**Big Data Analytics**

****

**Assignment #4**

Finding Frequent Itemsets using SON Algorithm

**Group Members: Arhum Ali Sheikh 18L-1870**

**Hussam-ul-Hussain 18L-1827**

**Submitted To: Dr. Zareen Alamgir**

**Question # 1**

1. **Data Pre-Processing**

We have selected 10 attributes from 17 attributes.

We have removed these 7 attributes:

1. Day
2. Month
3. Pdays
4. Campaign
5. Previous
6. Poutcome
7. Contact

These attributes were impacting badly on the lift and confidence of the rules and never appeared in the top rules.

Then we applied Discretization technique on the remaining attributes, using 10 bins. After discretization, all numeric attributes were now converted into different ranges from -inf to inf. Then we applied Apriori algorithm generating 100 rules of minimum lift 2. The minimum support was set to 0.15 by default.

1. **Association Rule Mining**

Apriori algorithm generated 100 rules, and pruned all those rules which had less than 0.15 support.

Out of those 80 rules, we have selected the following 10 rules, which have highest lift and confidence and also made sense.

1. **job=management balance='(-inf-2995.6]' loan=no y=no 5846 ==> education=tertiary default=no 4698 conf:(0.8) < lift:(2.77)> lev:(0.07) [3003] conv:(3.61)**

This rule implies:

If client’s job is in management, balance is in the range of – infinity to 2995, has not got a loan, and has not subscribed to a term deposit, then we can imply that client’s education can be tertiary and has not defaulted on payments.

We can target such clients for they will not be liability for the bank and prove to be a lucrative deal.

1. **job=management default=no balance='(-inf-2995.6]' y=no 6751 ==> education=tertiary 5508 conf:(0.82) < lift:(2.77)> lev:(0.08) [3521] conv:(3.83)**

This rule implies:

If client’s job is in management, has not defaulted, has a balance is in the range of – infinity to 2995, and has not subscribed to a term deposit, then we can imply that client’s education can be tertiary.

Those who have done higher studies are likely not to default on payments, keep a balance in the said range which is good for business.

1. **job=management loan=no y=no 6996 ==> education=tertiary default=no duration='(-inf-491.8]' 5215 conf:(0.75) < lift:(2.94)> lev:(0.08) [3440] conv:(2.93)**

This rule implies:

If client’s job is in management, and has not got a loan while obviously not to a term deposit, then we can imply that client’s education can be tertiary and can be contacted for a duration of up to 491 seconds.

People falling on these criteria might be business prospects. They will listen to campaign calls lasting almost 491, meaning higher chance of making them a customer.

1. **job=management 9458 ==> education=tertiary 7801 conf:(0.82) < lift:(2.8)**

This rule implies:

If job is management, then education is tertiary. There is more than 80% chance of this to happen from the data.

1. **job=management loan=no 8205 ==> education=tertiary 6787 conf:(0.83) < lift:(2.81)> lev:(0.1) [4373] conv:(4.08)**

This rule implies:

If job is management and loan has not been taken, then education must be tertiary. This rule has a high weight with 83% confidence and a lift of 2.81

1. **job=management y=no 8157 ==> education=tertiary 6667 conf:(0.82) < lift:(2.78)> lev:(0.09) [4267] conv:(3.86)**

This rule implies:

If job is management and the client has subscribed to a term deposit, then we can imply that the client’s education is tertiary. This rule has an 82% confidence with a high lift of 2.78.

1. **job=management balance='(-inf-2995.6]' duration='(-inf-491.8]' 6980 ==> education=tertiary y=no 5188 conf:(0.74) < lift:(2.97)> lev:(0.08) [3442] conv:(2.92)**

This rule implies:

If client’s job is in management with a balance in the range of 0-2995, and was contacted for a duration in the range 0-491 seconds, then we can imply that client’s education can be tertiary and has not subscribed to a term deposit.

This rule has a high percent chance of recurring while has a good lift.

1. **job=management duration='(-inf-491.8]' 8303 ==> education=tertiary default=no y=no 6037 conf:(0.73) < lift:(2.96)> lev:(0.09) [3994] conv:(2.76)**

This rule implies:

If client’s job is in management, and was contacted for a duration in the range 0-491 seconds, then we can imply that client’s education can be tertiary and has not defaulted on payments, also has not subscribed to a term deposit.

This rule has a high percent chance of recurring while has a good lift.

1. **job=management balance='(-inf-2995.6]' 7929 ==> education=tertiary y=no 5624 conf:(0.71) < lift:(2.84)> lev:(0.08) [3641] conv:(2.58)**

This rule implies:

If client’s job is in management, and has balance up till 2995, then we can imply that client’s education can be tertiary and has not subscribed to a term deposit.

This rule has a high percent chance of recurring while has a good lift.

1. **job=management y=no 8157 ==> education=tertiary duration='(-inf-491.8]' 6146 conf:(0.75) < lift:(2.93)> lev:(0.09) [4045] conv:(3.01)**

This rule implies:

If client’s job is in management, has not subscribed to a term deposit, then we can imply that client’s education can be tertiary and will listen to a call lasting a duration of up to 491 seconds.

This rule has a high percent chance of recurring while has a good lift.