A

Practical File

On

Data ware house and data mining


Paper Code: ITD08



Bachelor of Engineering

(2016-2020)

Computer Science Engineering

**Submitted By:**

Kavita Maurya

Batch-3

Group-1

2016UCO1579

**DIVISION OF COMPUTER SCIENCE ENGINEERING**
**NETAJI SUBHAS UNIVERSITY OF TECHNOLOGY, DWARKA**

# 1.

<u>**AIM:**</u>

Write a procedure for cross-validation using J48 Algorithm for weather table.

<u>**DESCRIPTION:**</u>

Cross-validation, sometimes called rotation estimation, is a technique for assessing how the results of a statistical analysis will generalize to an independent data set. It is mainly used in settings where the goal is prediction, and one wants to estimate how accurately a predictive model will perform in practice. One round of cross-validation involves partitioning a sample of data into complementary subsets, performing the analysis on one subset (called the training set), and validating the analysis on the other subset (called the validation set or testing set).
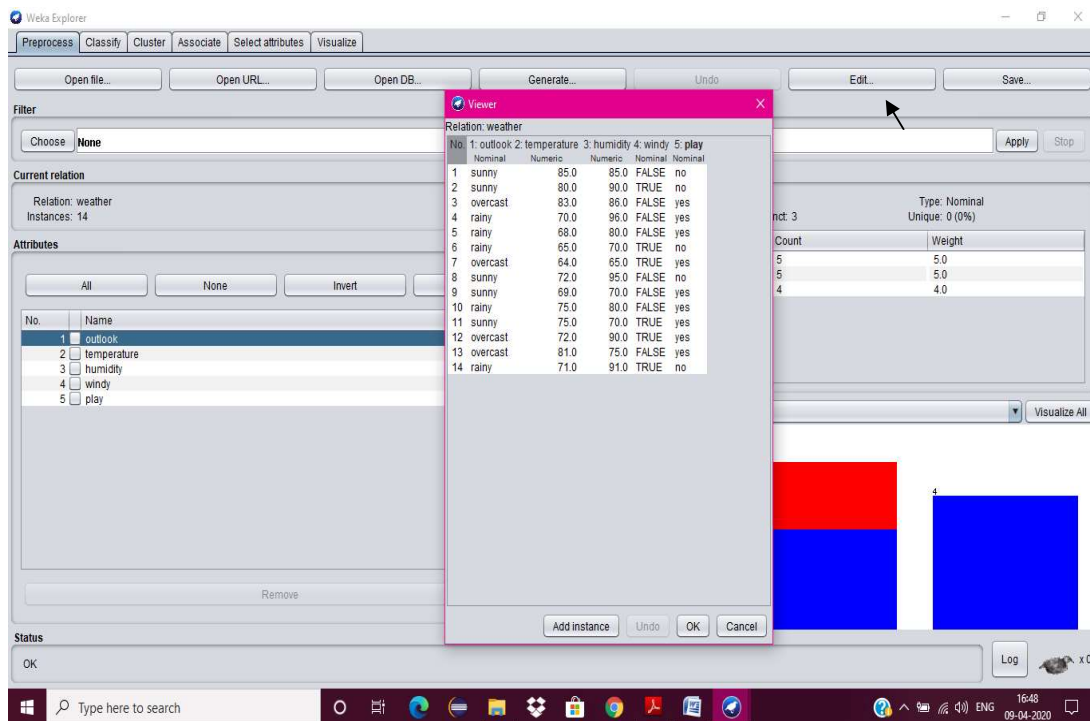
<u>**PROCEDURE:**</u>

