b) Use your algorithm to construct a decision tree for the given data. **(6)**

c) Given a data tuple having the values "system", "26 ...30", and "46-50K" for the attributes department, age and salary, respectively, what would a Naive Bayaesian classification of the status for the tuple be ? **(5)**

15. a) Explain any four significant applications of data mining. **(16)**

(OR)

b) Describe any four approaches to clustering with suitable examples. **(16)**

18/11/17 AN

Reg. No. :

Question Paper Code : 50770

B.E./B.Tech. DEGREE EXAMINATION, NOVEMBER/DECEMBER 2017

Sixth/Seventh Semester

Information Technology

IT 6702 – DATA WAREHOUSING AND DATA MINING

(Common to : Computer Science and Engineering/Information Technology)

(Regulations 2013)

Time : Three Hours Maximum : 100 Marks

Answer ALL questions

PART – A **(10×2=20 Marks)**

1. What is data warehouse metadata ?
2. Differentiate between fact table and dimension table.
3. List the functions of OLAP servers in the data warehouse architecture.
4. What are the uses of multi-feature cubes ?
5. How concept hierarchies are useful in data mining ?
6. Mention the various tasks to be accomplished as part of data pre-processing.
7. Mention few approaches to mining Multilevel Association Rules.
8. Give the difference between Boolean Association rule and quantitative association rule.
9. What are Bayesian classifiers ?
10. What is the purpose of clustering in data mining ?

PART – B (5×16=80 Marks)

11. a) i) Sketch and explain the three-tier architecture of data warehousing. (8)

ii) Explain any two data transformation engine tools. (8)

(OR)

b) i) Discuss the star schema design with suitable illustrations. (10)

ii) Brief the three important steps in the design of a data warehouse. (6)

12. a) i) List and explain the five categories of decision support tools that operate on data warehouse. (10)

ii) Write about the three functional components of Fortè that provide comprehensive life-cycle support. (6)

(OR)

b) i) Give a critical comparison between OLTP and OLAP systems. (10)

ii) Illustrate any three OLAP operations on multidimensional data. (6)

13. a) Define each of the data mining functionalities : prediction, discrimination, association and evolution analysis. Give examples of each data mining functionality, using a real-life database of your choice. (16)

(OR)

b) Describe with suitable examples, the forms of data pre-processing : data cleaning, data integration, data transformation and data reduction. (16)

14. a) i) Illustrate the Apriori algorithm for finding frequent itemsets of any transaction database. (8)

ii) A database of table-14a has four transactions. Let min_sup = 60% and min_conf = 80%.

Table – 14a

Cust-id	T-id	Items-bought (brand-item_category)
01	T100	{King's-Crab, Sunset-Milk, Dairyland-Cheese, Best-Bread}
02	T200	{Best-Cheese, Dairyland-Milk, Goldenfarm-Apple, Tasty-Pie, Wonder-Bread}
01	T300	{Westcoast-Apple, Dairyland-Milk, Wonder-Bread, Tsty-Pie}
03	T400	{Wonder-Bread, Sunset-Milk, Dairyland-Cheese}

a) At the granularity of item_category (e.g. $item_i$ could be "Milk"), for the following rule template, (4)

$\forall X \in$ transaction, buys (X, $item_1$) $\wedge$ buys (X, $item_2$) $\Rightarrow$ buys (X, $item_3$) [s, c]

List the frequent k-itemset for the largest k, and all of the strong association rules (with their support s and confidence c) containing the frequent k-itemset for the largest k.

b) At the granularity of brand-item_category (e.g. $item_i$ could be "Sunset-Milk"), for the following rule template, (4)

$\forall X \in$ customer, buys(X, $item_1$) $\wedge$ buys (X, $item_2$) $\Rightarrow$ buys (X, $item_3$)

List the frequent k-termset for the largest k.

(OR)

b) The Table, Table-14b consists of training data from an employee database. The data have been generalized. For example, "31 ... 35" for age represents the age range of 31 to 35. For a given row entry, count represents the number of data tuples having the values for department, status, age, and salary given in that row.

Table-14b

Department	Status	Age	Salary	Count
Sales	Senior	31 ... 35	46K ... 50K	30
Sales	Junior	26 ... 30	26K ... 30K	40
Sales	Junior	31 ... 35	31K ... 35K	40
Systems	Junior	21 ... 25	46K ... 50K	20
Systems	Senior	31 ... 35	66K ... 70K	5
Systems	Junior	26 ... 30	46K ... 50K	3
Systems	Senior	41 ... 45	66K ... 70K	3
Marketing	Senior	36 ... 40	46K ... 50K	10
Marketing	Junior	31 ... 35	41K ... 45K	4
Secretary	Senior	46 ... 50	36K ... 40K	4
Secretary	Junior	26 ... 30	26K ... 30K	6

Let status be the class label attribute.

a) How would you modify the basic decision tree algorithm to take into consideration the count of each generalized data tuple (i.e. of each row entry) ? (5)