Total No. of Questions : 5]

SEAT No. : ____

P-5789

[Total No. of Pages : 3

[6120]-112

S.Y. M.C.A. (Management)

IT-32 : DATA WAREHOUSING AND DATAMINING

(2020 Pattern) (Semester - III)

Time : 2½ Hours] ***[Max. Marks : 50***

Instructions to the candidates:

1) ***All questions are compulsory.***
2) ***Draw neat labelled diagram wherever necessary.***

Q1) a) Describe different tools used for data warehouse development. **[5]**

b) Differentiate ER modelling Vs. Dimensional modelling. **[5]**

OR

c) What is OLAP? Explain different operations of OLAP. **[5]**

d) Discuss components of data warehouse architecture. **[5]**

Q2) a) What are the different cleaning tasks in ETL? **[5]**

b) Explain discretization & concept hierarchy generation steps in ETL. **[5]**

OR

c) Brief data integration & reduction methods in detail. **[5]**

d) Categorize ETL extraction methods. **[5]**

Q3) a) What is datamining? Discuss architecture of datamining with neat diagram. **[5]**

b) Define Text mining and discuss text mining process. **[5]**

OR

c) Write a short note on web context mining. **[5]**

d) Enlist the steps used is predictive modelling. **[5]**

P.T.O.

Q4) a) Consider following data set and find the frequent item sets with minimum support count 3 using FP. Tree algorithm. **[5]**

TID	Items
1	MONKEY
2	DONKEY
3	MAKE
4	COOKJE
5	CAKE
6	MUKCY

b) Consider the same data set in Q.4 a) and calculate support & confidence of following rules. **[5]**

i) $\{M, O\} \rightarrow \{Y\}$

ii) $\{K, E\} \rightarrow \{O, Y\}$

iii) $\{K, E\} \rightarrow \{M, O\}$

iv) $\{M\} \rightarrow \{K, E, Y\}$

v) $\{D\} \rightarrow \{O, N\}$

OR

c) Apply NB classifier on below dataset for the given instance. **[5]**

S.No.	Weather Condition	Vehical Condition	Traffic Problem	Accident
1	Rain	bad	high	yes
2	snow	average	normal	yes
3	clear	bad	light	no
4	clear	good	light	yes
5	snow	good	normal	no
6	rain	average	light	no
7	rain	good	normal	no
8	snow	bad	high	yes
9	clear	good	high	no
10	clear	bad	high	yes

Instance =

[Weather condition = “clear”, vehical condition = “bad”,

Traffic condition = “light”, Accident = ?]

d) Write the algorithm for decision tree. Generate a Decision tree for classification of loan approval or rejection. Consider the below attributes. 1) age, 2) Income group, 3) CIBIL Score **[5]**

Q5) a) Divide the data into high & low income group using k-mean clustering. **[5]**

D = (20k, 25k, 22k, 23k, 30k, 35k, 65k, 80k, 70k, 90k, 100k, 92k, 94k, 96k, 78k, 60k, 65k, 35k, 25k, 32k)

b) Write a note on Hierarchical clustering. **[5]**

OR

c) Given the dataset of age of people. Form K = 2. Clusters. **[5]**

D = (20, 25, 15, 35, 42, 41, 30, 56, 61, 62, 75, 80, 72, 75, 85, 55, 45, 43, 35, 78)

d) Explain metadata collection strategies. **[5]**