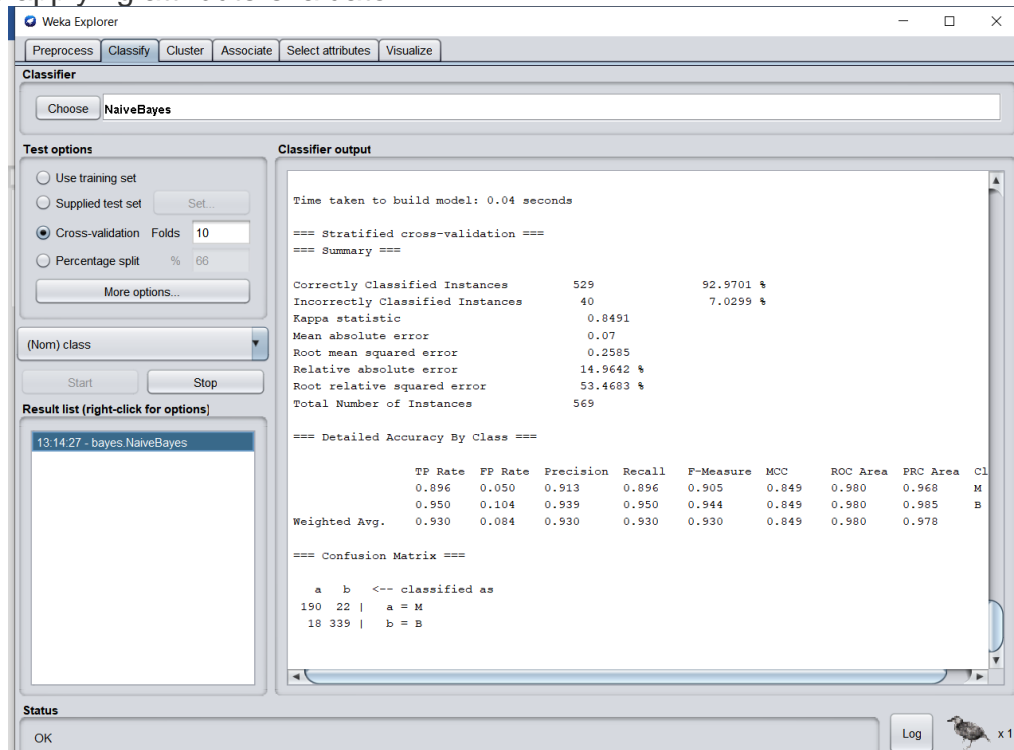
First open the data file (csv format or .arff format) that you have created in the previous experiment and import that in the weka tool using open file option .

a) Wrapper method :-

We can apply the Wrapper method as follows:-

- In the select attribute section,click on choose attribute evaluator> Classifier subset evaluator
- click on Classifier subset evaluator>> a dialog box will appear >> chose naïve bayes
- select the attribute nom(class)
- click on apply option.

Before applying attribute evaluator:-



The screenshot shows the Weka Explorer interface with the NaiveBayes classifier selected. The 'Test options' section shows 'Cross-validation' with 'Folds' set to 10. The 'Classifier output' section displays the following results:

```
Time taken to build model: 0.04 seconds
=== Stratified cross-validation ===
=== Summary ===
Correctly Classified Instances      529      92.9701 %
Incorrectly Classified Instances    40       7.0299 %
Kappa statistic                    0.8491
Mean absolute error                 0.07
Root mean squared error             0.2585
Relative absolute error             14.9642 %
Root relative squared error         53.4683 %
Total Number of Instances          569

=== Detailed Accuracy By Class ===
          TP Rate  FP Rate  Precision  Recall  F-Measure  MCC  ROC Area  PRC Area  Cl
          0.896   0.050   0.913   0.896   0.905   0.849   0.980   0.968   M
          0.950   0.104   0.939   0.950   0.944   0.849   0.980   0.985   B
Weighted Avg.   0.930   0.084   0.930   0.930   0.930   0.849   0.980   0.978

=== Confusion Matrix ===
  a  b  <-- classified as
190 22 | a = M
 18 339 | b = B
```

The 'Result list' on the left shows '13.14.27 - bayes.NaiveBayes'.



DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

Discover. Learn. Empower.

