Institute of Information Technology

Jahangirnagar University

Professional Masters in IT

1st Semester Final Examination, Fall 2022

Duration: 3 Hours | **Intake: Fall 2022, Summer 2022**

Full Marks: 60

Course Code: PMIT – 6307 | **Course Title: Data Mining and Knowledge Discovery**

Do not write anything on the question paper.
There are **7 (Seven)** questions. Answer any **5 (Five)** of them.
Figures in the right margin indicate marks.
[Follow the question order]

---

1. a) Define the term "Data Mining" and give an example where human always use data mining techniques. 4
   b) What arse the difference between classification and clustering? 4
   c) In the real life, where do you use association rule? 4

2. a) If you have number data i.e. income, you have to standardize them. Write the process of it. 4
   b) For ordinal and nominal data, how do you calculate dissimilarity for a single attribute? 4
   c) In what aspect, you may like decision tree classification? 4

3. a) How a tree-based classification works? What are the advantages of tree-based classification? 4
   b) Suppose you have two variables (Gender and Marital status) to select one in constructing a tree. 4

Gender	Marital	Class
M	M	Good
M	U	Bad
M	M	Good
F	U	Bad
F	U	Bad

Gender	Marital	Class
M	U	Good
M	U	Bad
F	M	Good
M	M	Good
F	U	Bad

   Which variable you should select? Use Gini coefficient or entropy to give your answer.
   c) Define the terms: True positive and False Negative and F-measure 4

4. a) When do we use learning curve and ROC curve? 4
   b) Explain the Ensemble Method of classification. 4
   c) What is OLAP? Why do researcher use Slicing and Dicing in data analysis? 4

5. a) What are the steps of KNN classification? 4

b) Why do the most researchers prefer SVM classification? 4

c) If you have a nominal attribute which has 3 categories and for binary tree classification you have to merge any two categories. Which two will you merge? 4

6. a) Define hierarchical clustering and give an example of this clustering where it is the most useful technique. 4

b) What are the different types of clustering? 4

c) Write some limitations of k-means clustering. How can you overcome these limitations? 4

7. a) What is the basic principle of DBSCAN clustering? Write the usefulness of this clustering. 4

b) Explain with a pictorial example of Core Point, Noise Point and Border Point. 4

c) The distance of k-th neighbor of data points are almost equal" – explain this comment. 4