1. Create or download **weather.arff** file having weather data.
2. Click on weka and then click on explorer.



3. Explorer shows many options. In that click on 'open file' and select the arff file.
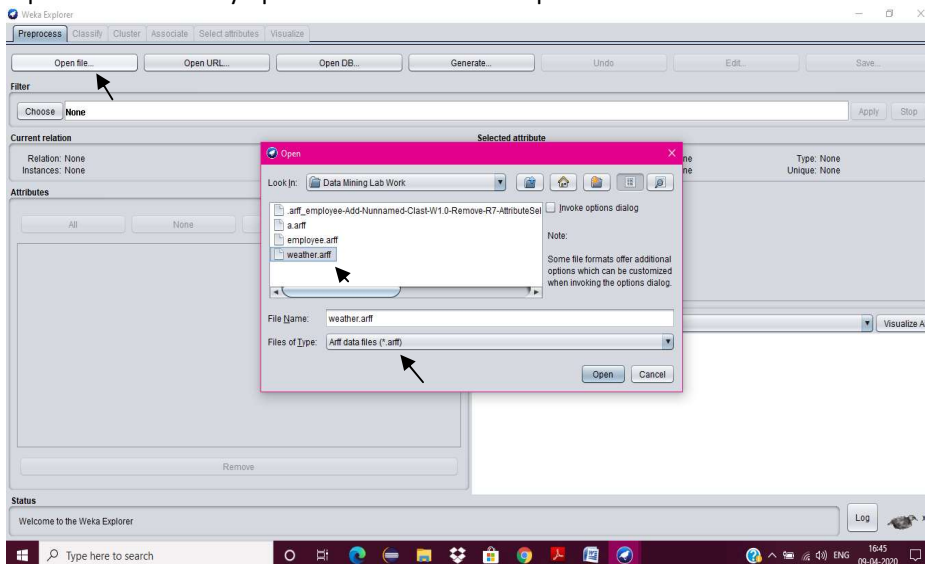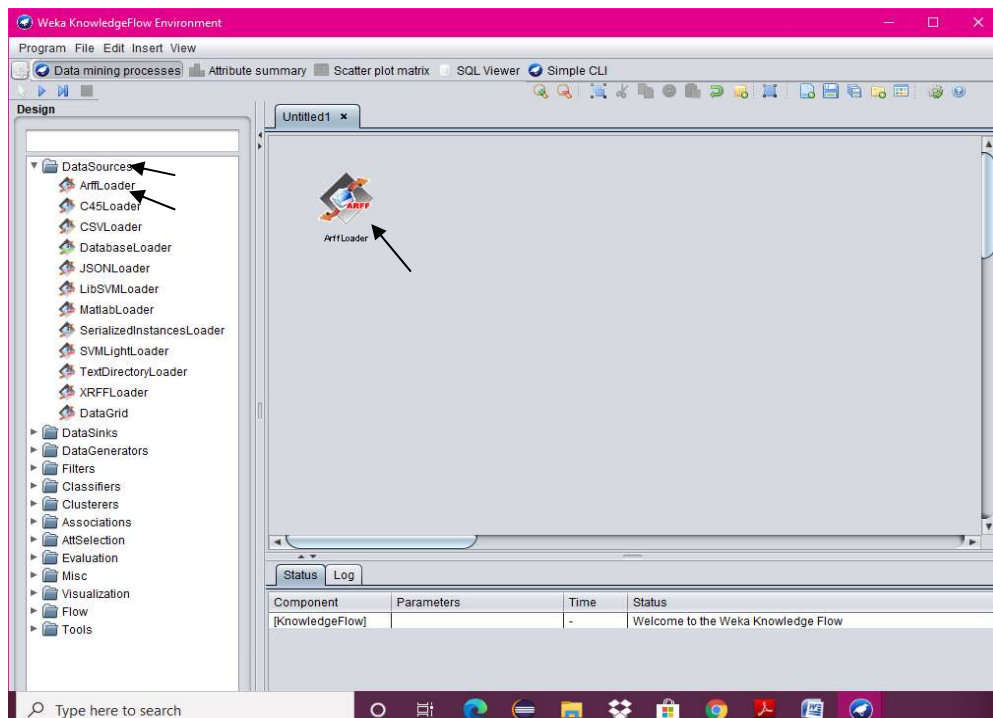
4.  Click on edit button which shows weather table on weka.
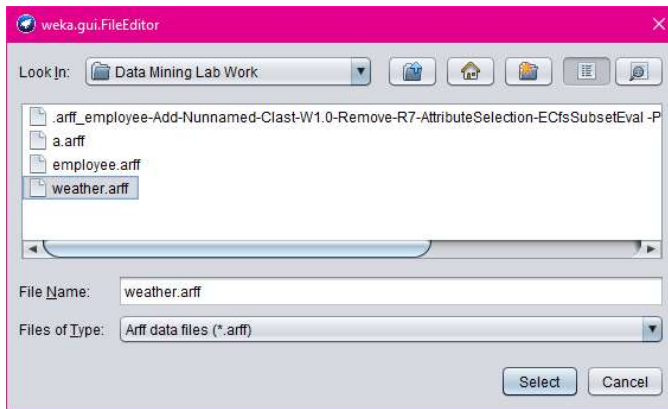


