**PA 3: Association Analysis - Apriori Algorithm**

**Report on Bread Basket items**

**Student Details**

Shivani Manojkumar Panchiwala – 1001982478

Kuldip Rameshbhai Savaliya – 1001832000

Meghaben Ghanshyambhai Patel – 1002006777

**Apriori Algorithm:** To extract frequent items from the itemset and derive association rules, the apriori dataset is used.

The Apriori algorithm has the following two parameters:

1. Support: This refers to how frequently the element appears in the dataset.
2. Confidence: A conditional Probability of that item is confidence.

**The Given Dataset:**

|  |  |
| --- | --- |
| Transaction ID | Items |
| 1 | Bread, Scandinavian |
| 2 | Scandinavian, Cookies |
| 3 | Scandinavian, Hot chocolate |
| 4 | Bread, Scandinavian, Cookies |
| 5 | Bread, Hot chocolate |
| 6 | Scandinavian, Hot chocolate |
| 7 | Bread, Hot chocolate |
| 8 | Bread, Scandinavian, Hot chocolate, Muffin |
| 9 | Bread, Scandinavian, Hot chocolate |

Minimum Support = 2, Minimum Confidence = 50%

Find the frequently occurring item sets and create association rules now.

**Step 1:**

We'll refer to this below table as S1.

|  |  |
| --- | --- |
| ITEMS | SUPPORT COUNT |
| Bread | 6 |
| Scandinavian | 7 |
| Hot chocolate | 5 |
| Cookies | 2 |
| Muffin | 1 |

The following step is to eliminate every element whose support count is less than the minimal support.

* We'll refer to this below table as T1.

|  |  |
| --- | --- |
| ITEMS | SUPPORT COUNT |
| Bread | 6 |
| Scandinavian | 7 |
| Hot chocolate | 5 |
| Cookies | 2 |

**Step 2:**

* We will now produce S2. In this, we will first create the pair and then record their frequency.
* We'll refer to this below table as S2.

|  |  |
| --- | --- |
| ITEMS | SUPPORT COUNT |
| {Bread, Scandinavian} | 4 |
| {Bread, Hot chocolate} | 4 |
| {Bread, Cookies} | 1 |
| {Scandinavian, Hot chocolate} | 4 |
| {Scandinavian, Cookies} | 2 |
| {Cookies, Muffin} | 0 |

* We will now eliminate all rows with support counts below the minimal support count.
* We'll refer to this below table as T2.

|  |  |
| --- | --- |
| ITEMS | SUPPORT COUNT |
| {Bread, Scandinavian} | 4 |
| {Bread, Hot chocolate} | 4 |
| {Scandinavian, Hot chocolate} | 4 |
| {Scandinavian, Cookies} | 2 |

**Step 3:**

* We will now create an S3 table. This will provide the cumulative frequency of the three things.

|  |  |
| --- | --- |
| ITEMS | SUPPORT COUNT |
| {Bread, Scandinavian, Hot chocolate} | 2 |
| {Scandinavian, Hot chocolate, Cookies} | 1 |
| {Bread, Hot chocolate, Cookies} | 0 |
| {Bread, Scandinavian, Cookies} | 0 |

* We will now once again eliminate any rows with a support count below the required minimum.
* There is only one row left after that.

|  |  |
| --- | --- |
| {Bread, Scandinavian, Hot chocolate} | 2 |

**Step 4:**

Finding the association rules for the subsets:

|  |  |  |
| --- | --- | --- |
| RULES | SUPPORT | CONFIDENCE |
| Bread ^ Scandinavian → Hot chocolate | 2 | Sup {(Bread ^ Scandinavian) ^ Hot chocolate}/Sup (Bread ^ Scandinavian) = 2/4 = 0.5 = 50% |
| Scandinavian ^ Hot chocolate → Bread | 2 | Sup {(Scandinavian ^ Hot chocolate) ^ Bread}/Sup (Scandinavian ^ Hot chocolate) = 2/4 = 0.5 = 50% |
| Bread ^ Hot chocolate → Scandinavian | 2 | Sup {(Bread ^ Hot chocolate) ^ Scandinavian}/Sup (Bread ^ Hot chocolate)  = 2/4 = 0.5 = 50% |
| Hot chocolate → Bread ^ Scandinavian | 2 | Sup {(Hot chocolate ^ (Bread ^ Scandinavian)}/Sup (Hot chocolate) = 2/5 = 0.4 = 40% |
| Bread → Scandinavian ^ Hot chocolate | 2 | Sup {(Bread ^ (Scandinavian ^ Hot chocolate)}/Sup (Bread) = 2/6 = 0.33 = 33.33% |
| Scandinavian → Scandinavian ^ Hot chocolate | 2 | Sup {(Scandinavian ^ (Scandinavian ^ Hot chocolate)}/Sup (Scandinavian) = 2/7 = 0.28  = 28% |

As a result, the stated minimum level of confidence is 50%, and the above table reveals that only three rows have a level of confidence more than or equal to the specified minimum level of confidence. So, the *Bread ^ Scandinavian → Scandinavian, Scandinavian ^ Scandinavian → Bread and Bread ^ Hot chocolate → Scandinavian,* can be considered as the **Strong association rules.**

**Evaluated Results:**

Case 1:

A picture containing graphical user interface

Description automatically generated

Text, letter

Description automatically generated

A picture containing text

Description automatically generated

Case 2:

A picture containing text

Description automatically generated

Text, letter

Description automatically generated A picture containing text

Description automatically generated

Case 3:

Graphical user interface, text, application

Description automatically generated

Text

Description automatically generated

Text

Description automatically generated

Text

Description automatically generated

References:

<https://www.kaggle.com/datasets/mittalvasu95/the-bread-basket>

<https://www.geeksforgeeks.org/apriori-algorithm/>

<https://www.softwaretestinghelp.com/apriori-algorithm>

<https://www.educative.io/edpresso/what-is-the-apriori-algorithm>