

**Mawlana Bhashani Science and Technology University**  
**Department of Information and Communication Technology**  
**3rd Year 1st Semester B.Sc. (Engg.) Final Examination 2023**

Course Title: Data Mining  
Course Code: ICT 3101

Marks: 70  
Time: 3 hours

Answer any 05 (FIVE) questions

1. a) What is data type in Data Mining? What types of data are found in Data Mining? Name those data types. 3
- b) How do you collect and clean data? Is there any specific way to accomplish this work? 5
- c) What are the techniques to fill-in the missing values? 6
  
2. a) It is known that learning is divided into two major classes: Supervised learning and Unsupervised learning. How these learning types influence data mining works? 4
- b) What are the Data Mining programming languages used in the current time? What are the common data repositories available? 4
- c) Write about Information Gain and Gini Index. 6
  
3. a) Why are test and training data separated? What is the benefit of doing so? 5
- b) Why do we need cross validation in Data Mining operations? 3
- c) "Leave-one-out cross-validation is a good way to validate any data" – Do you agree or disagree? Why? 6
  
4. a) Write a note on incremental Association Rule Mining. 3
- b) Explain Mining Multiple-Level Association Rules using example. 4
- c) State the Apriori Property. Generate large itemsets and association rules using Apriori algorithm on the following data set with minimum support value and minimum confidence value set as 50% and 75% respectively: 7

TID	Items Purchased
T 101	Cheese, Milk, Cookies
T 102	Butter, Milk, Bread
T 103	Cheese, Butter, Milk, Bread
T 104	Butter, Bread

5. a) Compare association and classification. 2  
    Briefly explain associative classification with suitable example. 3
- b) What is an attribute? How is the Attribute-Relation File Format file constructed? 3
- c) How is the IF – THEN logic achieved? Provide an example. 6

Table – 01

Outlook	Temperature	Humidity	Windy	Play
Sunny	85	85	FALSE	No
Sunny	80	90	TRUE	No
Overcast	83	86	FALSE	Yes
Rainy	70	96	FALSE	Yes
Rainy	68	80	FALSE	Yes
Rainy	65	70	TRUE	No
Overcast	64	65	TRUE	Yes
Sunny	72	95	FALSE	No
Sunny	69	70	FALSE	Yes
Rainy	75	80	FALSE	Yes
Sunny	75	70	TRUE	Yes
Overcast	72	90	TRUE	Yes
Overcast	81	75	FALSE	Yes
Rainy	71	91	TRUE	No

Question below may be referred to the Table – 01.

6. a) What are the different item sets in association learning? 3  
b) Refer to the table – 01: How many one-item and four-item sets can be prepared? 8  
c) Why is the decision set important in data mining? Give one example. 3
  
7. a) What is Decision Tree? Explain how classification is done using decision tree induction. 7  
b) Generate decision tree using CART algorithm using the table – 01. 7
  
8. a) How is clustering accomplished? 2  
b) How many clusters are possible for the table – 01? 6  
c) How many nodes are possible for the table – 01, if you consider a decision tree? 6