5.  Close Weka Explorer and then select KnowledgeFlow from GUI.

6.  Select Data Source tab & choose Arff Loader.
7.  Place Arff Loader component on the layout area by clicking on that component.



8.  Specify an Arff file to load by right clicking on Arff Loader icon, and then a pop-up menu will appear. In that select Configure & browse to the location of weather.arff

9. Click on the Evaluation tab & choose Class Assigner & place it on the layout.

10. Now connect the Arff Loader to the Class Assigner by right clicking on Arff Loader, and then select Data Set option, now a link will be established.
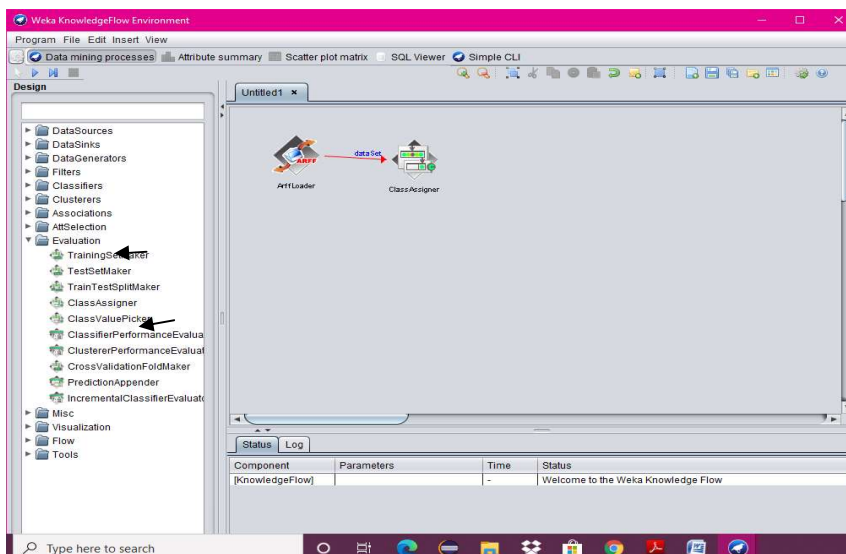


11. Right click on Class Assigner & choose Configure option, and then a new window will appear & specify a class to our data.

