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

Examination	**: Third Sessional (Written Exam)**	**Seat No.**	**: _____**
Date	**: 14/10/2021**	**Day**	**: Thursday**
Duration	**: 45 minutes**	**Max. Marks**	**: 16**

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.

Q.2 Attempt ***Any Two*** from the following questions. **[8]**

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

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

(c) Give an example of hierarchical clustering using dendrogram? **[4]**

Q.3 Training data tuples of XYZ Cos. customer database are as follows:

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	no

a) Using Naive Bayesian Classifier to predict the class label of the following sample: **[6]**

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

b) Give the limitation of algorithm & solution. **[2]**

OR

Q.3 Suppose that we would like to mine the general characteristics describing graduate students at Big-university using analytical characterization. Given are the attributes:

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

a) Perform the analytical characterization (Find the information gain for the attribute major.). Assume suitable data. (Threshold Value = 0.1) **[7]**

b) Give uses of Information Gain. **[1]**