# Experiment - 3

Aim:- Perform data preprocessing tasks and Demonstrate performing association rule mining on data sets

→ Explore various options available in Weka for preprocessing data and apply Unsupervised filters like Discretization, Resample Filter etc on each dataset

→ Load weather, nominal, Iris, Glass datasets into Weka and run Apriori. Algorithm with different support and confidence values

→ study the rules generated. Apply different discretization filters on numerical attributes and run the Apriori association rule algorithm. study the rules generated.

→ Derive interesting insights and observe the effect of discretization in the rule generation process.

Objectives:-

Data preprocessing is essential before its actual use The dataset is preprocessed in order to check missing values, noisy data and other inconsistencies before executing it to the algorithm

Unsupervised Filter:-

A filter that adds a new nominal attribute representing the cluster assigned to each instance by the specified clustering algorithm.

Apriori Algorithm:-

Apriori algorithm refers to an algorithm that is used in mining frequent product sets and relevant association rules. Generally the apriori algorithm operates on a database containing a huge number of transactions.

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 free software available under the GNU 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 algorithm.

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 Locator address for where the data is stored

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

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

Current Relation: Once some data has been loaded, the Preprocess panel shows a variety of information.

→ Relation:- The name of the relation, as given in the file it was loaded from. Filters modify the name of a relation.

→ Instances:- The no. of instances in the data

→ Attributes:- the no. of attributes in the data.

Association Rule:-

An association rule has two parts, an antecedent and a consequent. An antecedent is an item found in the data. A consequent is an item that is found in combination with the antecedent.

Association rule are created by analyzing data for frequent if then patterns and using the criteria support and confidence to identify the most important

relationships. support is an indication of how frequently the items appear in the database. confidence indicates the number of times the statements have been found to be true.

Support and Confidence Values:-

- Support count:-

The support count of an itemset X, denoted by X.count, in a data set T is the number of transactions in T that contain X. Assume T has n transactions.

Then

$$\text{support} = \frac{(X \cup Y).count}{n}$$

$$\text{confidence} = \frac{(X \cup Y).count}{X.count}$$

support = support(dAUCy)

confidence = support(dAUCy) / support(dAy)

Steps Required:-

1. Open WEKA Tool.
2. click on WEKA Explorer
3. click on Preprocessing tab button.
4. click on open file button.
5. choose WEKA folder in C drive
6. select and click on data option button.
7. choose labor data set and open file.
8. choose filter button and select the Unsupervised - Discretize option and apply
9. click on Associate tab and choose Aprior algorithm
10. click on start button.

Output:

| Preprocess | classify | cluster | Associate | select Attributes | Visualize |
|---|---|---|---|---|---|
| open file | open URL | open DB | Generate | Undo | Edit | save |

Filter

Choose | None | Apply | Stop

| Preprocess | classify | cluster | Associate | Select Attributes | Visualize |
|---|---|---|---|---|---|
| open file | open URL | open DB | Generate | undo | edit | save |

Filter

Choose | Discretize -B 10 -M -1.0 -R first-last -precision 6 | apply | Stop

Current relation

| No. | Name |
|---|---|
| 1 | outlook |
| 2 | temperature |
| 3 | humidity |
| 4 | windy |
| 5 | play |

Select attribute

| No. | label | count | weight |
|---|---|---|---|
| 1 | Sunny | 5 | |
| 2 | overcast | | |
| 3 | rainy | | |