After applying attribute evaluator:-

The screenshot shows the 'Attribute Evaluator' window in Weka Explorer. The 'Choose' button is selected, and the 'ClassifierSubsetEval' evaluator is chosen. The 'Search Method' is set to 'BestFirst -D 1 -N 5'. The 'Attribute Selection Mode' is set to 'Use full training set'. The 'Search Method' output shows the search results, including the 'Selected attributes: 13,18,20,23,24,26 : 6'. The 'Result list' shows the selected attributes.

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)

13.14.41 - BestFirst + ClassifierSubsetEval

Attribute selection output

Search Method:
Best first.
Start set: no attributes
Search direction: forward
Stale search after 5 node expansions
Total number of subsets evaluated: 336
Merit of best subset found: 0.974

Attribute Subset Evaluator (supervised, class (nominal): 1 class):
Classifier Subset Evaluator
Learning scheme: weka.classifiers.bayes.NaiveBayes
Scheme options:
Hold out/test set: Training data
Subset evaluation: classification error

Selected attributes: 13,18,20,23,24,26 : 6
texture_se
concavity_se
symmetry_se
texture_worst
perimeter_worst
smoothness_worst

Status

OK Log x0

The screenshot shows the main Weka Explorer window. The 'Filter' tab is selected, and the 'Current relation' is 'Copy of data-weka.filters.unsupervised.attribute.Re...'. The 'Attributes' list shows the selected attributes: class, texture_se, concavity_se, texture_worst, perimeter_worst, and smoothness_worst. The 'Selected attribute' table shows the distribution of the 'smoothness_worst' attribute.

Weka Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Open file... Open URL... Open DB... Generate... Undo Edit... Save...

Filter

Choose: None Apply Stop

Current relation

Relation: Copy of data-weka.filters.unsupervised.attribute.Re... Attributes: 6
Instances: 569 Sum of weights: 569

Attributes

All None Invert Pattern

No.	Name
1	class
2	texture_se
3	concavity_se
4	texture_worst
5	perimeter_worst
6	smoothness_worst

