**Sanyasi Naidu Pyla**

**801077940**

8. The Apriori algorithm uses a generate-and-count strategy for deriving frequent

itemsets. Candidate itemsets of size k+1 are created by joining a pair

of frequent itemsets of size k (this is known as the candidate generation step).

A candidate is discarded if any one of its subsets is found to be infrequent

during the candidate pruning step. Suppose the Apriori algorithm is applied

to the data set shown in Table 6.3 with minsup = 30%, i.e., any itemset

occurring in less than 3 transactions is considered to be infrequent.

Table 6.24. Example of markel basket transactions.

|  |  |
| --- | --- |
| Transaction ID | Items Bought |
| 1 | {a,b,d,e} |
| 2 | {b, c,d} |
| 3 | {a,b,d,e} |
| 4 | {a.c.d,e} |
| 5 | {b,c,d,e} |
| 6 | {b, d, e} |
| 7 | {c,d} |
| 8 | {a,b.c} |
| 9 | {a, d, e} |
| 10 | {b,d} |

**(a) Draw an itemset lattice representing the data set given in Table 6.24.**

**Label each node in the lattice with the following letter(s):**

**• N: If the itemset is not considered to be a candidate itemset by**

**the A priori algorithm. There are two reasons for an itemset not to**

**he considered as a candidate itemset: (1) it is not generated at all**

**during the candidate generation step, or (2) it is generated during**

**the candidate generation step hut is subsequently removed during**

**the candidate pruning step because one of its subsets is found to be**

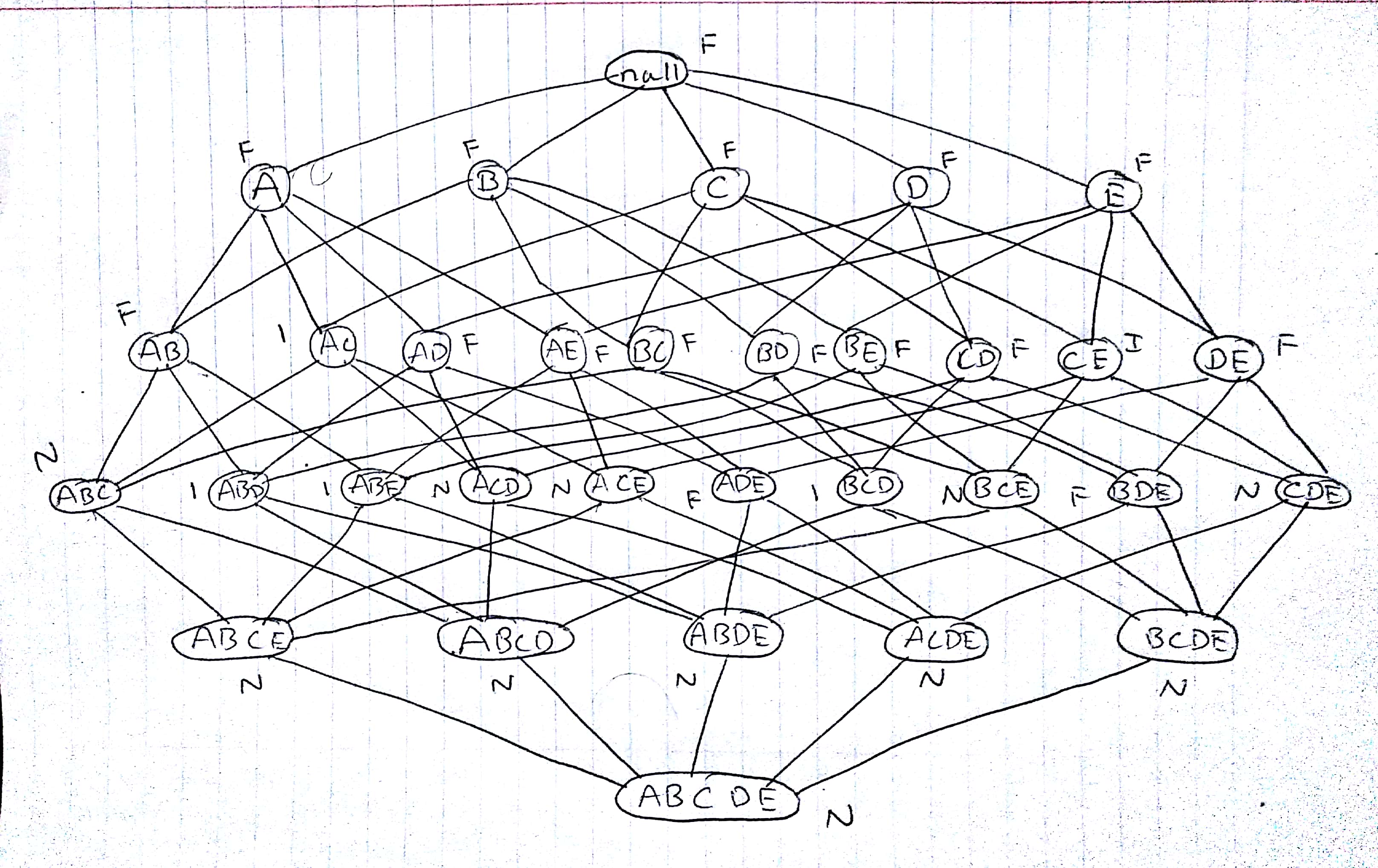
**infrequent.**

**• F: If the candidate itemset is found to be frequent by the A priori**

**algorithm.**

**• 1: If the candidate itemset is found to be infrequent. after support**

**Counting.**

****

**(b) What is the percentage of frequent itemsets (with respect to all itemsets**

**in the lattice)?**

The frequent itemsets are the number of itemsets which are not pruned and that has the support greater than 30%.

Total No.of frequent itemsets = 16

Total No.of itemsets = 32

Percentage of frequent itemsets = 16/32 => 0.5 or 50.0%

**(c) What is the pruning ratio of the Apriori algorithm on this data set?**

**(Pruning ratio is defined as the percentage of itemsets not considered**

**to be a candidate because ( 1) they are not generated during candidate**

**generation or (2) they rule pruned during the candidate pruning step.)**

Total No.of pruned itemsets = 11

Total No.of itemsets = 32

Percentage of frequent itemsets = 11/32 => 0.344 or 34.4%

**(d) What is the false alarm rate (i.e, percentage of candidate itemsets that**

**are found to be infrequent after performing support counting)?**

Total no.of infrequent itemsets = 5

Total no.of itemsets = 32

Percentage of frequent itemsets = 5/32 => 0.156 or 15.6%