|  |  |
| --- | --- |
| TransactionID | 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 **FP Growth analysis** with **minimum support count of 2** to produce **frequent itemsets.**

Step 1: count the frequency of each item

This is called C1 (candidate 1-itemsets)

|  |  |
| --- | --- |
| ItemID | Support Count |
| I1 |  |
| I2 |  |
| I3 |  |
| I4 |  |
| I5 |  |

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

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

Itemset L1 (K=1)

|  |  |
| --- | --- |
| ItemID | Frequency |
| I1 |  |
| I2 |  |
| I3 |  |
| I4 |  |
| I5 |  |

Step 3: Sort in descending order

|  |  |
| --- | --- |
| ItemID | Frequency |
| I2 |  |
| I1 |  |
| I3 |  |
| I4 |  |
| I5 |  |

Step 4: For each transaction, sort the itemID based on the frequency

|  |  |
| --- | --- |
| TransactionID | ItemID |
| T100 |  |
| T200 |  |
| T300 |  |
| T400 |  |
| T500 |  |
| T600 |  |
| T700 |  |
| T800 |  |
| T900 |  |

Step 5: Make Tree

1. The tree starts with a root, where its value is NULL
2. Add the items based on the already sorted itemID
   1. If the node of the entered item has not been created, then add a new node of the entered item with its number count as 1.
   2. If the node of the entered item has been created, then increase the number count of the node by one.

A diagram of a team

Description automatically generated

Step 6: make Conditional Pattern Base

* Conditional Pattern Base is the set of paths that end with the selected item, followed by the number of times the path ends with that item.

|  |  |
| --- | --- |
| Items | Conditional Pattern Base |
| I2 | - |
| I1 |  |
| I3 |  |
| I4 |  |
| I5 |  |

Step 7: make Conditional FP-Tree

* Conditional FP-Tree is a set of paths from the **Conditional Pattern Base** with a support count **greater than or equal** to the **minimum support**.
* If they start with the same path, we can add them together
  + For example, I3 🡪{I1:2}, {I2:2}, {I2,I1:2}
  + For {I2:2} and {I2,I1:2}, we can add the I2 together since they both start with I2
  + {I2:2} and {I2,I1:2} 🡪 {I2:4, I1:2}
  + The I1 is separated because it is a different branch
* If the frequency is below the minimum support count, we can throw them away
  + For example, I4 🡪 {I2:1}, {I2,I1:1}
  + we can add the I2 together since they both start with I2
  + {I2:1} and {I2,I1:1} 🡪 {I2:2, I1:1}
  + The I1 is separated because it is a different branch
  + However, since I1 is below the minimum support count, we will throw it away
  + {I2:1} and {I2,I1:1} 🡪 {I2:2}

|  |  |  |
| --- | --- | --- |
| Items | Conditional Pattern Base | Conditional FP-Tree |
| I2 | - | - |
| I1 | {I2:4} |  |
| I3 | {I1:2}, {I2:2}, {I2,I1:2} |  |
| I4 | {I2:1}, {I2,I1:1} |  |
| I5 | {I2,I1:1}, {I2,I1,I3:1} |  |

Step 8: Determine Frequent Pattern Generated

* It is how many times does a pair appear together
* Make every combination of the Conditional FP-Tree
* Add the item at the end
* The frequency is from the Conditional FP-Tree
* If the items are the same, add them

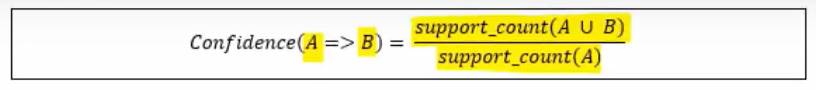
|  |  |  |
| --- | --- | --- |
| Items | Conditional FP-Tree | Frequent Pattern Generated |
| I2 | - |  |
| I1 | {I2:4} |  |
| I3 | {I2:4, I1:2}, {I1:2} |  |
| I4 | {I2:2} |  |
| I5 | {I2:2,I1:2} |  |

