Transaction Data

|  |  |
| --- | --- |
| **TID** | **ItemID** |
| T100 | I1, I2, I5 |
| T200 | I2, I4 |
| T300 | I2, I3 |
| T400 | I1, I2, I4 |
| T500 | I1, I3 |
| T600 | I2, I3 |
| T700 | I1, I3 |
| T800 | I1, I2, I3, I5 |
| T900 | I1, I2, I3 |

Perform **Apriori analysis** with **minimum support count of 2**

* Minimum support is a **threshold** that determines the minimum frequency at which an itemset must appear in a dataset to be considered **significant**.
* It helps filter out itemsets that are **too rare** to be useful for association rule mining.
* Minimum support count or absolute minimum support refers to the frequency directly.
* There is also **relative minimum support**, which is in **percentage**. It is the minimum number of times an itemset must appear, divided by the total amount of transactions.

Step 1: count the frequency of each item

This is called C1 (candidate 1-itemsets)

|  |  |
| --- | --- |
| **Item** | **Support Count** |
| I1 | 6 |
| I2 | 7 |
| I3 | 6 |
| I4 | 2 |
| I5 | 2 |

Step 2: remove the items that are below the minimum support count

* Since all is above the minimum support count (2), we will take all
* K is the amount of items in each itemset

Itemset L1 (K=1)

|  |  |
| --- | --- |
| **Item** | **Support Count** |
| I1 | 6 |
| I2 | 7 |
| I3 | 6 |
| I4 | 2 |
| I5 | 2 |

Step 3: we will pair them up (2 items per set)

* This is done by L1 join L1🡪 combination

This is called C2 (candidate 2-itemsets)

|  |
| --- |
| **Itemset** |
| I1, I2 |
| I1, I3 |
| I1, I4 |
| I1, I5 |
| I2, I3 |
| I2, I4 |
| I2, I5 |
| I3, I4 |
| I3, I5 |
| I4, I5 |

Step 4: count the frequency of each itemset

|  |  |
| --- | --- |
| **Itemset** | **Support Count** |
| I1, I2 | 4 |
| I1, I3 | 4 |
| I1, I4 | 1 |
| I1, I5 | 2 |
| I2, I3 | 4 |
| I2, I4 | 2 |
| I2, I5 | 2 |
| I3, I4 | 0 |
| I3, I5 | 1 |
| I4, I5 | 0 |

Step 5: remove the itemsets that are below the minimum support count

Itemset L2 (K=2)

|  |  |
| --- | --- |
| **Itemset** | **Support Count** |
| I1, I2 | 4 |
| I1, I3 | 4 |
| I1, I5 | 2 |
| I2, I3 | 4 |
| I2, I4 | 2 |
| I2, I5 | 2 |

Step 6: make groups of 3 🡪 use apriori generation method

1. We will look at the first K-1 items that are the same. Since K = 2, therefore we will look at the itemsets that have the same first item. There are two first items: I1 & I2

We will look at I1 first

|  |
| --- |
| **Itemset** |
| I1, I2 |
| I1, I3 |
| I1, I5 |

1. We will remove the first K-1 item.

Since K = 2, then delete the first item ({I1}) 🡪 {I2}, {I3}, {I5}

1. List all possible combinations

|  |
| --- |
| **Combinations** |
| I2, I3 |
| I2, I5 |
| I3, I5 |

1. Lastly, add the previously deleted item

|  |  |
| --- | --- |
| **Combinations** | **Itemsets** |
| I2, I3 | I1, I2, I3 |
| I2, I5 | I1, I3, I5 |
| I3, I5 | I1, I2, I5 |

Now do this for I2 also

|  |  |  |  |
| --- | --- | --- | --- |
| **Itemset** | **Remove K-1 item** | **Combination** | **Itemset** |
| I2, I3 | I3 | I3, I4 | I2, I3, I4 |
| I2, I4 | I4 | I3, I5 | I2, I3, I5 |
| I2, I5 | I5 | I4, I5 | I2, I4, I5 |

This is called C3 (Candidate 3-itemsets)

|  |  |
| --- | --- |
| **Itemset** | **Support Count** |
| I1, I2, I3 | 2 |
| I1, I2, I5 | 2 |
| I1, I3, I5 | 1 |
| I2, I3, I4 | 0 |
| I2, I3, I5 | 1 |
| I2, I4, I5 | 0 |

Step 7: remove the itemsets that are below the minimum support count

Itemset L3 (K = 3)

|  |  |
| --- | --- |
| **Itemset** | **Support Count** |
| I1, I2, I3 | 2 |
| I1, I2, I5 | 2 |

Step 8: make groups of 4 🡪 use apriori generation method

1. We will look at the first K-1 items that are the same. Since K = 3, therefore we will look at the first 2 items that are the same. See it from itemset L3. Therefore, the first 2 items that are the same are {I1, I2}
2. We will remove the first K-1 items

Remove the first 2 items {I1, I2} 🡪 {I3}, {I5}

1. Get the combinations 🡪 {I3, I5}
2. Add the previously deleted items 🡪 {I1, I2, I3, I5}
3. Count the frequency

This is called C3 (Candidate 3-itemsets)

|  |  |
| --- | --- |
| **Itemset** | **Support Count** |
| I1, I2, I3, I5 | 1 |

1. Remove those that are below the minimum support

Therefore, L4 does not exist

Step 9: calculate support

* Support = support count / total transactions

|  |  |  |
| --- | --- | --- |
| **Itemset** | **Support Count** | **Support** |
| I1 | 6 | 6/9 = 0.667 |
| I2 | 7 | 7/9 = 0.778 |
| I3 | 6 | 6/9 = 0.667 |
| I4 | 2 | 2/9 = 0.222 |
| I5 | 2 | 2/9 = 0.222 |
| I1, I2 | 4 | 4/9 = 0.444 |
| I1, I3 | 4 | 4/9 = 0.444 |
| I1, I5 | 2 | 2/9 = 0.222 |
| I2, I3 | 4 | 4/9 = 0.444 |
| I2, I4 | 2 | 2/9 = 0.222 |
| I2, I5 | 2 | 2/9 = 0.222 |
| I1, I2, I3 | 2 | 2/9 = 0.222 |
| I1, I2, I5 | 2 | 2/9 = 0.222 |