Q1.) Apriori

1. Show the steps of running the Apriori algorithm to get frequent 1-itemsets F1, candidate 2-itemsets C2, and frequent 2-itemsets F2;

Ans.) given,

|  |  |
| --- | --- |
| Transaction ID`s | Items |
| T1 | I1, I2, I5. |
| T2 | I2, I4. |
| T3 | I2, I3. |
| T4 | I1, I2, I4. |
| T5 | I3, I4. |
| T6 | I1, I3. |
| T7 | I1, I2, I3, I5. |
| T8 | I2, I3, I4. |
| T9 | I2, I3, I5. |
| T10 | I3, I5. |

Minimum Support Threshold = 30% = 0.3 \* 10 = 3.

Frequency Itemset F1:

|  |  |
| --- | --- |
| Itemset | Support Count |
| I1 | 4. |
| I2 | 7. |
| I3 | 7. |
| I4 | 4. |
| I5 | 4. |

Now from Frequency Itemset F1, find out the occurrences of 2-itemset.

C2:

|  |  |
| --- | --- |
| Itemset | Support Count |
| {I1, I2} | 3. |
| {I1, I3} | 2. |
| {I1, I4} | 1. |
| {I1, I5} | 2. |
| {I2, I3} | 4. |
| {I2, I4} | 3. |
| {I2, I5} | 3. |
| {I3, I4} | 2. |
| {I3, I5} | 3. |
| {I4, I5} | 0. |

Since the minimum support threshold is 3, any itemset < 3 is cancelled.

Frequency Itemset F2:

|  |  |
| --- | --- |
| Itemset | Support Count |
| {I1, I2} | 3. |
| {I2, I3} | 4. |
| {I2, I4} | 3. |
| {I2, I5} | 3. |
| {I3, I5} | 3. |

1. From F2, derive all the association rules in the form of α 🡪 β (α ≠ 0 and β ≠ 0 ) that satisfy the confidence threshold;

Ans.) (I1, I2)

{I1} 🡪 {I2}

count (I1, I2) : 3.

count (I1) : 4.

We know that,

Therefore, confidence = ¾ = 0.75.

{I2} 🡪 {I1}

Count (I1, I2) : 3.

Count (I2) : 7.

Therefore, confidence = 3/7 = 0.42.

(I2, I3)

{I2} 🡪 {I3}

Count (I2, I3) : 4.

Count (I2) : 7.

Therefore, confidence = 4/7 = 0.57.

{I3} 🡪 {I2}

Count (I2, I3) : 4.

Count (I3): 7.

Therefore, confidence = 4/7 = 0.57.

(I2, I4)

{ I2} 🡪 {I4}

Count (I2, I4) : 3.

Count (I2) : 7.

Therefore, confidence = 3/7 = 0.42.

{I4} 🡪 {I2}

Count (I2, I4) : 3.

Count (I4) : 4.

Therefore, confidence = ¾ = 0.75.

(I2, I5)

{I2} 🡪 {I5}

Count (I2, I5) : 3.

Count ( I2) : 7.

Therefore, confidence = 3/7 = 0.42.

{I5} 🡪 {I2}

Count (I2, I5) : 3.

Count (I5) : 4.

Therefore, confidence = ¾ = 0.75.

(I3, I5)

{I3} 🡪 {I5}

Count (I3, I5) : 3.

Count (I3) : 7.

Therefore, confidence = 3/7 = 0.42.

{I5} 🡪 {I3}

Count (I3, I5) : 3.

Count (I5) : 4.

Therefore, confidence = ¾ = 0.75.

Therefore, the association rules that are greater than the minimum confidence threshold are:

{I1} 🡪 {I2}

{I4} 🡪 {I2}

{I5} 🡪 {I2}

{I5} 🡪 {I3}

1. Draw the hash tree for the candidate 2-itemsets by using the hash function x mod 3 where x is the digit in an item Ix. This hash tree does not need to be a full hash tree. You just need to create ONE two-level branch and all the other branches should contain only one level.

Ans.) Diagram

Description automatically generated

Q2.) FP – Growth

1. Show the steps of constructing the FP-tree.

Ans.) Step 1: Scan the DB once, and find the frequent 1- itemset and sort the items in frequency descending order.

<(I2 : 7), (I3 : 7), (I1 : 4), (I4 : 4), (I5 : 4)>

And then order the itemset.

|  |  |  |
| --- | --- | --- |
| TID | Items | Ordered Frequent Items |
| T1 | I1, I2, I5 | I2, I1, I5 |
| T2 | I2, I4 | I2, I4 |
| T3 | I2, I3 | I2, I3 |
| T4 | I1, I2, I4 | I2, I1, I4 |
| T5 | I3, I4 | I3, I4 |
| T6 | I1, I3 | I3, I1 |
| T7 | I1, I2, I3, I5 | I2, I3, I1, I5 |
| T8 | I2, I3, I4 | I2, I3, I4 |
| T9 | I2, I3, I5 | I2, I3, I5 |
| T10 | I3, I5 | I3, I5 |

Step 2: Scan the DB again, and construct the FP- tree.

Header Table:

|  |  |
| --- | --- |
| Items | Head of node-links |
| **I2** |  |
| **I3** |  |
| **I1** |  |
| **I4** |  |
| **I5** |  |

Background pattern

Description automatically generated

1. Derive the frequent item sets from the SECOND item in the header table. Note that you do NOT need to show all frequent item sets (or patterns).

Ans.) Find the conditional pattern base.

|  |  |
| --- | --- |
| Item | Conditional Pattern Base |
| I3 | I2 : 4 |
| I1 | I2 : 2, I2I3 : 1 |
| I4 | I2 : 1, I2I1 : 1, I3 : 1, I2I3 : 1 |
| I5 | I2I1 : 1, I2I3I1 : 1, I2I3 : 1, I3 : 1 |

Frequent item sets (Patterns):

All frequent Patterns,

Frequent Patterns relate to I5:

I2I1I5, I2I3I1I5, I2I3I5, I3I5, I5.

Frequent Patterns relate to I4:

I2I4, I2I1I4, I3I4, I2I3I4, I4.

Frequent Patterns relate to I1:

I2I1, I1.

Frequent Patterns relate to I3:

I2I3, I3.