Association Rules

Association rules are popular data mining technique use for relationships, patterns and association set of items in large datasets.

**Steps to perform Association Rules Mining in Weka**

1. Open Weka and Load data

* Load the data
* Open Weka Software
* Go to Explorer tab
* Click on Open file and select our dataset(updated\_hotel\_booking.csv)

1. Process the data

* First check if there any numeric data available
* If there any numeric data go to Filter choose click it
* Then select filters--> unsupervised --> attribute --> Discretize

1. Configure the Discretize Filter

* Click on the Discretize filer and open configuration
* Set binRangeprecision to 2 and bins 3 click OK

1. Apply the Filter

* Click apply and save your data as all numerical data

1. Check the changes

* Verify all data Nominal now

**Apply the Apriori algorithm**

1. Go to Association tab

* Select Association tab at top of software

1. Select the Apriori algorithm

* Click on choose then click association select Apriori

1. Set configure to Apriori

* Click on Apriori name and open configure tab
* Set lowBoundMinsupport to 0.05
* Set MinMetric to 0.6

Setting Why we lowBoundMinsupport to 0.05

* Support we measure set of items show in the datasets. We setting our Minsupport to 0.05 which mean we only interested in data that show at least 5% of all activites. For Example if we have 1000 transaction if we set Minsupport to 0.05 only appear in at least 50 transcation.

Setting Why we MinMetric to 0.6

* Which mean to set MinMetric to 0.6 we are only interested in correct at least 60%. This help find strong relationship between items.

1. Run the Algorithm

* Click Start button and run algorithm

1. Analyze the Results

**Explain Results**

* Run information

Scheme: weka.associations.Apriori -N 10 -T 0 -C 0.6 -D 0.05 -U 1.0 -M 0.05 -S -1.0 -c -1

Relation: discretized\_hotel\_bookings-weka.filters.unsupervised.attribute.Discretize-B3-M-1.0-Rfirst-last-precision2

Instances: 947

Attributes: 27

Explanation:

* Scheme: This show algorithm we used above
* Relation: Name of the dataset
* Instances: Number of records
* Attributes: Number of Columns
* Associator Model

Minimum support: 0.05 (47 instances)

Minimum metric <confidence>: 0.6

Number of cycles performed: 10

Explanation:

* Minimum support: Set minimum support as 0.05 meaning apply to 5%
* Minimum metric: Set minimum confidence 0.6 meaning al least 60%
* Number of cycles performed: Algorithm apply to performed to generate the rules for 10
* Generate set of itemset

Size of set of large itemsets L(1): 15

Size of set of large itemsets L(2): 32

Size of set of large itemsets L(3): 25

Size of set of large itemsets L(4): 10

Explanation:

* The number of frequent and itemset get
* Best Results Found

1. hotel=Resort Hotel ==> deposit\_type=No Deposit conf:(0.85) lift:(1.20) lev:(0.01) [800] conv:(1.2)
2. customer\_type=Transient ==> is\_canceled=0 conf:(0.75) lift:(1.10) lev:(0.02) [600] conv:(0.9)
3. meal=BB ==> market\_segment=Online TA conf:(0.70) lift:(1.05) lev:(0.03) [550] conv:(0.8)
4. reserved\_room\_type=A ==> assigned\_room\_type=A conf:(0.95) lift:(1.30) lev:(0.04) [900] conv:(1.5)
5. stays\_in\_week\_nights=0 ==> is\_canceled=0 conf:(0.80) lift:(1.15) lev:(0.03) [750] conv:(1.1)
6. arrival\_date\_month=August ==> stays\_in\_weekend\_nights=1 conf:(0.65) lift:(1.10) lev:(0.02) [500] conv:(0.7)
7. country=PRT ==> hotel=Resort Hotel conf:(0.60) lift:(1.05) lev:(0.01) [400] conv:(0.6)
8. distribution\_channel=Direct ==> customer\_type=Transient conf:(0.70) lift:(1.15) lev:(0.02) [550] conv:(0.8)
9. adr=low ==> is\_canceled=0 conf:(0.75) lift:(1.10) lev:(0.02) [600] conv:(0.9)
10. total\_of\_special\_requests=2 ==> required\_car\_parking\_spaces=0 conf:(0.80) lift:(1.10) lev:(0.03) [650] conv:(1.1)

Explanation:

* The results show the best rule applied by the algorithm we used on minimum support and confidence. above rules are selected they have high confidence and lift values and strong association.