12. Select Evaluation tab & select Cross-Validation Fold Maker & place it on the layout.

13. Now connect the Class Assigner to the Cross-Validation Fold Maker.
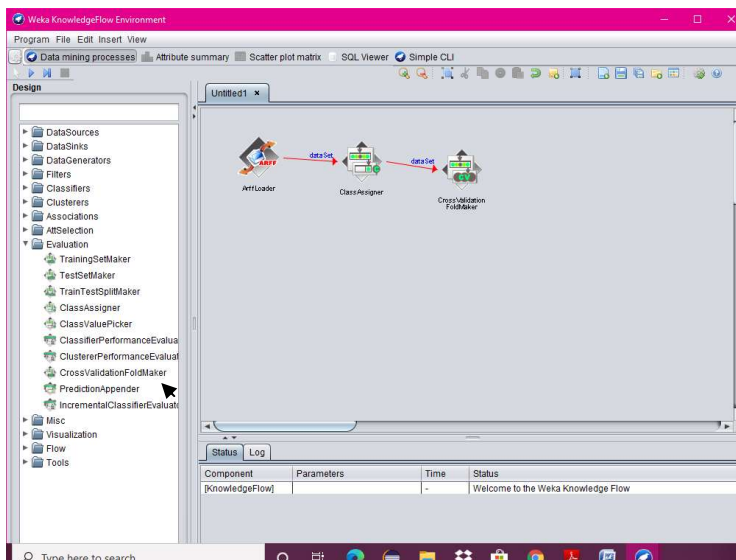
14. Select Classifiers tab & select J48 component & place it on the layout.

15. Now connect Cross-Validation Fold Maker to J48 twice; first choose Training Data Set option and then Test Data Set option.
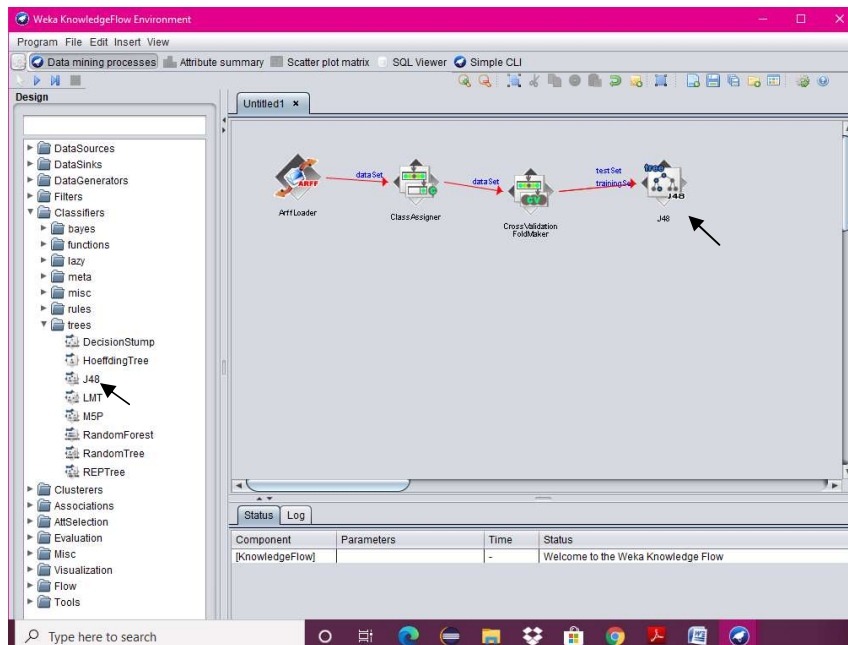


16. Select Evaluation Tab & select Classifier Performance Evaluator component & place it on the layout.

17. Connect J48 to Classifier Performance Evaluator component by right clicking on J48 & selecting Batch Classifier.
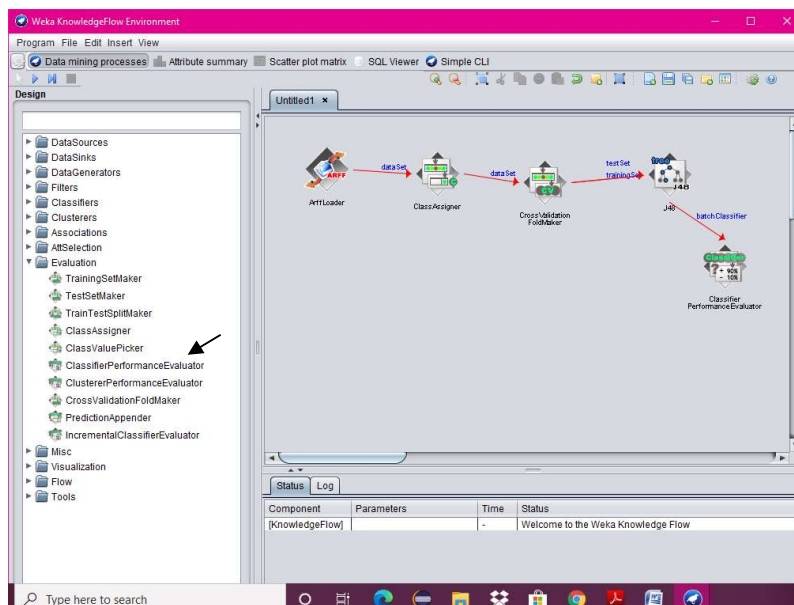


