Reg. No. | | | | | | | | | |

STS 557

Fourth Semester M.Sc. Degree Examination, September/October 2022
STATISTICS
Data Mining Techniques

Time : 3 Hours Max. Marks : 70

Note : *Question number **1** is **compulsory**.*
*Answer **any four** questions from the remaining seven questions.*

1. Answer **any six** questions. **(6×3=18)**
   a) Discuss any two applications of data mining in industry.
   b) Which are the tools used for data integration in data mining ?
   c) Discuss various kind of data used in data mining.
   d) Differentiate between Online Transaction Processing (OLTP) and Online Analytical Processing (OLAP).
   e) Explain CART approach of construction of decision trees.
   f) What is the difference between supervised learning and unsupervised learning ?
   g) Mention any three "similarity and distance measures" and its characteristics.
   h) Explain with examples crossover and mutation with reference to genetic algorithm.

2. a) Describe the steps involved in data mining when viewed as a process of knowledge discovery.
   b) Explain the different schemes of multidimensional data modelling with examples. **(7+6)**

3. a) Taking an example of applications of artificial intelligence in data mining, describe the method and use of the same.
   b) Explain with the help of an example, the different kinds of OLAP operations performed in a data cube. **(7+6)**

P.T.O.

4. a) Explain data integration and data transformation of data pre-processing methods.

   b) Explain the steps involved in Andrews plots and explain Chernoff faces. Write down the difference among them. **(7+6)**

5. a) Explain the principle of ID3 algorithm and using an application describe the method.

   b) Using an example describe decision tree problems. How the classifications are carried out ? **(6+7)**

6. a) Explain in detail the regression based classification.

   b) Explain the working theory of k-means clustering method. **(7+6)**

7. a) Discuss partitioning algorithm and its applications.

   b) How Does DBSCAN quantify the neighbourhood of an object and how clusters are formed ? **(7+6)**

8. a) Explain how Bootstrap method is used to generate an empirical estimate of the sampling distribution of an estimate.

   b) Explain Gibbs sampler. Further explain how it is used in Markov Chain Monte Carlo. **(6+7)**
