**Postlab 4**

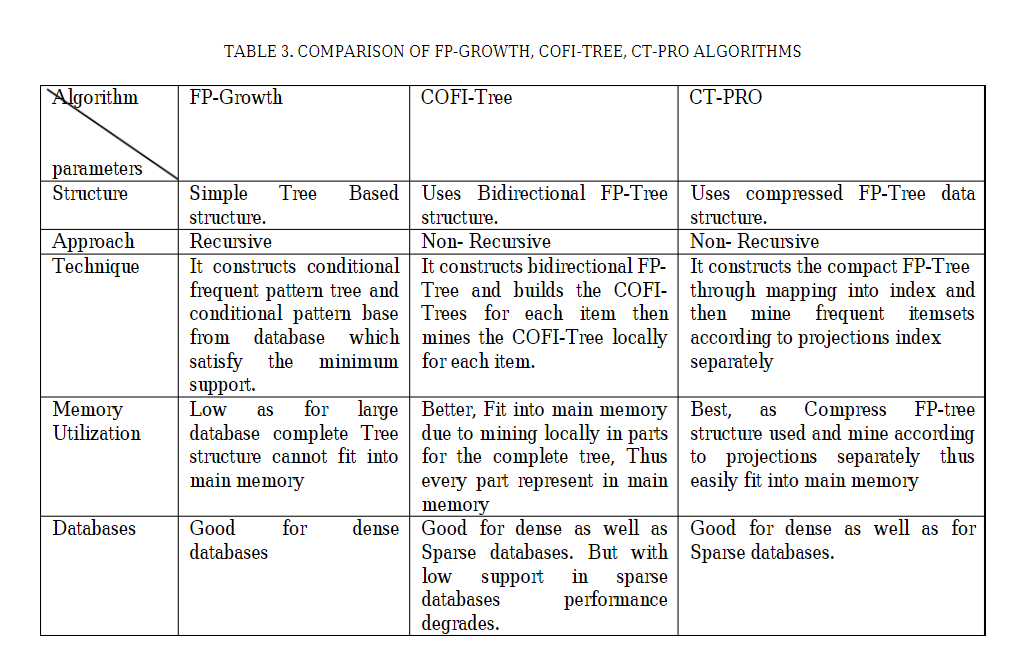
Q1) Give limitations of Apriori algorithms

Ans: -

Apriori algorithm suffers from some weakness in spite of being clear and simple. The main limitation is costly wasting of time to hold a vast number of candidates sets with much frequent item sets, low minimum support or large item sets. For example, if there are 104 from frequent 1- item sets, it needs to generate more than 107 candidates into 2-length which in turn they will be tested and accumulate. Furthermore, to detect frequent pattern in size 100 (e.g.) v1, v2… v100, it has to generate 2100 candidate item sets that yield on costly and wasting of time of candidate generation. So, it will check for many sets from candidate item sets, also it will scan database many times repeatedly for finding candidate item sets. Apriori will be very low and inefficiency when memory capacity is limited with large number of transactions.

Q2) Write different variations of FP tree algorithm

Ans: -



Q3) Explain FP tree Algorithms with one example

### Ans: -

### Frequent Pattern Growth Algorithm

This algorithm is an improvement to the Apriori method. A frequent pattern is generated without the need for candidate generation. FP growth algorithm represents the database in the form of a tree called a frequent pattern tree or FP tree.This tree structure will maintain the association between the itemsets. The database is fragmented using one frequent item. This fragmented part is called “pattern fragment”. The itemsets of these fragmented patterns are analyzed. Thus, with this method, the search for frequent itemsets is reduced comparatively.

**FP Tree**

Frequent Pattern Tree is a tree-like structure that is made with the initial itemsets of the database. The purpose of the FP tree is to mine the most frequent pattern. Each node of the FP tree represents an item of the itemset.The root node represents null while the lower nodes represent the itemsets. The association of the nodes with the lower nodes that is the itemsets with the other itemsets are maintained while forming the tree.

**Frequent Pattern Algorithm Steps**

The frequent pattern growth method lets us find the frequent pattern without candidate generation.

Let us see the steps followed to mine the frequent pattern using frequent pattern growth algorithm**:**

**#1)** The first step is to scan the database to find the occurrences of the itemsets in the database. This step is the same as the first step of Apriori. The count of 1-itemsets in the database is called support count or frequency of 1-itemset.

**#2)** The second step is to construct the FP tree. For this, create the root of the tree. The root is represented by null.

**#3)**The next step is to scan the database again and examine the transactions. Examine the first transaction and find out the itemset in it. The itemset with the max count is taken at the top, the next itemset with lower count and so on. It means that the branch of the tree is constructed with transaction itemsets in descending order of count.

**#4)** The next transaction in the database is examined. The itemsets are ordered in descending order of count. If any itemset of this transaction is already present in another branch (for example in the 1st transaction), then this transaction branch would share a common prefix to the root.This means that the common itemset is linked to the new node of another itemset in this transaction.

**#5)** Also, the count of the itemset is incremented as it occurs in the transactions. Both the common node and new node count is increased by 1 as they are created and linked according to transactions.