Remove

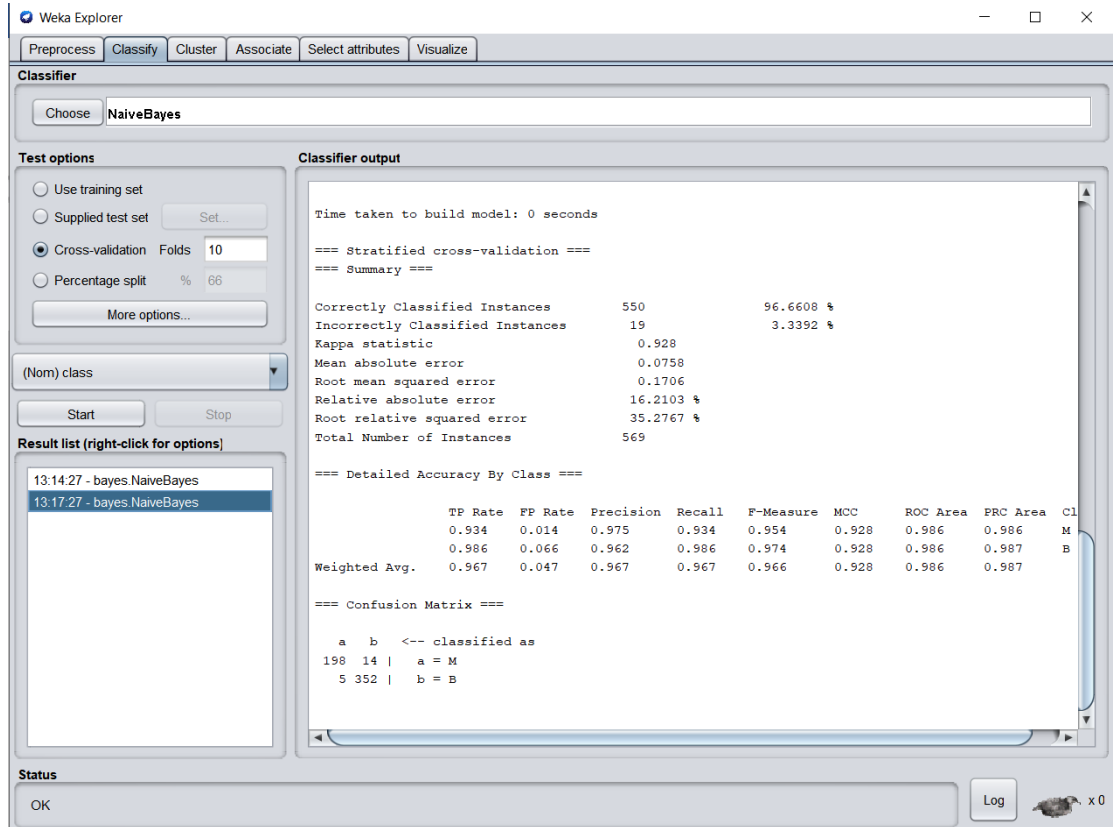
Selected attribute

No.	Label	Count	Weight
1	M	212	212.0
2	B	357	357.0

Class: smoothness_worst (Num) Visualize All

Status

OK Log x0



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)

- 13:14:27 - bayes.NaiveBayes
- 13:17:27 - bayes.NaiveBayes

Classifier output

```
Time taken to build model: 0 seconds

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

Correctly Classified Instances      550           96.6608 %
Incorrectly Classified Instances    19            3.3392 %
Kappa statistic                    0.928
Mean absolute error                 0.0758
Root mean squared error             0.1706
Relative absolute error             16.2103 %
Root relative squared error         35.2767 %
Total Number of Instances          569

=== Detailed Accuracy By Class ===

      TP Rate  FP Rate  Precision  Recall  F-Measure  MCC   ROC Area  PRC Area  Cl
      -----  -
      0.934    0.014    0.975    0.934    0.954    0.928    0.986    0.986    M
      0.986    0.066    0.962    0.986    0.974    0.928    0.986    0.987    B
Weighted Avg.  0.967    0.047    0.967    0.967    0.966    0.928    0.986    0.987

=== Confusion Matrix ===

  a  b  <-- classified as
198 14 | a = M
  5 352 | b = B
```

Status

OK Log x 0

b) Filtering method:-

We can apply the filtering method as follows:-

- In the select attribute section, click on choose attribute evaluator> InfoGainAttributeEval
- click on search method>>choose ranker>>click on ranker>> a dialog box will appear >> keep the value as -1 on num to select
- select the attribute nom(class)
- click on apply option.



DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

Discover. Learn. Empower.

Weka Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Attribute Evaluator

Choose InfoGainAttributeEval

Search Method

Choose Ranker -T-1.7878931348623157E308 -N-1

Attribute Selection Mode

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

(Nom) class

Start Stop

Result list (right-click for options)

- 13:14:41 - BestFirst + ClassifierSubsetEv
- 14:40:49 - Ranker + InfoGainAttributeEval
- 14:41:21 - Ranker + InfoGainAttributeEval

