1st ~~2nd~~ Trimester Final Examination, Summer 2022

Duration: 3 Hours

Intake: Summer 2022, Spring 2022

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.
Figures in the right margin indicate marks.

---

1. a) Define the term "Data Mining" and give an example where human always use data mining techniques. — 4
   b) What are the difference between classification and clustering? — 4
   c) What are the sequential steps in doing classification of a set of data? — 4

2. a) List different types of attributes with their general properties — 4
   b) For different types of data how do you calculate dissimilarity for a single attribute? — 4
   c) What is 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 (A and B) to select one in constructing a tree. Under variable A distribution of class variable are 5 Cheat and 15 No Cheat in one hand and in other hand it is 3, 10 whereas under variable B it is 3, 7 in one hand and 6 , 8 in other hand respectively. Which variable you should select? Use Gini coefficient or entropy to give your answer. — 4
   c) Define the terms: True positive and False Negative and F-measure — 4

4. a) Data may affected by various kind of reasons. These reasons we may define as data quality problems. — 4
   b) Why EDA is necessary for data mining? Why do we use cross validation in 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 most researchers do 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) 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 principal 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