Step 9: Calculate the support

|  |  |  |
| --- | --- | --- |
| Itemset | Support Count | Support |
| I1 |  |  |
| I2 |  | /9 = 0. |
| I3 |  | /9 = 0. |
| I4 |  | /9 = 0. |
| I5 |  | /9 = 0. |
| I2, I1 |  | /9 = 0. |
| I2, I3 |  | /9 = 0. |
| I1, I3 |  | /9 = 0. |
| I2, I4 |  | /9 = 0. |
| I2, I5 |  | /9 = 0. |
| I1, I5 |  | /9 = 0. |
| I2, I1, I3 |  | /9 = 0. |
| I2, I1, I5 |  | /9 = 0. |

Calculate **Association Rules** if minimum confidence is **0.8**

* Minimum confidence: the threshold that a rule must meet to be considered significant
* Confidence (A => B): what are the chances that if A happens, B also happens
* The probability that item B is present in transactions that contain item A.



Step 10: split the itemset to premise (A) and conclusion (B)

* If the customer buys the premise, what are the chances it will buy the conclusion

|  |  |  |
| --- | --- | --- |
| Itemset | Premise (A) | Conclusion (B) |
| I2, I1 |  |  |
| I2, I3 |  |  |
| I1, I3 |  |  |
| I2, I4 |  |  |
| I2, I5 |  |  |
| I1, I5 |  |  |
| I2, I1, I3 |  |  |
|  |  |
|  |  |
| I2, I1, I5 |  |  |
|  |  |
|  |  |

Step 11: Calculate the support and confidence

* Also swap the premise and conclusion

|  |  |  |  |
| --- | --- | --- | --- |
| Premise (A) | Conclusion (B) | Support | Confidence |
| I2 | I1 |  |  |
| - |  |  |  |
| I2 | I3 |  |  |
| - |  |  |  |
| I1 | I3 |  |  |
| - |  |  |  |
| I2 | I4 | 2/9 = 0.222 | 2/7 = 0.286 |
| I4 | I2 | 2/9 = 0.222 | 2/2 = 1 |
| I2 | I5 | 2/9 = 0.222 | 2/7 = 0.286 |
| I5 | I2 | 2/9 = 0.222 | 2/2 = 1 |
| I1 | I5 | 2/9 = 0.222 | 2/6 = 0.333 |
| I5 | I1 | 2/9 = 0.222 | 2/2 = 1 |
| I2 | I1,I3 | 2/9 = 0.222 | 2/7 = 0.286 |
| I1,I3 | I2 | 2/9 = 0.222 | 2/4 = 0.5 |
| I1 | I2,I3 | 2/9 = 0.222 | 2/6 = 0.333 |
| I2,I3 | I1 | 2/9 = 0.222 | 2/4 = 0.5 |
| I3 | I2,I1 | 2/9 = 0.222 | 2/6 = 0.333 |
| I2,I1 | I3 | 2/9 = 0.222 | 2/4 = 0.5 |
| I2 | I1,I5 | 2/9 = 0.222 | 2/7 = 0.286 |
| I1,I5 | I2 | 2/9 = 0.222 | 2/2 = 1 |
| I1 | I2,I5 | 2/9 = 0.222 | 2/6 = 0.333 |
| I2,I5 | I1 | 2/9 = 0.222 | 2/2 = 1 |
| I5 | I2,I1 | 2/9 = 0.222 | 2/2 = 1 |
| I2,I1 | I5 | 2/9 = 0.222 | 2/4 = 0.5 |

Step 12: remove those that is below the minimum confidence

|  |  |  |  |
| --- | --- | --- | --- |
| Premise (A) | Conclusion (B) | Support | Confidence |
| I4 | I2 | 2/9 = 0.222 | 2/2 = 1 |
| I5 | I2 | 2/9 = 0.222 | 2/2 = 1 |
| I5 | I1 | 2/9 = 0.222 | 2/2 = 1 |
| I1,I5 | I2 | 2/9 = 0.222 | 2/2 = 1 |
| I2,I5 | I1 | 2/9 = 0.222 | 2/2 = 1 |
| I5 | I2,I1 | 2/9 = 0.222 | 2/2 = 1 |

fpgrowth.exe documentation

<https://borgelt.net/doc/fpgrowth/fpgrowth.html>