



Institute of Information Technology
Jahangirnagar University
Professional Masters in IT

1st Semester Final Examination, Spring 2023

Duration: 3 Hours

Intake: Spring 2023, Fall 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 we frequently use data mining techniques. 4
b) Write an example of descriptive data mining. Explain why it is different from predictive data mining. 4
c) What are the sequential steps in doing classification of a set of data? 4
2. a) What do you mean by discretization of data? If you have occupation data like Job, Business, unemployed etc. How will you make it binary? 4
b) What do you mean by similarity of two data point? 4

Name	Age	Income
Asad	45	10,000
Karim	43	9,997
Rahim	44	9,995

Who is more similar with Asad?

- c) In case of ordinal and nominal data what is the calculation of similarity? 4
3. a) How a tree-based classification works? What are the advantages of tree-based classification? 4
b) Suppose you have two variables (occupation and gender to select one in constructing a tree. 4

Occupation	Gender	Cheat
Job	Male	Yes
Business	Female	No
----	----	----
Business	Male	No

Data from the above table are summarized as follows:

Occupation		Gender	
Job	Business	Male	Female
Cheat =Yes: 5 Cheat= No: 15	Cheat =Yes: 3 Cheat= No: 10	Cheat =Yes: 3 Cheat= No: 7	Cheat =Yes: 6 Cheat= No: 8

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) Data may be affected by various kind of reasons. These reasons we may define as data quality problems. 4
- b) Why EDA is necessary for data mining? What is the main advantage of using learning curve? 4
- c) What is OLAP? Why do researcher use Slicing and Dicing in data analysis? 4
5. a) If you have a data set and you have to perform KNN classification, what are the steps you will follow? 4
- b) Why do most researchers prefers SVM classification? 4
- c) Among Tree-based classification and Rule-based classification, which one will you prefer and why? 4
6. a) Define hierarchical clustering and give an example of this clustering where it is the most useful technique. 4
- b) Write the steps of k-means 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) In DBSCAN clustering, how will you decide EPS and max points? 4