18. Select Visualization tab & select Text Viewer component & place it on the layout.

19. Connect Classifier Performance Evaluator to Text Viewer by right clicking on Classifier Performance Evaluator & by selecting Text option.
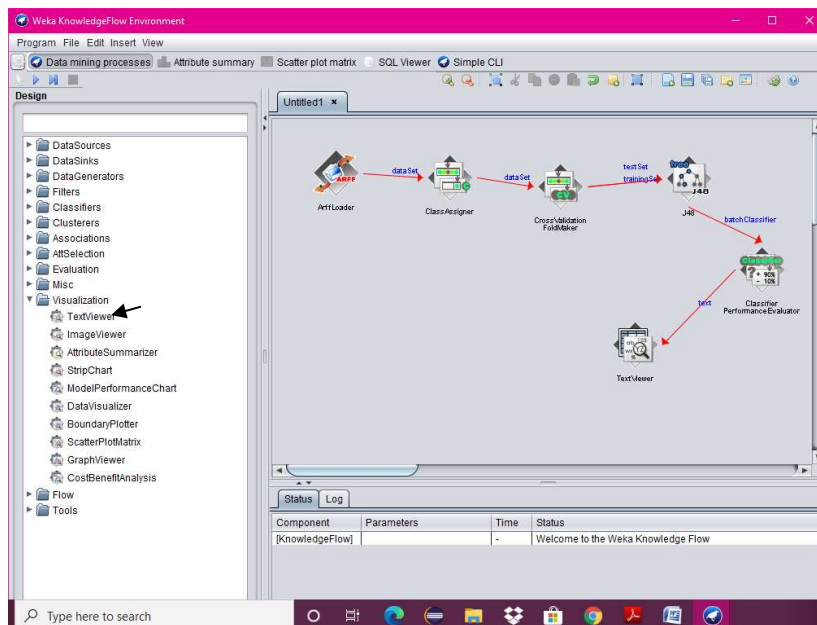
20. Start the flow of execution by selecting Start Loading from Arff Loader.

21. For viewing result, right click on Text Viewer & select the Show Results, and then the result will be displayed on the new window.



=== Evaluation result ===

Scheme: J48
Options: -C 0.25 -M 2
Relation: weather

| | | |
|---|---|---|
| Correctly Classified Instances | 9 | 64.2857 % |
| Incorrectly Classified Instances | 5 | 35.7143 % |
| Kappa statistic | 0.186 | |
| Mean absolute error | 0.2857 | |
| Root mean squared error | 0.4818 | |
| Relative absolute error | 60 % | |
| Root relative squared error | 97.6586 % | |
| Total Number of Instances | 14 | |

=== Detailed Accuracy By Class ===

| | TP Rate | FP Rate | Precision | Recall | F-Measure | MCC | ROC Area | PRC Area | Class |
|---|---|---|---|---|---|---|---|---|---|
| | 0.778 | 0.600 | 0.700 | 0.778 | 0.737 | 0.189 | 0.789 | 0.847 | yes |
| | 0.400 | 0.222 | 0.500 | 0.400 | 0.444 | 0.189 | 0.789 | 0.738 | no |
| Weighted Avg. | 0.643 | 0.465 | 0.629 | 0.643 | 0.632 | 0.189 | 0.789 | 0.808 | |

