Sample Questions

1. Briefly define the Bayes' rule and explain its importance. Write its equation and explain each term.
2. Define concept hierarchies and explain why they are very useful in data mining (Give at least three reasons).
3. Table 1 is the transaction database of a grocery store. The table contains five transactions and five items. Answer the following questions using the data given in Table 1.

Table 1

TID	Items
t1	Bread, Jelly, PeanutButter
t2	Bread, PeanutButter
t3	Bread, Milk, PeanutButter
t4	Beer, Bread
t5	Beer, Milk

   a. What is the total number of non-empty subsets of items in Table 1?
   b. The support of an item (or set of items) is the percentage of transactions in which that item (or items) occurs. Find the support for all subsets of items in Table 1.
   c. A frequent itemset is an itemset whose number of occurrences is above a threshold. Suppose that the minimum support and minimum confidence thresholds are **30%** and **50%**, respectively. Refer to data in Table 1 with associated supports you found in part **b** and list the frequent itemsets.
   d. Generate **all** association rules from the frequent itemsets that you listed in part **c** and determine the strong ones.
4. Explain FP-Growth algorithm.