DEDAN KIMATHI UNIVERSITY OF TECHNOLOGY

UNIVERSITY EXAMINATIONS 2021/2022

M.Sc. (BUSINESS ANALYTICS) YEAR ONE SEMESTER TWO

CIT 6113: BUSINESS DATA MINING

DATE: **TIME:**

Instructions: Answer Question 1 and Any Other Two.

Question 1: (30 Marks)

a) By giving appropriate examples describe TWO different types of attributes one may come across in data mining sets **[4 Marks]**

b) Describe the FIVE different visualization techniques that can be used in data mining **[5 Marks]**

c) Data pre-processing involves more that 80% of any data mining task. Discuss the different steps involved in data pre-processing **[5 Marks]**

d) Missing values cannot be looked over in a data set. They must be handled. Also, a many data mining models do not accept missing values. Discuss FIVE techniques used to handle missing data **[5 Marks]**

e) Discuss different distance measurers that can be used to compute distance between two clusters **[6 Marks]**

f) Using appropriate diagram discuss Data Mining as a step in Knowledge Discovery in Databases **[5 Marks]**

Question 2: (15 Marks)

Consider the following transaction database:

Transaction	List of items
T1	I1,I2,I3
T2	I2,I3,I4
T3	I4,I5

Transaction	List of items
T4	I1,I2,I4
T5	I1,I2,I3,I5
T6	I1,I2,I3,I4

Apply the Apriori algorithm with **Support threshold=50% and Confidence= 60%** and find all the association rules in the dataset **[20 Marks]**

Question 3: (15 Marks)

Consider an example dataset of the last 10 days weather data with the following attributes; outlook, temperature, wind, and humidity. The outcome variable will be playing cricket or not.

Day	Outlook	Temperature	Humidity	Wind	Play cricket
1	Sunny	Hot	High	Weak	No
2	Sunny	Hot	High	Strong	No
3	Overcast	Hot	High	Weak	Yes
4	Rain	Mild	High	Weak	Yes
5	Rain	Cool	Normal	Weak	Yes
6	Rain	Cool	Normal	Strong	No
7	Overcast	Cool	Normal	Strong	Yes
8	Sunny	Mild	High	Weak	No
9	Sunny	Cool	Normal	Weak	Yes
10	Rain	Mild	Normal	Weak	Yes

a) Describe the steps of creating a decision tree **[5 Marks]**

b) Construct a decision tree to demonstrate how to arrive at the decision of either to play cricket or not **[10 Marks]**

Question 4: (15 Marks)

a) *Naive Bayes Classifiers* are a family of simple probabilistic classifiers based on applying *Bayes' theorem* with strong independence assumptions between the features. Describe the theorem **[5 Marks]**

b) Consider the following
   - the probability of dangerous fires are rare (1%)
   - but smoke is fairly common (10%) due to barbecues
   - 90% of dangerous fires make smoke

   Using Bayes theorem, find the probability of dangerous Fire when there is Smoke **[5 Marks]**

c) Discuss the difference between Supervised Learning and UnSupervised Learning as used in Data Mining **[2 Marks]**

d) Discuss the steps for performing classification tasks in Data Mining **[8 Marks]**

END.