**#6)** The next step is to mine the created FP Tree. For this, the lowest node is examined first along with the links of the lowest nodes. The lowest node represents the frequency pattern length 1. From this, traverse the path in the FP Tree. This path or paths are called a conditional pattern base.

Conditional pattern base is a sub-database consisting of prefix paths in the FP tree occurring with the lowest node (suffix).

**#7)** Construct a Conditional FP Tree, which is formed by a count of itemsets in the path. The itemsets meeting the threshold support are considered in the Conditional FP Tree.

**#8)** Frequent Patterns are generated from the Conditional FP Tree.

Example Of FP-Growth Algorithm

**Support threshold=50%, Confidence= 60%**

**Table 1**

| **Transaction** | **List of items** |
| --- | --- |
| T1 | I1,I2,I3 |
| T2 | I2,I3,I4 |
| T3 | I4,I5 |
| T4 | I1,I2,I4 |
| T5 | I1,I2,I3,I5 |
| T6 | I1,I2,I3,I4 |

**Solution:**

Support threshold=50% => 0.5\*6= 3 => min\_sup=3

**1. Count of each item**

**Table 2**

| **Item** | **Count** |
| --- | --- |
| I1 | 4 |
| I2 | 5 |
| I3 | 4 |
| I4 | 4 |
| I5 | 2 |

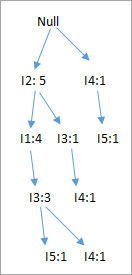
**2. Sort the itemset in descending order.**

| **Item** | **Count** |
| --- | --- |
| I2 | 5 |
| I1 | 4 |
| I3 | 4 |
| I4 | 4 |

**Table 3**

**3. Build FP Tree**

1. Considering the root node null.
2. The first scan of Transaction T1: I1, I2, I3 contains three items {I1:1}, {I2:1}, {I3:1}, where I2 is linked as a child to root, I1 is linked to I2 and I3 is linked to I1.
3. T2: I2, I3, I4 contains I2, I3, and I4, where I2 is linked to root, I3 is linked to I2 and I4 is linked to I3. But this branch would share I2 node as common as it is already used in T1.
4. Increment the count of I2 by 1 and I3 is linked as a child to I2, I4 is linked as a child to I3. The count is {I2:2}, {I3:1}, {I4:1}.
5. T3: I4, I5. Similarly, a new branch with I5 is linked to I4 as a child is created.
6. T4: I1, I2, I4. The sequence will be I2, I1, and I4. I2 is already linked to the root node, hence it will be incremented by 1. Similarly, I1 will be incremented by 1 as it is already linked with I2 in T1, thus {I2:3}, {I1:2}, {I4:1}.
7. T5:I1, I2, I3, I5. The sequence will be I2, I1, I3, and I5. Thus {I2:4}, {I1:3}, {I3:2}, {I5:1}.
8. T6: I1, I2, I3, I4. The sequence will be I2, I1, I3, and I4. Thus {I2:5}, {I1:4}, {I3:3}, {I4 1}.

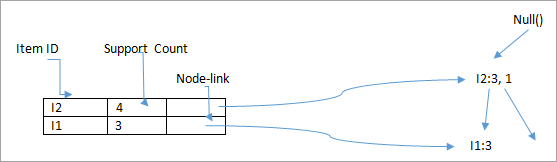
[](https://cdn.softwaretestinghelp.com/wp-content/qa/uploads/2019/09/FP-Tree.png)

**4. Mining of FP-tree is summarized below:**

1. The lowest node item I5 is not considered as it does not have a min support count, hence it is deleted.
2. The next lower node is I4. I4 occurs in 2 branches, {I2,I1,I3:,I41},{I2,I3,I4:1}. Therefore, considering I4 as suffix the prefix paths will be {I2, I1, I3:1}, {I2, I3: 1}. This forms the conditional pattern base.
3. The conditional pattern base is considered a transaction database, an FP-tree is constructed. This will contain {I2:2, I3:2}, I1 is not considered as it does not meet the min support count.
4. This path will generate all combinations of frequent patterns: {I2,I4:2},{I3,I4:2},{I2,I3,I4:2}
5. For I3, the prefix path would be: {I2,I1:3},{I2:1}, this will generate a 2 node FP-tree : {I2:4, I1:3} and frequent patterns are generated: {I2,I3:4}, {I1:I3:3}, {I2,I1,I3:3}.
6. For I1, the prefix path would be: {I2:4} this will generate a single node FP-tree: {I2:4} and frequent patterns are generated: {I2, I1:4}.

| **Item** | **Conditional Pattern Base** | **Conditional FP-tree** | **Frequent Patterns Generated** |
| --- | --- | --- | --- |
| I4 | {I2,I1,I3:1},{I2,I3:1} | {I2:2, I3:2} | {I2,I4:2},{I3,I4:2},{I2,I3,I4:2} |
| I3 | {I2,I1:3},{I2:1} | {I2:4, I1:3} | {I2,I3:4}, {I1:I3:3}, {I2,I1,I3:3} |
| I1 | {I2:4} | {I2:4} | {I2,I1:4} |

The diagram given below depicts the conditional FP tree associated with the conditional node I3.

[](https://cdn.softwaretestinghelp.com/wp-content/qa/uploads/2019/09/conditional-FP-tree-associated-with-conditional-node-I3.png)