=== Confusion Matrix ===

```
 a b   <-- classified as
 7 2 | a = yes
 3 2 | b = no
```

# 2.

Write a procedure for Clustering Buying data using Cobweb Algorithm.
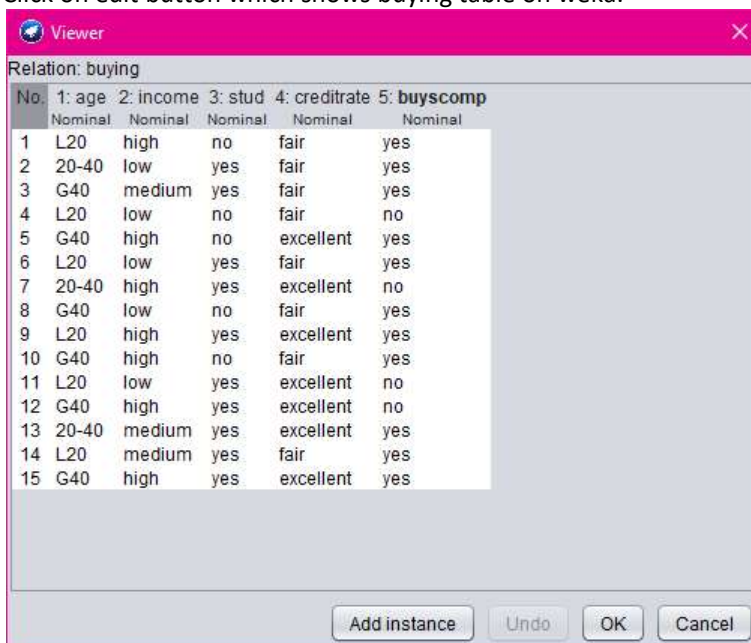
**DESCRIPTION:**

Cluster analysis or clustering is the task of assigning a set of objects into groups (called clusters) so that the objects in the same cluster are more similar (in some sense or another) to each other than to those in other clusters. Clustering is a main task of explorative data mining, and a common technique for statistical data analysis used in many fields, including machine learning, pattern recognition, image analysis, information retrieval, and bioinformatics.

**PROCEDURE:**

1. Create or download **buying.arff** file having buying data.
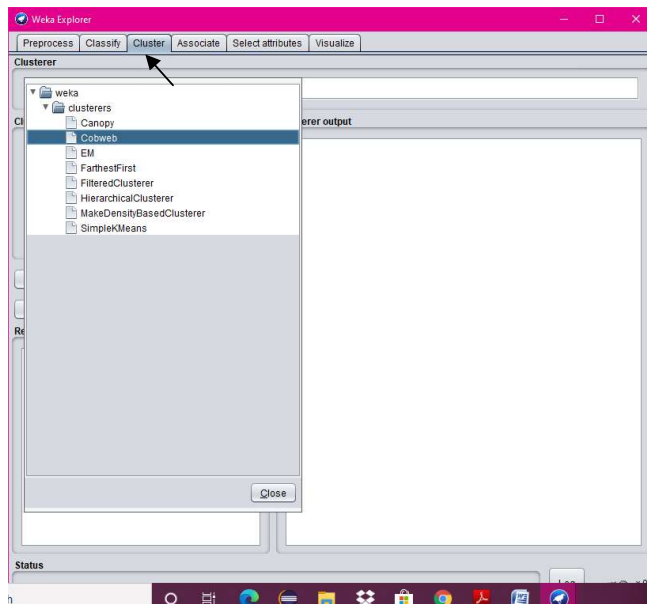2. Click on weka and then click on explorer.



3. Explorer shows many options. In that click on 'open file' and select the arff file.

4. Click on edit button which shows buying table on weka.

5.  Close the file.

6.   Click on Cluster menu. In this there are different algorithms are there.

7.  Click on Choose button and then select cobweb algorithm.



8.  Click on Start button and then output will be displayed on the screen.