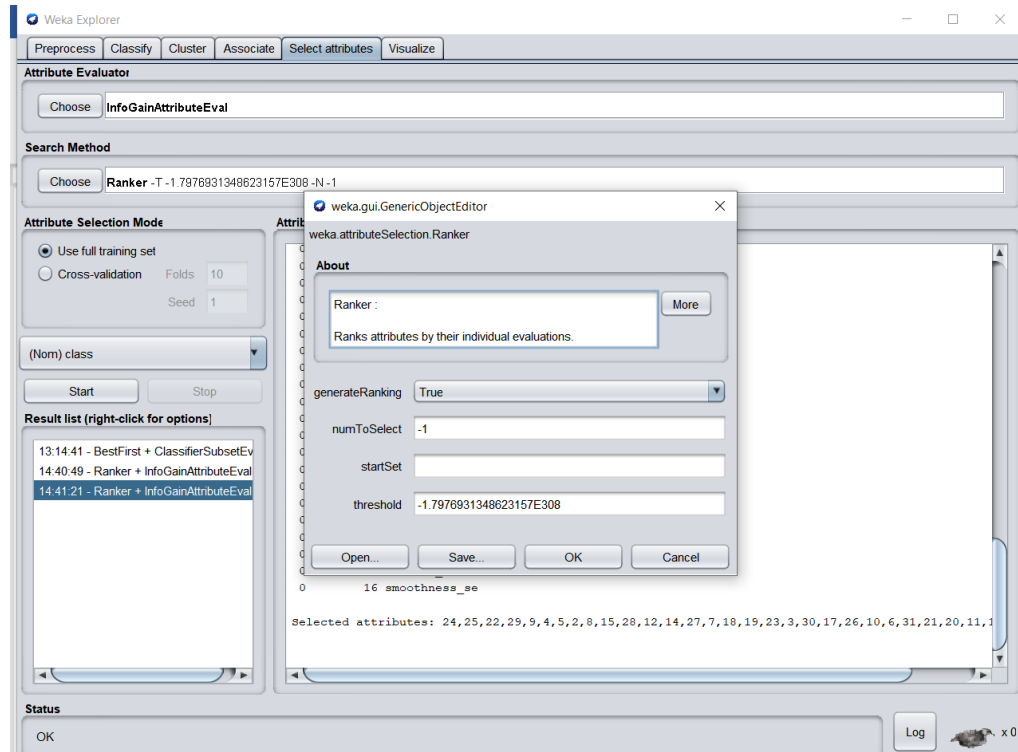
Attribute selection output

0.517	15	area_se
0.4735	28	concavity_worst
0.3679	12	radius_se
0.3663	14	perimeter_se
0.3204	27	compactness_worst
0.304	7	compactness_mean
0.2225	18	concavity_se
0.197	19	concave_points_se
0.1881	23	texture_worst
0.1593	3	texture_mean
0.1492	30	symmetry_worst
0.1303	17	compactness_se
0.1235	26	smoothness_worst
0.0988	10	symmetry_mean
0.0971	6	smoothness_mean
0.0747	31	fractal_dimension_worst
0.0346	21	fractal_dimension_se
0.0228	20	symmetry_se
0	11	fractal_dimension_mean
0	13	texture_se
0	16	smoothness_se

Selected attributes: 24,25,22,29,9,4,5,2,8,15,28,12,14,27,7,18,19,23,3,30,17,26,10,6,31,21,20,11,1

Status

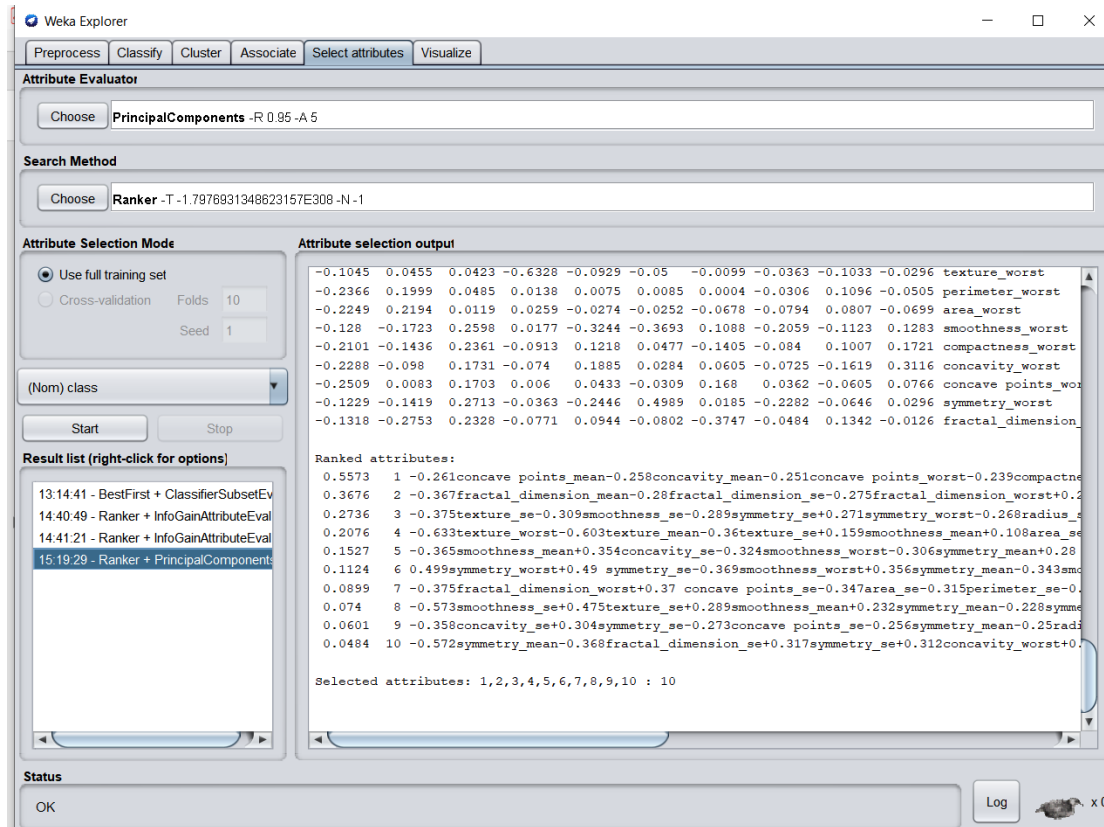
OK Log



c) PCA (principal component analysis):-

We can apply the PCA (principal component analysis) as follows:-

- In the select attribute section, click on choose attribute evaluator> Principal components
- click on search method>>choose ranker>>click on ranker>> a dialog box will appear >> keep the value as -1 on num to select
- select the attribute nom(class)
- click on apply option.



