Experiment - 2

Aim: Explore Machine learning tool "WEKA"

→ Explore WEKA Data Mining / Machine Learning Toolkit.

→ Downloading and/or installation of WEKA data mining toolkit.

→ understand the features of WEKA toolkit such as Explorer, knowledge Flow interface, Experimenter, command-line interface.

→ Navigate the options available in the WEKA (ex. select attributes panel, preprocess panel, classify panel etc.)

→ Study the arff file format Explore the available data sets in WEKA. Load a data set (ex. Weather dataset etc)

→ Load each dataset and observe the following:

1. List the attribute names and they types

2. Number of records in each dataset

3. Identify the class attribute (if any)

4. plot Histogram.

5. Determine the number of records for each class.

6. Visulize the data in Various dimensions.

Objectives:

Data Ware housing is a technique of gathering and Analyzing data from many sources to get valuable business insights. Typically a data ware house integrates and analyzes business data from many sources Data Ware housing is a vital component of business intelligence.

Preprocesser:

The data that is collected from the field contains many unwanted things that leads to wrong analysis. Thus, the data must be preprocessed to meet the requirements of the type of analysis you are seeking. This is the done in the preprocessing module.

classifiers:

classifiers in WEKA are the models for predicting nominal & numeric quantities. The learning schemes available in WEKA include decision trees and lists, instance - based classifiers, classifiers include bagging, boosting, stacking, erner - connecting output codes and locally weighted learning.

WEKA:

WEKA (Waikota environment for knowledge Analysis) is a popular site of Machine learning software written in Java developed at the university of Waikota. New zealand. Weka is free software available under the GNO General Public License.

Weka - an open source software provides tools for data preprocessing implementation of several Machine learning Algorithms and visualization tools so that you can develop machine learning techniques and apply them to real-world data mining problems.

Explorer : It is an environment for exploring data explorer Consists of Several tools. They are :-

→ **preprocess :**

It is the first step in machine learning is to preprocess the data. It is used to select the data file, process it and make it fit for applying the various machine learning Algorithms.

→ **classify :**

The classify tab provides you Several machine learning algorithms for the classification of your data. Such as linear Regression, logistic Regression.

→ **cluster :**

under the cluster tab there are Several Clustering Algorithm provided - Such as Simple k means, filtered cluster, Hierarchical cluster.

→ **Associate :**

under the Associate tab you would find Apriori filtered Associator and FP Growth.

→ **Select Attributes Tab :**

Select Attributes allows you feature Selections based on Several algorithms Such as classifier Subset eval, principal Component.

→ **Visualize Tab :**

The Visualize option allows you to visualize your processed data for analysis.

Simple CLI :

It provides a Simple Command-line interface and allows direct execution of Weka Commands.

Experimenter :

It is an environment for performing experiments and Conducting statistical tests between learning Schemes.

Knowledge flow:

It is a Java-Beans based interface for setting up and running machine Learning experiments.

Trees J48 classifier:

It is an Algorithm to generate a decision tree that is generated by C4.5. It is also known as Statistical classifier. for decision tree classification, we need a database.

Weather nominal:

In Weka, attributes Can be nominal or numeric. The value of a nomial attribute is represented by a word: Sunny, overcast and rainy for the outlook attribute: yes and no for the play attribute.
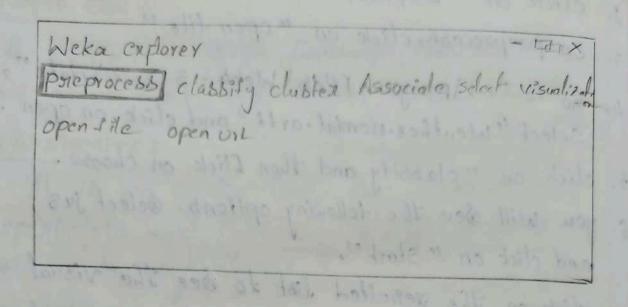
Steps Required :

1. open WEKA you can see 5 tabs on the right side of the application. They are: explorer, experimentor, knowledge flow, Work Bench, Simple CLI

2. click on "Explorer"

3. on preprocess. click on "open file"

4. Go to "c:\program files\Weka-3-8-6\data", Select "Weather.nomial.arff" and click on open.

5. click on "classify and then click on choose.

6. you will see the following options. select j48 and click on "Start",

7. click on the resulted list to see the visual

8. click on the resulted list and click on visualize tree option.

Outcome of the experiment :

Program	visualization	Tools	Help	$-\ \square\ \times$
				Application
				Explorer
	WEKA			experimenter
				knowledge -flow
				Work bench
				Simple CLI

Weka explorer   — □ X

[Preprocess] classify cluster | Associate select visualiz.

open file   open url

---

open

Look in: Weka 3-8-6

changelogs

data ——→ Weather.nominal

doc.

jre

File name: Weather.nominal.arff

[Open] [Cancel]

Preprocess classify cluster Associate select Attribute Visualization

openfile   openURL   openpb   Gen.

File	Select attribute		
Chose None.	NO label	count	Weight
	1 Sunny	5	5
	2 Windy	4	4
	3 Rainy	5	5