Sixth Semester B.Sc. Degree Examination, April 2022

Career Related First Degree Programme under CBCSS

Group 2(b) — Computer Science

Core Course

CS 1641 — DATA MINING AND WAREHOUSING

(2018 & 2019 Admission)

Time : 3 Hours

Max. Marks : 80

SECTION – A

(Very Short Answer Type)

(**One** word to maximum of **one** sentence. Answer **all** questions)

1. What do you mean by data mining?
2. What does OLTP stand for?
3. What is a spatial database?
4. Define confidence.
5. What is a multimedia database?
6. What is a frequent itemset? Give an example.
7. What is an outlier?
8. What is the purpose of data cleaning?
9. What is a dimension table?
10. What is a categorical variable? Give example.

(10 × 1 = 10 Marks)

P.T.O.

SECTION – B

(Short Answer)

Not to exceed **one** paragraph, answer any **eight** questions. Each question carries **2** marks.

11. What is background knowledge? Give an example.
12. List the four categories of data preprocessing.
13. What is numerosity reduction?
14. What are the differences between a data warehouse and an operational database?
15. What is ROLAP?
16. What is technical metadata in a data warehouse?
17. Differentiate between characterization and discrimination.
18. What do you mean by generalization?
19. Which are the two methods for dimensionality reduction?
20. What are the business skills needed in building a data warehouse?
21. What is market basket analysis?
22. What are hybrid association rules? Give example.
23. Mention any four methods for classification.
24. What is lift?
25. What do you mean by accuracy of a rule?
26. What is an asymmetric binary variable? Give example.

(8 × 2 = 16 Marks)

SECTION – C

(Short Essay)

Not to exceed **120** words, answer any **six** questions. Each question carries **4** marks.

27. Differentiate between OLTP versus OLAP.
28. Explain prediction methods in data mining.
29. Explain data warehouse with a diagram.

30. How will you handle missing data?

31. Explain decision tree and its uses.

32. Draw the architecture of a typical data mining system.

33. What is cluster analysis? Mention different methods for clustering.

34. Explain various data mining classification systems.

35. Explain concept hierarchy with examples.

36. Differentiate between supervised learning and unsupervised learning.

37. Mention the conditions for stopping partitioning in Decision Tree Induction.

38. Mention different types of association rules and give examples for each rule.

(6 × 4 = 24 Marks)

SECTION – D

(Long Essay)

(Answer any **two** questions. **Each** question carries **15** marks)

39. Explain various data mining functionalities.

40. Explain the requirements of clustering in data mining.

41. Explain with diagram, various schema for multidimensional data model.

42. With neat diagrams, explain typical OLAP operations.

43. Explain apriori algorithm with example.

44. Explain partitioning methods for clustering.

(2 × 15 = 30 Marks)