DHARMSINH DESAI UNIVERSITY, NADIAD
FACULTY OF TECHNOLOGY
B.TECH. SEMESTER VII [I.T]
SUBJECT: (IT-704) Data Analysis & Information Extraction

Examination	**: External (Written Exam)**	**Seat No.**	**: _____**
Date	**: 25/11/2021**	**Day**	**: Thursday**
Duration	**: 90 minutes**	**Max. Marks**	**: 30 Time: 11:30 to 1:00 PM**

INSTRUCTIONS:
1. Figures to the right indicate maximum marks for that question.
2. The symbols used carry their usual meanings.
3. Assume suitable data, if required & mention them clearly.
4. Draw neat sketches wherever necessary.

Section –II

Q.2 Attempt ***Any Three*** from the following questions. **[15]**

(a) What are the import issues do we have to consider for design of any clustering Algorithm? **[5]**

(b) Explain Hierarchical base clustering using suitable examples. **[5]**

(c) A database has four transactions.Let *min_sup*=2 & *min_conf* = 80%. **[5]**

TID	***Date***	***Items_bought***
T100	01/09/2003	{K, A, D, B}
T200	01/09/2003	{D, A, C, E, B}
T300	03/09/2003	{C, A, B, E}
T400	08/09/2003	{B, A, D}

List all frequent itemsets using Apriori and list all strong association rules for given type of patterns. E.g X ^ Y => Z.

(d) Suppose that we would like to mine the general characteristics describing graduate students at Big-university using analytical characterization. Given are the attributes: **[5]**

Target class		Contrasting class	
Major	Count	Major	Count
Science	18	Science	16
Science	22	Business	20
Engineering	17	Business	22
Science	25	Science	24
Science	21	Engineering	22
Engineering	16	Engineering	24

Perform the analytical characterization (Find the information gain for the attribute major.).

(e) Training data tuples of XYZ Cos. customer database are as follows: **[5]**

RID	*age*	*income*	*student*	*credit_rating*	*Class: buys_computer*
1	<=30	high	no	fair	no
2	<=30	high	no	excellent	no
3	31…40	high	no	fair	yes
4	>40	medium	no	fair	yes
5	>40	low	yes	fair	yes
6	>40	low	yes	excellent	no
7	31…40	low	yes	excellent	yes
8	<=30	medium	no	fair	no
9	<=30	low	yes	fair	yes
10	>40	medium	yes	fair	yes
11	<=30	medium	yes	excellent	yes
12	31…40	medium	no	excellent	yes
13	31…40	high	yes	fair	yes
14	>40	medium	no	excellent	yes

Using Naive Bayesian Classifier to predict the class label of the following sample:

X = (age = "<=30", income = "medium", student = "yes", credit_rating = "fair").

Q.3 (a) Explain three Tier data warehouse architecture with proper figure. **[4]**

(b) Define K-mean algorithm with proper example. What are the limitation? **[4]**

(c) For Airport Management System Company wants to do data analysis on Passagers.

1. Identify dimensions and fact table. **[2]**
2. Draw star schema diagram with proper measures. **[2]**
3. Design Data CUBE and show roll-up & pivot operation on it. **[3]**

OR

Q.3 (a) Explain KDD process with proper diagram. **[4]**

(b) How does decision tree classification work? Explain with on Example. **[4]**

(c) For E-Retailer like Jio-Mart Company wants to do data analysis.

1. Identify dimensions and fact table. **[2]**
2. Draw star schema diagram with proper measures. **[2]**
3. Design Data CUBE and show roll-up & pivot operation on it. **[3]**