Experiment No :- 03

19_Sanket Chandrashekhar Harvande

PAGE NO.
DATE / /

Experiment No. 03

Aim :- Classification via decision tree using WEKA Tool.

Theory and concept :-

Classification model predicts catagorical class labels, and prediction model predict continous valued function. For example, we can build a classification model to catagorize bank loan application as either safe or risky or a prediction model to predict the expenditure in dollars of potential customers on computer equipment given their income and occupation.

The core algorithm for building decision trees called ID3 by J.R. Quinlan which employs a top down, greedy search through the spade of possible branches with no backtracking. ID3 uses entropy and information Gain to construct a decision tree. In ZeroR model there is no predictor. In OneR model we try to find the single best predictor, naive Bayesian includes all predictors using Bayes rule & the independence assumptions between predictors but decision tree includes all predictors with the dependence assumptions between predictors.

Entropy :-

A decision tree is built top-down from a root node & involves partioning the data into subsets that contains instances with similar values. ID3 algorithm uses entropy to calculate the homogenerity of a sample. If the sample is completely homogeneous

the entropy is zero & if the sample is an equally divide it has entropy of one.

To build a decision tree, we need to calculate two types of entropy using frequency table as follows.

a) Entropy using the frequency table of one attribute

$$E(S) = \sum_{i=1}^{c} -p_i \log_2 p_i$$

b) Entropy using the frequency table of view attributes

$$E(T, x) = \sum_{c \in x} p(c) E(c)$$

Information Gain :-

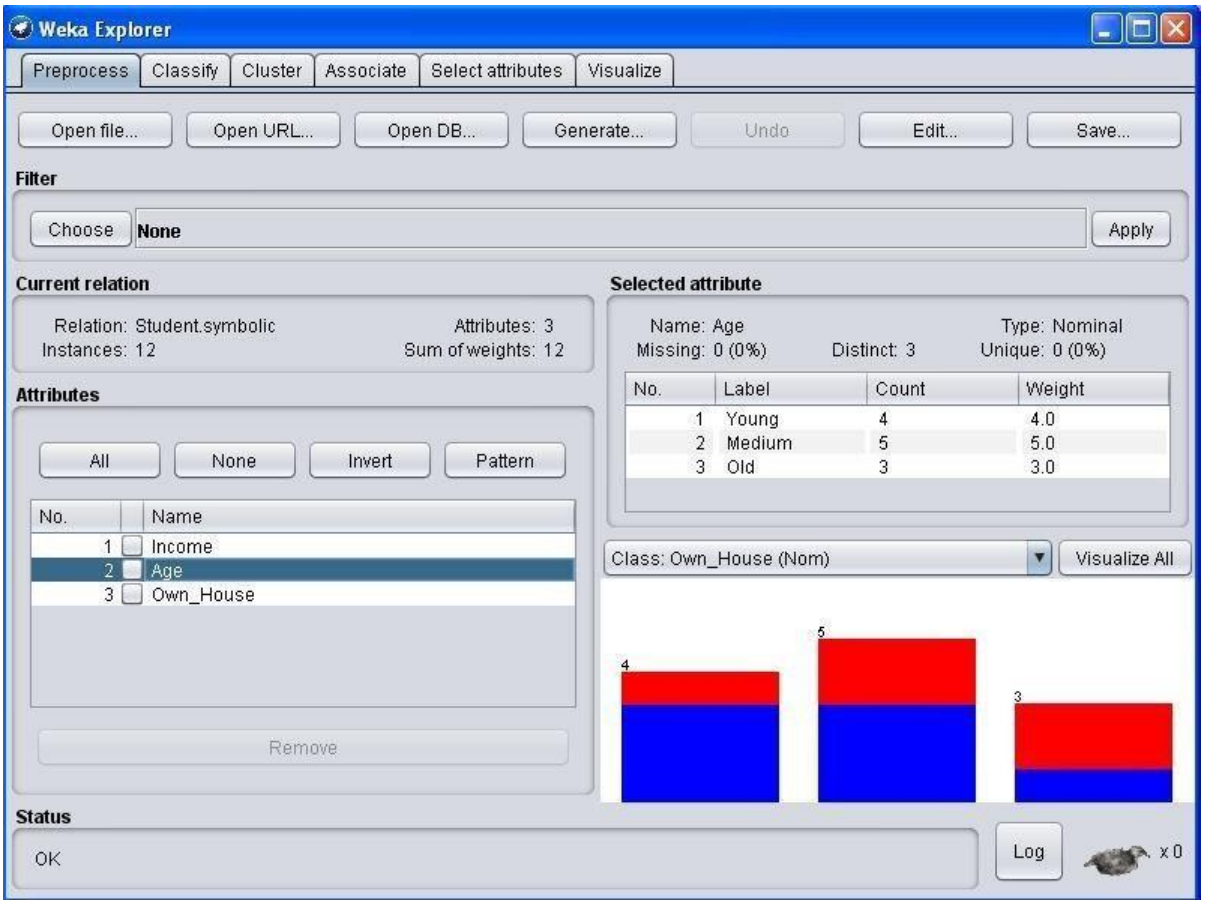
The information gain is based on the decrease in entropy after a dataset is split on an attribute constructing a decision tree is all about finding attribute that returns the highest information gain.