d). Correlation attribute evaluation:-

We can apply the Correlation attribute eval as follows:-

- In the select attribute section, click on choose attribute evaluator> CorrelationAttributeEval
- click on search method>>choose ranker>>click on ranker>> a dialog box will appear >> keep the value as -1 on num to select
- select the attribute nom(class)
- click on apply option.



DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

Discover. Learn. Empower.

Weka Explorer

Preprocess Classify Cluster Associate **Select attributes** Visualize

Attribute Evaluator

Choose **CorrelationAttributeEval**

Search Method

Choose **Ranker -T -1.7976931348623157E308 -N -1**

Attribute Selection Mode

☒ Use full training set
☐ Cross-validation
Folds: 10
Seed: 1

(Nom) class

Start Stop

Result list (right-click for options)

- 15.44.46 - BestFirst + WrapperSubsetEval
- 15.44.53 - GreedyStepwise + WrapperSubsetEval
- 15.45.25 - GreedyStepwise + WrapperSubsetEval
- 15.45.42 - GreedyStepwise + WrapperSubsetEval
- 15.45.49 - GreedyStepwise + WrapperSubsetEval
- 15.47.04 - Ranker + CorrelationAttributeEval**

Attribute selection output

0.65961	28	concavity_worst
0.59653	7	compactness_mean
0.591	27	compactness_worst
0.56713	12	radius_se
0.55614	14	perimeter_se
0.54824	15	area_se
0.4569	23	texture_worst
0.42146	26	smoothness_worst
0.41629	30	symmetry_worst
0.41519	3	texture_mean
0.40804	19	concave points_se
0.35856	6	smoothness_mean
0.3305	10	symmetry_mean
0.32387	31	fractal_dimension_worst
0.293	17	compactness_se
0.25373	18	concavity_se
0.07797	21	fractal_dimension_se
0.06702	16	smoothness_se
0.01284	11	fractal_dimension_mean
0.0083	13	texture_se
0.00652	20	symmetry_se

Selected attributes: 29,24,9,22,4,25,2,5,8,28,7,27,12,14,15,23,26,30,3,19,6,10,31,17,18,21,16,11,1

Status: OK Log x 0

Weka Explorer

Preprocess Classify Cluster Associate **Select attributes** Visualize

Attribute Evaluator

Choose **CorrelationAttributeEval**

Search Method

Choose **Ranker -T -1.7976931348623157E308 -N -1**

Attribute Selection Mode

☒ Use full training set
☐ Cross-validation
Folds: 10
Seed: 1

(Nom) class

Start Stop

Result list (right-click for options)

- 15.44.46 - BestFirst + WrapperSubsetEval
- 15.44.53 - GreedyStepwise + WrapperSubsetEval
- 15.45.25 - GreedyStepwise + WrapperSubsetEval
- 15.45.42 - GreedyStepwise + WrapperSubsetEval
- 15.45.49 - GreedyStepwise + WrapperSubsetEval
- 15.47.04 - Ranker + CorrelationAttributeEval**

weka.gui.GenericObjectEditor

weka.attributeSelection.Ranker

About

Ranker : More

Ranks attributes by their individual evaluations.

generateRanking: **True**

numToSelect: -1

startSet:

threshold: -1.7976931348623157E308

Open... Save... OK Cancel

0.00652 20 symmetry_se

Selected attributes: 29,24,9,22,4,25,2,5,8,28,7,27,12,14,15,23,26,30,3,19,6,10,31,17,18,21,16,11,1

Status: OK Log x 0



DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

Discover. Learn. Empower.

Learning outcomes (What I have learnt):

1. Learned how to use filtering method in WEKA.
2. Learned how to use the wrapper method in WEKA.
3. Learned how to use the principal component analysis in WEKA.
4. Learned how to use the Correlation attribute eval in WEKA.