A

Pages: 3

04000CS402052201

Reg No.:_____ Name:_____

APJ ABDUL KALAM TECHNOLOGICAL UNIVERSITY

Eighth Semester B.Tech Degree Supplementary Examination October 2022 (2015 Scheme)

Course Code: CS402

Course Name: DATA MINING AND WAREHOUSING

Max. Marks: 100 Duration: 3 Hours

PART A

Answer all questions, each carries 4 marks. Marks

1	List out any four applications of datamining in day-to-day life	(4)
2	Given unlimited storage, can an OLAP system be replaced by an OLTP system? Justify your answer.	(4)
3	Explain the process of automatic concept hierarchy generation with an example.	(4)
4	Describe any two methods for attribute subset selection	(4)
5	Explain any four well known issues regarding classification and prediction.	(4)
6	Define support vectors. What is the significance of maximum marginal hyperplane in SVM classification?	(4)
7	Describe any four methods to improve the efficiency of apriori algorithm.	(4)
8	Define support and confidence. With the help of an example, show how these values are calculated.	(4)
9	Write any two differences between Agglomerative and Divisive Hierarchical Clustering. Give examples for each method.	(4)
10	Define clustering feature (CF) in the context of BIRCH algorithm. Explain how each attribute of CF is calculated.	(4)

PART B

Answer any two full questions, each carries 9 marks.

11	a)	Draw a fact constellation schema for a data warehouse that stores the details of a Department Library of a college. There are at least two fact tables. Each fact tables should have minimum two measures and four dimension tables connected to it. Some dimension tables are shared by both fact tables.	(5)
	b)	Describe any four differences between a traditional database system and a data warehouse	(4)

12 a) What is the significance of data pre-processing? Explain any four data pre-processing methods. (5)

b) A set of data is given:{235,453,675,964}. Normalize the data by Min-max normalization (range: [0.0,1.0]). (4)

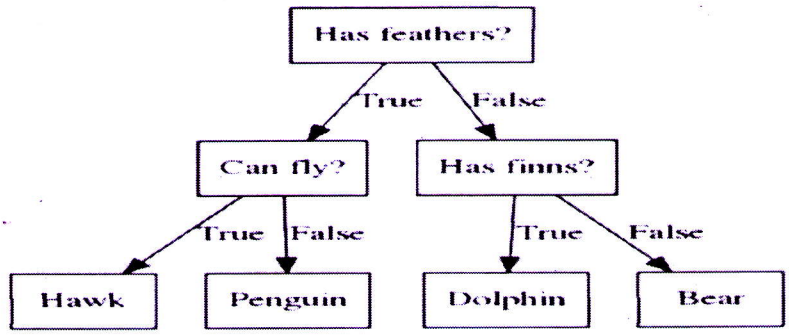
13 a) Explain the importance of data reduction in data warehousing. Describe the techniques used for dimensionality reduction and numerosity reduction. (5)

b) Show how typical OLAP operations are performed on a sample data cube. (4)

PART C

Answer any two full questions, each carries 9 marks.

14 Give an algorithm for extracting rules from a decision tree. Demonstrate the working of the algorithm in the following decision tree (9)

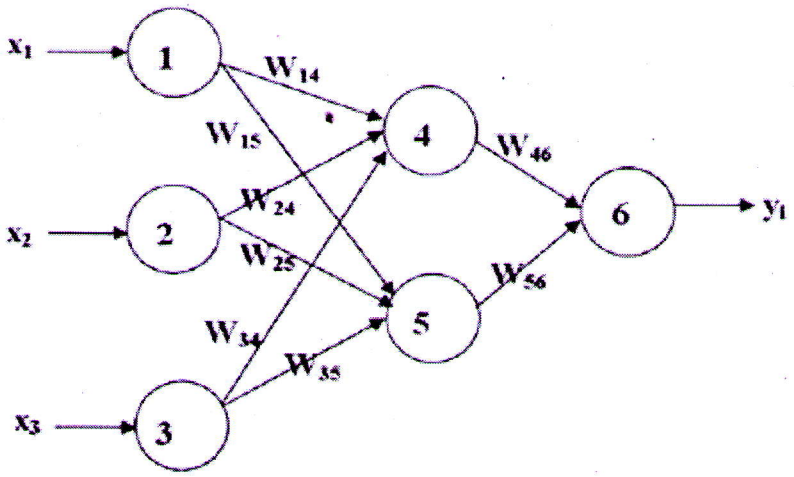


15 a) With the help of an example, demonstrate the working of KNN classifier (5)

b) Compare the information gain calculation methods of C4.5 and ID3 algorithm. Which method is better and why? (4)

16 Using back propagation algorithm, show the updating of weights and biases in the first round using the first training sample (0,1,1) with class label '1' in the given multi-layered feed forward neural network. The initial bias values and weights are given in the table below. Activation function used is 'Sigmoid function' and the learning rate is 0.7 (9)

W_{14}	W_{15}	W_{24}	W_{25}	W_{34}	W_{35}	W_{46}	W_{56}	θ_4	θ_5	θ_6
0.2	-0.1	0.5	-0.3	0.4	0.1	-0.2	0.4	0.1	-0.2	0.1



PART D

Answer any two full questions, each carries 12 marks.

17 Find out all the frequent patterns in the following transaction table with minimum support count 2, using FP-Growth algorithm. Show each step in detail. (12)

TID	List of Items
TID1	biscuit, juice, chocolate
TID2	juice, chocolate, bun
TID3	bun, biscuit, juice, jam
TID4	biscuit, juice, bun
TID5	biscuit, bun, juice, chocolate
TID6	juice, bun, chocolate
TID7	biscuit, juice, jam

18 a) Describe any one of the density-based clustering algorithms. List out the advantages of the same. (6)

b) Explain k-means clustering algorithm with the help of a neat diagram. Give any two disadvantages of the algorithm. (6)

19 a) With the help of a suitable example demonstrate the working of apriori based approach for mining frequent subgraphs. (6)

b) Using the given frequency matrix, calculate the TF-IDF value of the following (6)

i. Term T3 in Document D2

ii. Term T6 in Document D4

	T1	T2	T3	T4	T5	T6
D1	1	5	9	8	0	6
D2	6	11	7	14	2	0
D3	19	0	0	6	3	0
D4	22	7	12	0	4	8
D5	0	2	0	9	5	13

****