(2 ½ Hours) [Total Marks: 75]

N.B. 1) All questions are compulsory.
2) Figures to the right indicate marks.
3) Illustrations, in-depth answers and diagrams will be appreciated.
4) Mixing of sub-questions is not allowed.

Q. 1 Attempt ANY FOUR from the following: (20M)

(a) Describe the framework of Data Warehouse with suitable diagram.
(b) How would you describe Dimensional Model. Also explain steps to design Dimensional Model.
(c) Compare OLTP V/S OLAP.
(d) Elaborate parts of Enterprise Data Warehouse (EDW) .
(e) Design a Star Schema for Data Warehouse of Bank Management System.
(f) When do you prefer Hybrid Data Warehouse model? Justify your answer.

Q. 2 Attempt ANY FOUR from the following: (20M)

(a) What kind of data can be mined form Data Mining?
(b) Describe various Data Mining metrics.
(c) What do you understand from Statistical Description of data.
(d) Why is data cleaning needed?
(e) For the given following data, apply Apriori algorithm. (Min. Support= 50%, Confidence threshold = 70%)

TID	Items
100	1 3 4
200	2 3 5
300	1 2 3 5
400	2 5

Dataset D

(f) How will you construct FP growth tree for the following transactions containing set of item = {a,b,c,d……} (min support=3)

T_id	Items bought
1	f,a,c,d,g,i,m,p
2	a,b,c,f,l,m,o
3	b,f,j,h,o
4	b,c,k,s,p
5	a,f,c,e,l,p,m,n

X548Y38A01CX548Y38A01CX548Y38A01CX548Y38A01C

Q. 3 Attempt ANY FOUR from the following: (20M)

(a) Differentiate between classification and prediction

(b) Explain linear regression with example.

(c) How can measure performance of classifier? Explain in brief.

(d) List the steps involved in k-means clustering algorithm.

(e) Apply ID3 on the following Training Dataset from the given database and extract the classification from the tree

Age	Income	Student	Credit_Rating	Class:buys_computer
<=30	High	No	Fair	No
<=30	High	No	Excellent	No
31...40	High	No	Fair	Yes
>40	Medium	No	Fair	Yes
>40	Low	Yes	Fair	Yes
>40	Low	Yes	Excellent	No
31....40	Low	Yes	Excellent	Yes
<=30	Medium	No	Fair	No
<=30	Low	Yes	Fair	Yes
>40	Medium	Yes	Fair	Yes
<=30	Medium	Yes	Excellent	Yes
31...40	Medium	No	Excellent	Yes
31...40	High	Yes	Fair	Yes
>40	Medium	No	Excellent	No

(f) What are some real-world applications where clustering algorithms play a significant role in different industries and domains?

Q. 4 Attempt ANY FIVE from the following: (15M)

(a) Elaborate features of Data warehouse.

(b) Define attribute and Enlist its types

(c) What are the requirements a clustering algorithm should satisfy?

(d) What can you say about granularity of facts?

(e) Explain the terms Support ,Confidence and Risk in Association Rule metrics

(f) State the applications of Classification.

****

X548Y38A01CX548Y38A01CX548Y38A01CX548Y38A01C