Experiment No – 4

Name : Dhiraj Ravindra Bodake
Roll No : 18141216

Title : Learn to install open source data mining such as Weka. Experiments with preparing the data, Loading the data, Building a decision tree, Examining the output.

Theory : Weka is an open source collection of data mining takes which you can utilize in a number of different ways. It comes with Graphical user interface (GUI), but can also be called from your own Java code. You can even write your own batch files for tasks that you need to execute more than once, possibly with slightly different parameters each time.

1. Downloading and installing Weka -
   IF you want to use Weka on your own machine, Follow the instructions on Weka Homepage. Weka installed on all the CS machines. IF you are using Weka on your own machine.

2. Starting up the Weka Explore -
   From CS machines
   Open Command window and type
   Weka

On your own computer. Either double-click on weka-3-8-2-oracle-jvm icon in your weka installation folder or open command window and type -

java -Xmx 500M weka.gui.explorer. Explorer

3. Introduction to the Weka Explore -

The Preprocess panel is the panel opened after starting the WEKA explorer. Before changing to any of the other panels the explorer must have data set to work with. Load a dataset by clicking the open file button in the top left corner of the panel.
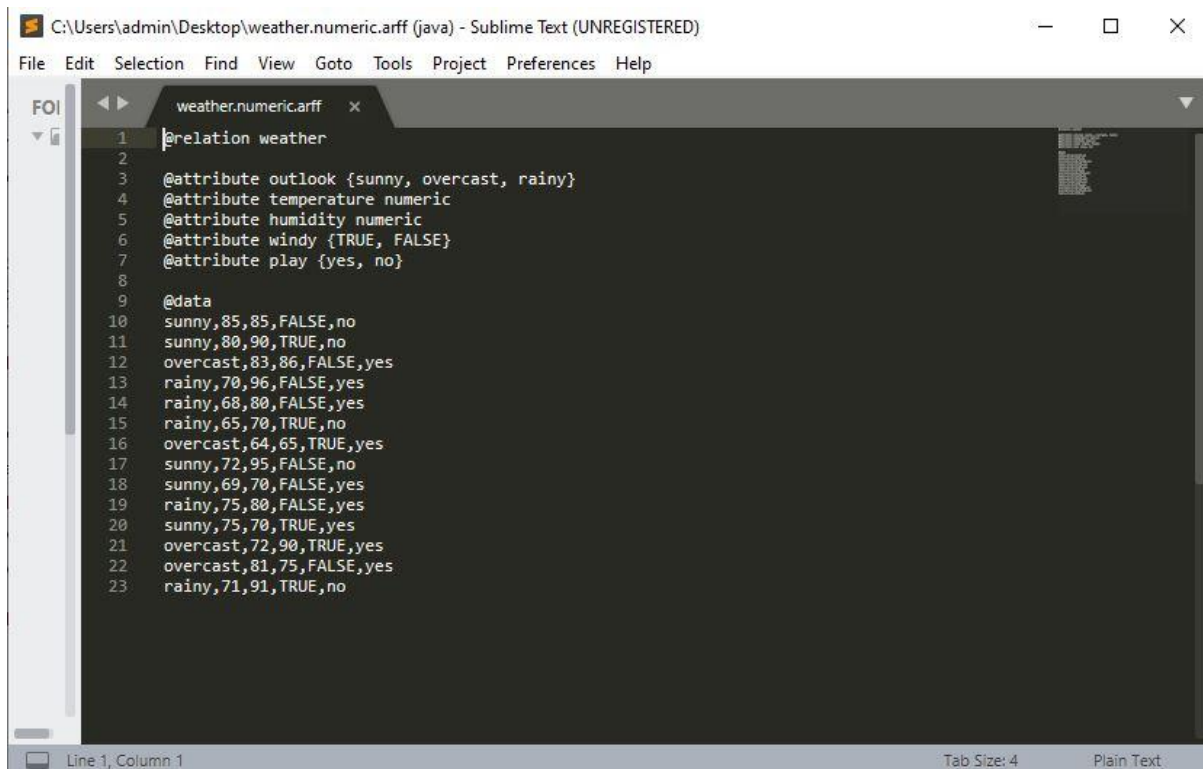
As the result shows, the wether data has 14 instances and 5 attributes called outlock, temperature, humidity, widy and play. Click on the name of an attribute in left subpanel to see information about the selected attribute on right.
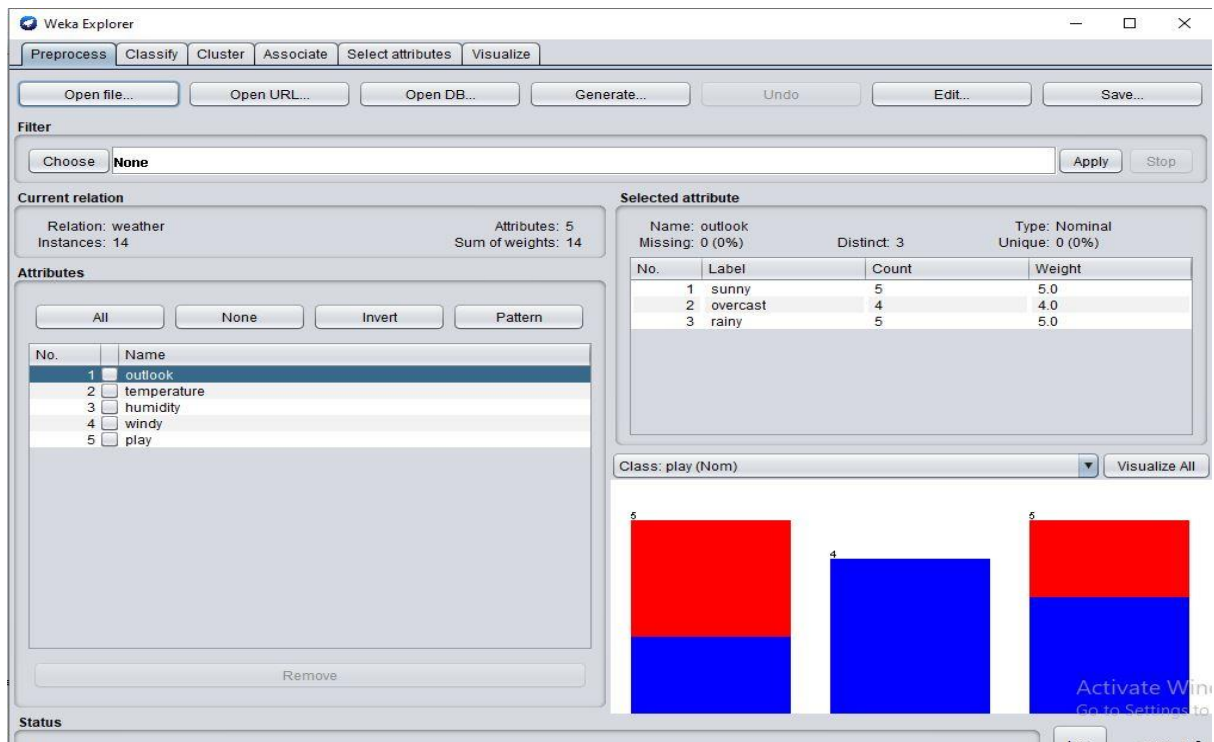
4. Dataset Editor -

