School of Computing Science and Engineering

Bachelor of Science in Computer Science
Semester End Examination - Aug 2024

Duration : 180 Minutes
Max Marks : 100

Sem V - E1UJ501B - Data Warehousing and Data Mining

General Instructions
Answer to the specific question asked
Draw neat, labelled diagrams wherever necessary
Approved data hand books are allowed subject to verification by the Invigilator

1)	Define Data warehouse metadata.	K1(2)
2)	List the key steps involved in the classification process.	K2(4)
3)	List some basic concepts associated with classification algorithms.	K2(6)
4)	Apply a decision tree algorithm to a dataset and explain the resulting model.	K3(9)
5)	Compare and contrast the Apriori algorithm and the FP-Growth algorithm for association rule mining.	K3(9)
6)	Explain about the Numerosity Reduction and Concept hierarchy generation	K5(10)
7)	Examine the factors that influence the choice between rule-based and neural network-based classification approaches	K4(12)
8)	Suppose that the data for analysis includes the attribute age. The age values for the data tuples are (in increasing order) 13, 15, 16, 16, 19, 20, 20, 21, 22, 22, 25, 25,25, 25, 30, 33, 33, 35, 35, 35, 35, 36, 40, 45, 46, 52,70. Compute the following: (a) Mean of the data? Median? (b) mode of the data? Comment on the data's modality(i.e.,bimodal,trimodal,etc.).	K5(15)
9)	Examine the following consider the following data for analysis includes the attribute age. The age values for the data tuples are (in increasing order) 13, 15, 16, 16, 19, 20, 20, 21, 22, 22, 25, 25, 25, 25, 30, 33, 33, 35, 35, 35, 35, 36, 40, 45, 46, 52, 70. (a) Usemin-maxnormalizationtotransformthevalue 35 for age on to the range [0.0, 1.0]. (b) Use z-score normalization to transform the value 35 for age, where the standard deviation of age is 12.94 years. (c) Use normalization by decimal scaling to transform the value 35 for age. (d) Comment on which method you would prefer to use for the given data, giving reasons as to why.	K5(15)
10)	Construct a data mining solution for a transportation company that optimizes routes and schedules based on real-time traffic data, weather conditions, and customer demand.	K6(18)