Experiment-7

Aim: Demonstrate zeroR technique on iris dataset.

Objectives:

IRis is an open access flower-based dataset is normally available on UCI dataset. The major objective of this research work is to examine the IRIS data using data mining techniques available supported in weka. In this work, four different classifier, viz Bayes, Network classifier, J48, Random forest and oneR has been successfully used to classify attributes the dataset.

ZeroR:

→ ZeroR is the simplest classification which relies on the largest and ignores all predictors.

→ ZeroR classifier simply predicts the majority category.

→ Although there is no predictability power in zeroR it is useful for determining a baseline performance as a benchmark for other classification methods.

Iris Dataset:

The Iris flower dataset is a famous dataset from statistics and is heavily borrowed by researches in machine learning. It contains 150 instances and 4 attributes and a class attribute for the species of iris flower.

WEKA:

WEKA (Waikota environment for knowledge Analysis is a popular suite of machine learning software written in java, developed at the university of waikota, New zealand weka is a free software available under the GNO, General Public license.

Preprocess tab:

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

Loading Data:- The first four buttons at the top of the preprocess section enable you to load data into WEKA.

→ open file:- Brings up a dialog box allowing you to browse for the data file on the local file system.

→ open URL... Asks for a uniform Resource locater address for where the data is stored.

→ open DB... Reads data from a database.

→ Generate.. Enables you to generate artificial data from variety of data generators. Using the open files... button you can read files in a variety of formats - WEKA's ARFF format, CSV format, C4.5 format.

classify:- The classify tab provide you several machine learning algorithms for the classification of your data - such as linear regression, logistic regression.

Test options:-

Before you run the classification algorithm, you need to set test options. Set test options in the Test options box. The test options that available are:

1. Use training set:- Evaluates the classifier on how well it predicts the class at the instances it was trained on.

2. Supplied test set:- Evaluates the classifier on how well it predicts the class of a set of instances loaded from a file. Clicking on the set... button brings up a dialog allowing you to choose the file to test on.

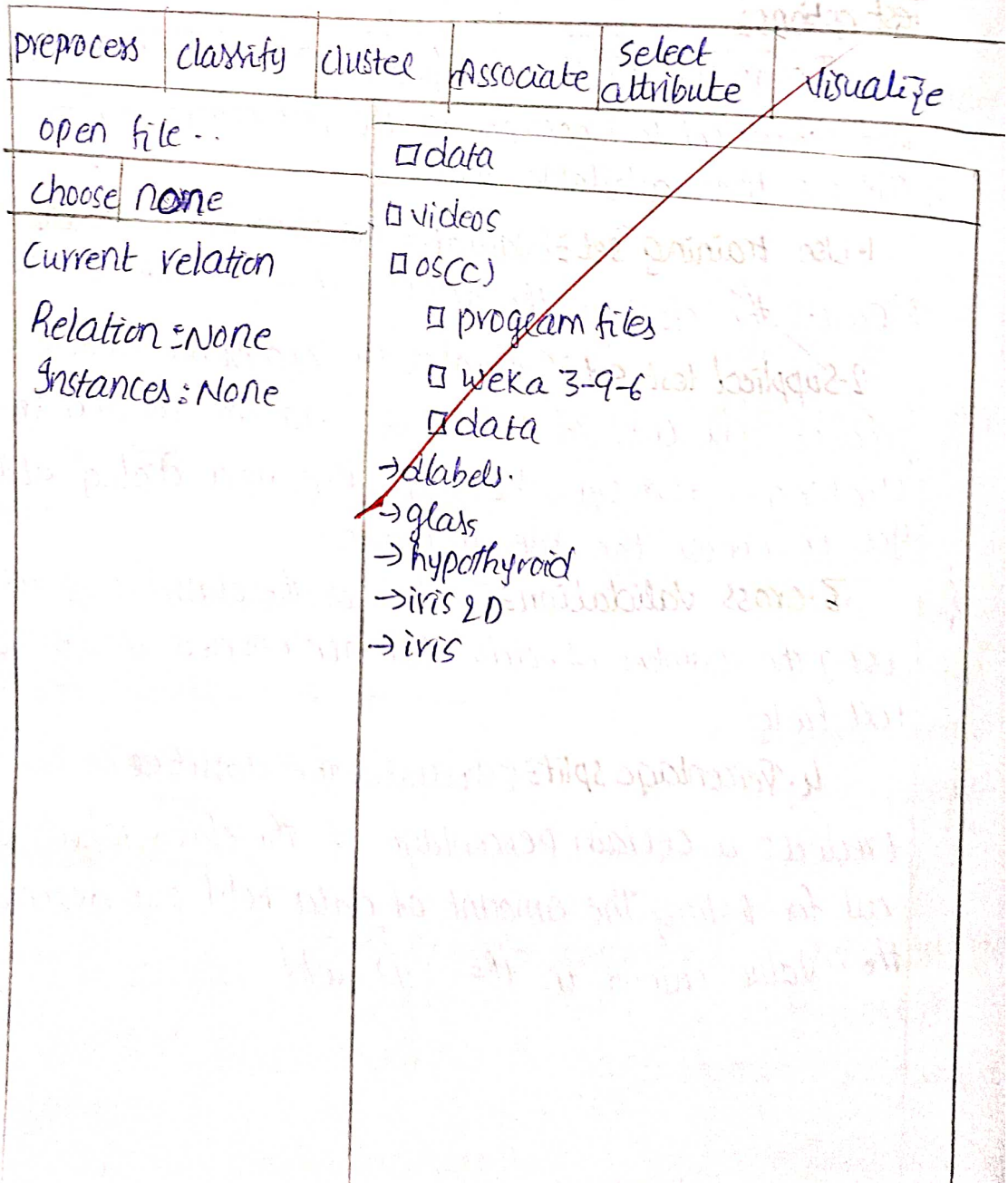
3. Cross validation:- Evaluates the classifier by cross-validation using the number of folds that are entered in the 'Folds' text field.

4. Percentage split:- Evaluates the classifier on how well it predicts a certain percentage of the data, which is held out for testing. The amount of data held out depends on the value entered in the '%' field.

Steps Required:

1. Open WEKA tool.
2. Click on WEKA explorel.
3. Click on the preprocess, and click on open file.
4. Open cdrive, then click on open file, then select WEKA 3.9.6
5. Select data file.
6. Select iris dataset.
7. You can see the Jaciance of the geaph.
8. Click on the classifiel, then zeroR
9. And set cross Jalidation fold as 10.
10. Click on submit.

Output:



preprocess | classify | cluster | Associate | select Attribute | visualize

Classifier

choose | ZeroR

Test options.

- Use training set
- Supplied test set
- Cross-Validation Folds 10
- percentage splits % 60

More options...

Start | Stop

Classifier output:

2/10/22