



Institute of Information Technology

Jahangirnagar University

Professional Masters in IT

1st Semester Final Examination- Spring 2022

Intake: Spring 2022, Fall 2021

Duration: 3 Hours

Full Marks: 60

Course Code: PMIT – 6307

Course Title: Data Mining & Knowledge
Discovery

Do not write anything on the question paper.

There are 7 (Seven) questions. Answer any 5 (Five) of them

-
1. a) Define Data Mining? What are the difference between descriptive data mining and predictive data mining? 4
b) A marketing officer of a Laptop Sales House is trying inspire customers to buy the special brand of laptop. He will have to use different data mining techniques in different phases. Explain it from your imagination. 4
c) What is “Association Rule Discovery”? give an appropriate example. 4
 2. a) There are a lot of challenges in doing data mining. One of them is “Scalability” – explain it clearly. 4
b) Explain different reasons for which data may be affected. 4
c) In nature, data is heterogeneous. In data mining, **nominal data** and **ordinal data** comes along with **ratio data**. Explain the bold terms with example. 4
 3. a) Why do we use data aggregation and dimension reduction? 4
b) Give an example of stratified sampling and write its purpose of use in data mining. 4
c) When do we need discretization and binarization of data? 4
 4. a) Define the term EDA. Why EDA is necessary for data mining? 4
b) Suppose, Asad, Kabir and Raihan bought groceries from shop. Calculate their similarities and evaluate who are very closest 4
Asad 5 kg rice, 1 kg fish, 200 gm chilly and 1 kg milk
Raihan 1 kg rice, 1 kg Meat, 200 gm chilly and 1 kg milk
Kabir 2 kg rice, 1 kg fish, 200 gm salt and 1 kg milk
c) What is the main principal of Gini index? – explain. When we have to use Gini index in splitting? 4
 5. a) Write the procedures of tree based classification? How does one can check the validity of a tree? 4
b) When do one can stop constructing classification tree? 4
c) Suppose you have two variables (A and B) to select one in constructing a tree. Under variable A distribution of class variable are 5 Cheat and 10 No Cheat in one hand and in other hand it is 3, 12 whereas under variable B it is 3, 7 in one hand and 12, 8 in other hand respectively. Which variable you should select? Use Gini coefficient or entropy to give your answer. 4

- | | | | |
|----|----|---|---|
| 6. | a) | What do you mean by clustering? What are the different types of clusters? | 4 |
| | b) | Shortly explain process of hierarchical clustering. | 4 |
| | c) | Write the steps of K-means clustering in your own words sequentially. | 4 |
| 7. | a) | How a KNN classification is performed? What are the steps of k-means clustering? | 4 |
| | b) | In density based clustering, how do you decide minPoints and Eps ? | 4 |
| | c) | Write two limitations of k-means clustering. How can we minimize these limitations? | 4 |