

Seventh Semester B.E. Degree Examination, Feb./Mar. 2022
Data Warehousing and Data Mining

Time: 3 hrs.

Max. Marks:100

Note: Answer any FIVE full questions, selecting at least TWO full questions from each part.

PART – A

- 1 a. Discuss the need for datawarehousing. With example, explain two schemes used to design datawarehouse. (10 Marks)
- b. Explain the architecture of ODS, with diagram. (05 Marks)
- c. List the major steps involved in ETL process. (05 Marks)

- 2 a. Explain different types of data cube operations with example using appropriate diagram. (10 Marks)
- b. Explain any 5 rules from Codd’s OLAP characteristics. (05 Marks)
- c. Explain the difference between MOLAP and ROLAP. (05 Marks)

- 3 a. Why it is necessary to preprocess the data? Describe various techniques used to preprocess the data. (10 Marks)
- b. Calculate SMC, Jaccard coefficients, cosine similarity, Euclidean and Manhattan distance between, below given two vectors X and Y.
 $X = (1, 1, 0, 1, 0, 1)$; $Y = (1, 1, 1, 1, 0, 1)$. (10 Marks)

- 4 a. Consider the following transaction database, as shown in Table 4.1.

Table 4.1

TID	Items-bought
T ₁₀₀	{M, O, N, K, E, Y}
T ₂₀₀	{D, O, N, K, E, Y}
T ₃₀₀	{M, A, K, E}
T ₄₀₀	{M, U, C, K, Y}
T ₅₀₀	{C, O, O, K, I, E}

- Let min-sup = 60% and min-conf = 80%. Find all frequency itemsets using FP – growth algorithm. Also generate all the strong rules from the frequent itemsets by assuming the minimum support of 60% and minimum confidence of 80%. (12 Marks)
- b. What is Apriori principle? Explain frequent itemset generation using Apriori algorithm. (08 Marks)

PART – B

- 5 a. How decision trees are used for classification? Explain decision tree induction algorithm for classification. (10 Marks)
- b. Explain nearest neighbor classifier and its characteristics. (10 Marks)

- 6 a. Explain the general approach to solve classification problem. Explain how the predictive accuracy of classification methods be estimated. (10 Marks)
- b. What are Bayesian classifiers? Explain Naïve Bayes classifier? Mention its characteristics. (10 Marks)

Important Note : 1. On completing your answers, compulsorily draw diagonal cross lines on the remaining blank pages.
 2. Any revealing of identification, appeal to evaluator and /or equations written eg, 42+8 = 50, will be treated as malpractice.

- 7 a. Explain the desired features of cluster analysis. (08 Marks)
b. What are the various types of cluster analysis methods? (05 Marks)
c. Briefly describe hierarchical methods to clustering by giving examples whenever necessary. (07 Marks)
- 8 a. With webmining taxonomy, explain various webmining tasks. (08 Marks)
b. Write a note on:
i) Text mining
ii) Temporal data mining
iii) Spatial data mining. (12 Marks)

* * * * *