3 (Sem–6/CBCS) CSC HE 4

2022

COMPUTER SCIENCE

(Honours Elective)

Paper : CSC–HE–6046

(Data Mining)

Full Marks : 60

Time : Three hours

The figures in the margin indicate full marks for the questions.

GROUP–A

1. Answer ***any seven*** of the following questions : 1×7=7

   (a) What is data mining ?

   (b) Define metadata.

   (c) What do you mean by optional data ?

   (d) Give *any two* applications of data mining.

(e) Define frequent set.

(f) What is meant by pattern ?

(g) What is star schema ?

(h) What is tree pruning ?

(i) What is partitioning ?

(j) What is a data cube ?

2. Answer ***any four*** of the following questions : 2×4=8

(a) What is data prediction ?

(b) Expand the following terms :
BIRCH, OPTICS

(c) What are technologies used in data mining ?

(d) List *any two* data mining tasks.

(e) What are the advantages of data mining ?

(f) What are the ways by which data mining algorithm are characterized ?

(g) What do you mean CART ?

(h) Write the use genetic algorithm.

3. Answer ***any three*** of the following questions : 5×3=15

(a) What is association analysis ?

(b) Distinguish between OLTP and OLAP.

(c) Write algorithm for K-nearest neighbor classification.

(d) What are the features of cluster analysis ?

(e) Write a computer program to implement the BIRCH algorithm.

(f) Write a computer program to implement the DIC (Dynamic Itemset Counting) algorithm.

(g) Discuss data visualization with reference to data mining.

(h) Illustrate the use of ID3 algorithm with an example.

GROUP–B

4. Answer ***any three*** of the following questions : 10×3=30

(a) Explain and draw the architecture of a typical data mining system.

(b) Develop the Apriori algorithm for generating frequent item set.

(c) Discuss the tasks of data mining with suitable examples.

(d) Explain shortly *any five* data pre-processing approaches.

(e) Explain the cluster analysis methods briefly.

(f) Explain with example the two approaches for extending the binary classifiers to handle multiclass problem.

(g) Discuss the decision tree induction with an algorithm.

(h) Write short notes on the following : ***(any two)***

(i) Border algorithm

(ii) Text timing

(iii) Spatial database mining

(iv) Hierarchical clustering

(v) Data visualization