It is possible to view and edit an entive dataset from within Weka. To do this, click the edit button from the row of buttons at the top of the preprocess panel. This opens a new window called viewer, which lists all instances of the weather data.

* Explorer has the following components:

1) Preprocess : The first component of explorer provides an option for data preprocessing. Various formats of data like ARPF, CSV, C4.5 binary etc. ARFF stands for attribute-relation File Format, and it was developed for use with the Weka machine learning software.

2) Classify : The next option in Weka explorer is the classifier, which is a model for predicting nominal or numeric quantities and includes various machine learning tehnique like decision trees and lists, instance-based classifiers, support vector machines, multi-laye perceptrons, logistics regression.

3) Cluster : The cluster panel is similar to the classify panel. Many techniques like k-Means, EM, Cobweb, X-mean. The output in this tab contains the confusion matrix, which shows how many errors the would be if the clusters were used instedd of the true class

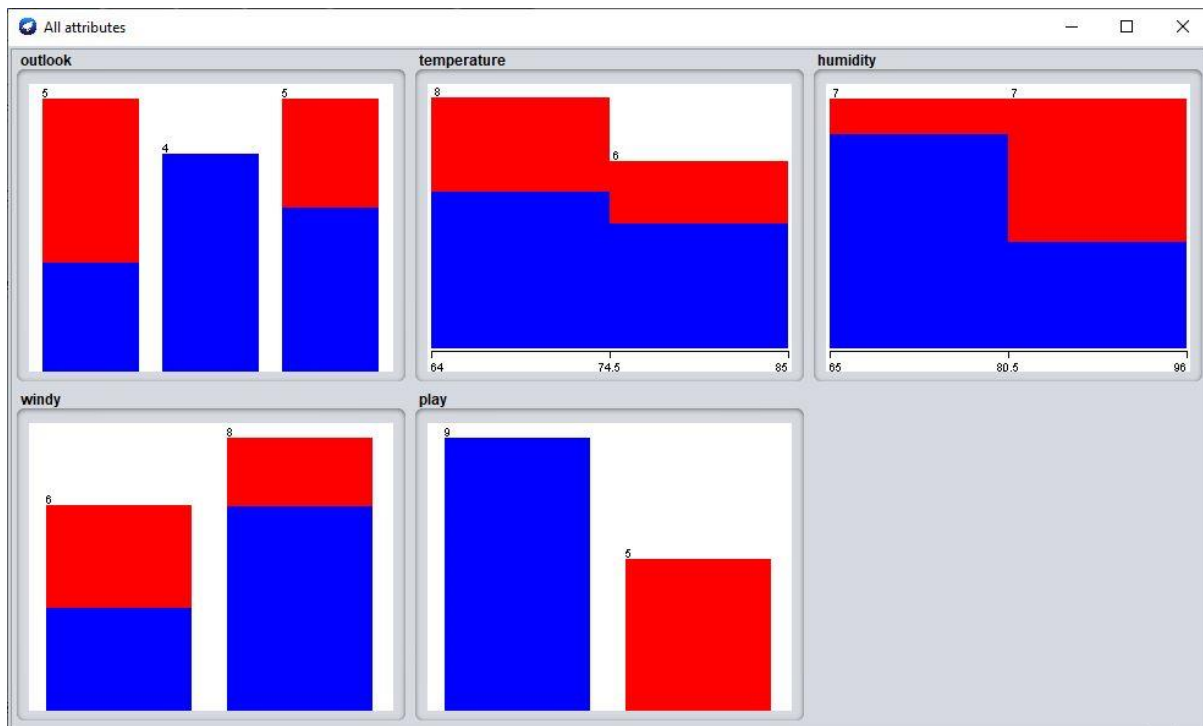4) Visualize : This tab can be used to visualize the result. It displays a scatte plot for every attribute.

Weka GUI Chooser

Program Visualization Tools Help

WEKA
The University
of Waikato

Applications

Explorer

Experimenter

KnowledgeFlow

Workbench

Simple CLI

Waikato Environment for Knowledge Analysis
Version 3.8.5
(c) 1999 - 2020
The University of Waikato
Hamilton, New Zealand

C:\Users\admin\Desktop\weather.numeric.arff (java) - Sublime Text (UNREGISTERED)

File Edit Selection Find View Goto Tools Project Preferences Help

weather.numeric.arff

```
1   @relation weather
2
3   @attribute outlook {sunny, overcast, rainy}
4   @attribute temperature numeric
5   @attribute humidity numeric
6   @attribute windy {TRUE, FALSE}
7   @attribute play {yes, no}
8
9   @data
10  sunny,85,85,FALSE,no
11  sunny,80,90,TRUE,no
12  overcast,83,86,FALSE,yes
13  rainy,70,96,FALSE,yes
14  rainy,68,80,FALSE,yes
15  rainy,65,70,TRUE,no
16  overcast,64,65,TRUE,yes
17  sunny,72,95,FALSE,no
18  sunny,69,70,FALSE,yes
19  rainy,75,80,FALSE,yes
20  sunny,75,70,TRUE,yes
21  overcast,72,90,TRUE,yes
22  overcast,81,75,FALSE,yes
23  rainy,71,91,TRUE,no
```

Line 1, Column 1   Tab Size: 4   Plain Text

All attributes

outlook | temperature | humidity
windy | play

Weka Explorer

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

Classifier

Choose | **J48** -C 0.25 -M 2

Test options

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

More options...

