

Anekant Education Society's
Tuljaram Chaturchand College of Arts, Science and Commerce, Baramati
(Autonomous)
Department of Computer Science

Class : M.Sc.(Computer Science) – I

Semester : II

Paper Title : Data Mining and Data Warehousing (C)

Paper Code : COMP4202

Solve The Examples And Case Studies

1. The following data is collected about students in a class

No. Of Lectures Attended	Marks Obtained in Maths
40	80
10	40
38	75
20	55
25	65
35	60
15	50
08	30
36	70
28	65

Using straight line regression analysis predict how many marks a student, who has Attended 37 lectures will score?

2. Suppose that the data mining task is to cluster the following eight points (with (x;y) representing location) into three clusters.

A1(2;10), A2(2;5), A3(8;4), B1(5;8), B2(7;5), B3(6;4), C1(1;2), C2(4;9)

The distance function is Euclidean distance. Suppose initially we assign A1, B1 and C1 as the center of each cluster, respectively. Use the K-mean algorithm to show only. The Euclidean distance function between two points $a=(x_1,y_1)$ and $b=(x_2,y_2)$ is given as $\sqrt{(x_1-x_2)^2+(y_1-y_2)^2}$ or $(\text{SQRT}(x_1-x_2)^2+(y_1-y_2)^2)$

- A. The three cluster centers after the first round of execution and
B. The final three clusters.

3. Consider the following transaction table and generate the candidate itemsets and frequent Itemsets, where the minimum support count is 2.

TID	List of items
1	Potato, Onion, Tomato
2	Potato, Onion, Brinjal, Tomato
3	Onion, Tomato
4	Potato, Onion, Tomato
5	Potato, Brinjal
6	Onion, Tomato
7	Potato, Onion, Brinjal
8	Onion, Brinjal
9	Garlic
10	Potato, Brinjal

Apply Apriori algorithm to find the candidate Itemset and frequent item set.
For above transaction table Confidence given is 70%.
Find out all association rules of this transaction table.

4. Suppose that a data warehouse of Jewellery Shop consists of the Four dimensions Branch, Customer, Gold Ornaments and Silver Ornaments. It has the three measures GoldRate, SilverRate and Bill where GoldRate is the price of gold as per the market price and price of SilverRate is the price of silver as per the market price. Bill is the measure which customer pays to branch manager. Enumerate and draw both star schema and snowflake schema for the above case study.

5. Construct an FP tree for the following data. Minimum Support Count = 3
Also Find Out all Association Rules for the following transaction table where minimum confidence is 70%.

TID	List of items
1	Mango, Orange, Apple
2	Watermelon, Mango, Apple, Orange
3	Apple, Mango, Orange
4	Orange, Mango, Watermelon
5	Watermelon
6	Watermelon, Orange
7	Mango, Watermelon, Orange
8	Orange, Apple
9	Watermelon
10	Mango, Watermelon, Orange

6. Consider the database given below

Person Age	Profession	Salary	Loan Amount	Loan Sanctioned
34	Employee	40K	5 Lakh	Yes
38	Employee	41K	15 Lakh	No
20	Student	NA	2 lakh	No
45	Employee	30K	2 Lakh	Yes
67	Business	35K	5 Lakh	Yes
80	Retired	10K	20 Lakh	No
37	Employee	20K	2 Lakh	Yes
14	Student	NA	1 Lakh	No
12	Student	NA	5 Lakh	No
75	Business	55K	8 Lakh	Yes
55	Employee	25K	3 Lakh	Yes
60	Business	50K	40 Lakh	No

Find out class label of the given tuple using Navie Bayesian Classification.

<Person Age : 35 , Profession : Employee, Salary : 30K, Loan amount : 4 Lakh>

7. The following table consists of training data. Construct a decision tree based on this data using the basic algorithm of decision tree induction. Classify the records by “status” attribute.

Department	Status	Age	Salary	Count
Production	Junior	22	46K-50K	25
Production	Senior	35	66K-70K	10
HR	Junior	31	41K-45K	9
Finance	Senior	34	46K-50K	35
Finance	Junior	26	26K-30K	45
Production	Junior	27	46K-50K	8
Production	Senior	43	66K-70K	8
HR	Senior	38	46K-50K	15
Sales	Senior	50	36K-40K	9
Finance	Junior	33	31K-35K	45
Sales	Junior	26	26K-30K	11

8. How data mining is used in the field of Telecommunication Industry? Explain in detail. Answer should be minimum 12 to 15 lines.
9. How data mining is used in the field of Retail Industry? Explain in detail. Answer should be minimum 12 to 15 lines.
10. Can we make use of data mining in Financial Data Analysis. If your answer is yes then explain how it is useful and if your answer is no then explain how it is not useful? Answer should be minimum 12 to 15 lines.