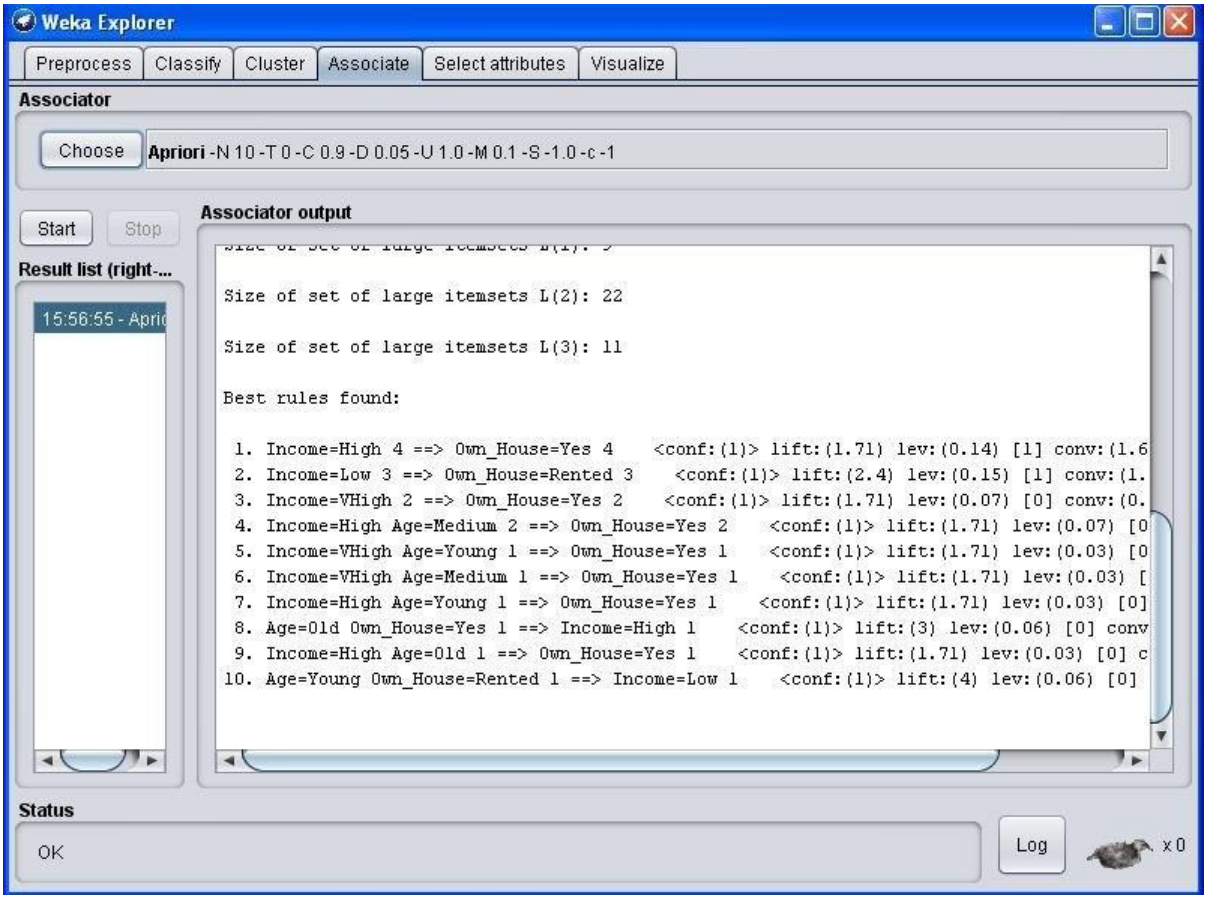
Conclusion :-

Thus we have implemented classification via decision tree using using WEKA Tool.

Output :-

19_Sanket Chandrashekhar Harvande





19_Sanket Chandrashekhar Harvande

student.arff - WordPad

File Edit View Insert Format Help

```
@relation Student.symbolic

@attribute Income {VHigh, High,Low, Medium}
@attribute Age{Young, Medium, Old}
@attribute Own_House{Yes, Rented}

@data
VHigh, Young, Yes
High, Medium, Yes
Low, Young, Rented
High, Medium, Yes
VHigh, Medium, Yes
Medium, Young, Yes
High, Old, Yes
Medium, Medium, Rented
Low, Medium, Rented
Low, Old, Rented
High, Young, Yes
Medium, Old, Rented
```