(Nom) play

Start | Stop

Result list (right-click for options)

01:04:39 - trees.J48

Classifier output

```
=== Summary ===

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
```
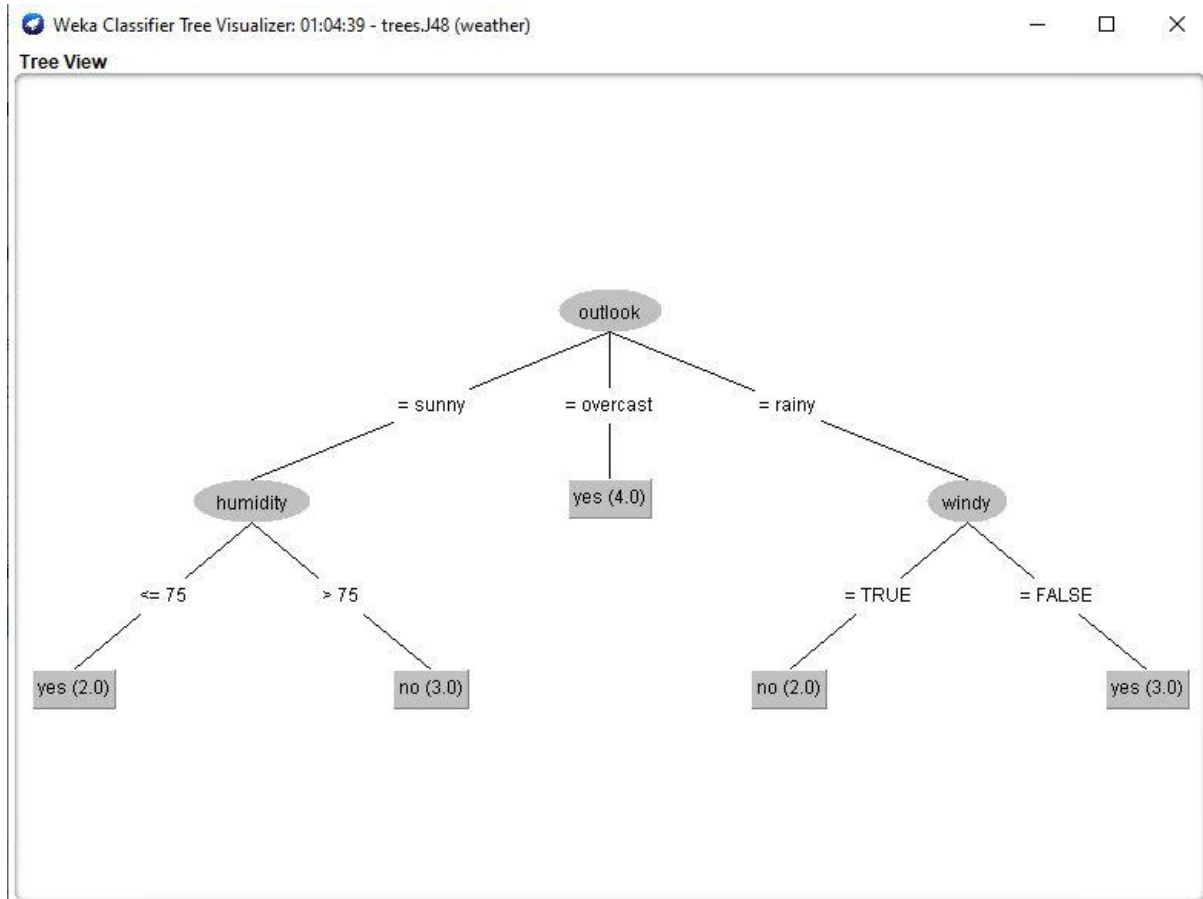
Status

OK | Log

Conclusion - In this experiment I Learnt to install open source data mining software such as Weka and done experiments with Preparing the data, Loading the data, Building a decision tree, Examining the output.ss