

Data mining assignment

Task 1

Use the Apriori method to generate association rules as follows:

Metric Type = Confidence, Minimum Metric Value = 0.9, Number of Rules = 10

Question 1: Explain the top 5 rules in simple English.

Instances: 4627 (transactions)

Attributes: 105

Best rules found:

1. biscuits=t frozen foods=t pet foods=t milk-cream=t vegetables=t 516 ==> bread and cake=t 475
conf:(0.92)
This rule implies that out of 4627 transactions, 516 of them contained biscuits, frozen foods, pet foods, milk-cream and vegetables along with other items. In these 516 transactions, 475 of them contained bread and cake. So, a customer who bought biscuits, frozen foods, pet foods, milk-cream and vegetables is likely to buy bread and cake with a confidence of $475/516 = 0.92$
2. baking needs=t biscuits=t milk-cream=t margarine=t fruit=t vegetables=t 505 ==> bread and cake=t 464 conf:(0.92)
This rule implies that out of 4627 transactions, 505 of them contained baking needs, biscuits, milk-cream, margarine, fruit and vegetables along with other items. In these 505 transactions, 464 of them contained bread and cake. So, a customer who bought baking needs, biscuits, milk-cream, margarine, fruit and vegetables is likely to buy bread and cake with a confidence of $464/505 = 0.92$
3. biscuits=t frozen foods=t milk-cream=t margarine=t vegetables=t 585 ==> bread and cake=t 537
conf:(0.92)
This rule implies that out of 4627 transactions, 585 of them contained biscuits, frozen foods, milk-cream, margarine and vegetables along with other items. In these 585 transactions, 537 of them contained bread and cake. So, a customer who bought biscuits, frozen foods, milk-cream, margarine and vegetables is likely to buy bread and cake with a confidence of $537/585 = 0.92$
4. biscuits=t canned vegetables=t frozen foods=t fruit=t vegetables=t 536 ==> bread and cake=t 492 conf:(0.92)
This rule implies that out of 4627 transactions, 536 of them contained biscuits, canned vegetables, frozen foods, fruit and vegetables along with other items. In these 536 transactions, 492 of them contained bread and cake. So, a customer who bought biscuits, canned vegetables,

frozen foods, fruit and vegetables is likely to buy bread and cake with a confidence of $492/536=0.92$

5. baking needs=t frozen foods=t milk-cream=t margarine=t fruit=t vegetables=t 517 ==> bread and cake=t 474 conf:(0.92)

This rule implies that out of 4627 transactions, 517 of them contained baking needs, frozen foods, milk-cream, margarine, fruit and vegetables along with other items. In these 517 transactions, 474 of them contained bread and cake. So, a customer who bought baking needs, frozen foods, milk-cream, margarine, fruit and vegetables is likely to buy bread and cake with a confidence of $474/517=0.92$

Task 2

Generate, using the Apriori method, rules that will predict the attribute **total** based on other attributes (Hint: Use the **car** option with a minimum confidence of 0.8. Note that this attribute has two values, namely **low** and **high**).

Question 2: Explain the top 5 rules in simple English.

Instances: 4627 (transactions)

Attributes: 106

Best rules found:

1. baking needs=t biscuits=t sauces-gravy-pkle=t frozen foods=t tissues-paper prd=t 574 ==> total=high 470 conf:(0.82)

The above rule implies that out of the 4627 transactions, 574 of them contained baking needs, biscuits, sauces-gravy-pkle, frozen foods and tissues-paper prd along with other items. The customer paid a total amount that was **high**, with a confidence of $470/574=0.82$.

2. bread and cake=t biscuits=t sauces-gravy-pkle=t frozen foods=t tissues-paper prd=t 600 ==> total=high 491 conf:(0.82)

The above rule implies that out of the 4627 transactions, 600 of them contained bread and cake, biscuits, sauces-gravy-pkle, frozen foods, tissues-paper prd along with other items. The customer paid a total amount that was **high**, with a confidence of $491/600=0.82$.

3. bread and cake=t baking needs=t sauces-gravy-pkle=t frozen foods=t tissues-paper prd=t. 620 ==> total=high 506 conf:(0.82)

The above rule implies that out of the 4627 transactions, 620 of them contained bread and cake, baking needs, sauces-gravy-pkle, frozen foods, tissues-paper prd along with other items. The customer paid a total amount that was **high**, with a confidence of $506/620=0.82$.

4. bread and cake=t baking needs=t biscuits=t sauces-gravy-pkle=t tissues-paper prd=t 595 ==> total=high 483 conf:(0.81)
The above rule implies that out of the 4627 transactions, 595 of them contained bread and cake, baking needs, biscuits, sauces-gravy-pkle, tissues-paper prd along with other items. The customer paid a total amount that was **high**, with a confidence of $483/595=0.81$.
5. bread and cake=t biscuits=t sauces-gravy-pkle=t tissues-paper prd=t vegetables=t 583 ==> total=high 469 conf:(0.8)
The above rule implies that out of the 4627 transactions, 583 of them contained bread and cake, biscuits, sauces-gravy-pkle, tissues-paper prd, vegetables along with other items. The customer paid a total amount that was **high**, with a confidence of $469/583=0.8$.

Question 3: In what way are the rules generated by this task useful?

1. It helps in discovery of the items that were purchased frequently.
2. It helps us to predict whether the total amount for the frequently bought items is **low, mid** or **high**. This helps us to know whether a business is flourishing well or not - Since **high** priced items are being bought frequently, it is safe to make a comment stating that the organization is making profit.