VIVEKANANDHA COLLEGE OF ENGINEERING FOR WOMEN
[AUTONOMOUS INSTITUTION AFFILIATED TO ANNA UNIVERSITY, CHENNAI]
Elayampalayam – 637 205, Tiruchengode, Namakkal Dt., Tamil Nadu.

Question Paper Code: 5024

M.E. / M.Tech. DEGREE END-SEMESTER EXAMINATIONS – JUNE / JULY 2024
Second Semester
Computer Science and Engineering
P23CS206 – DATA ANALYTICS
(Regulation 2023)

Time: Three Hours   Maximum: 100 Marks

Answer ALL the questions

Knowledge Levels	K1 – Remembering	K3 – Applying	K5 - Evaluating
(KL)	K2 – Understanding	K4 – Analyzing	K6 - Creating

PART – A

(10 x 2 = 20 Marks)

Q.No.	Questions	Marks	KL	CO
1.	What is Big Data? Why do we need to analyze Big Data?	2	K1	CO1
2.	Differentiate data analysis and data reporting.	2	K1	CO1
3.	Write the purpose of principal component analysis?	2	K2	CO2
4.	What are the parameters used to characterize fuzzy membership function?	2	K1	CO2
5.	Define stream computing?	2	K1	CO3
6.	How are moments estimated?	2	K1	CO3
7.	What is limited pass algorithm? Describe its significance.	2	K1	CO4
8.	Identify the alternative rules for controlling hierarchical clustering.	2	K2	CO4
9.	State the strength and weakness of map reduce.	2	K2	CO5
10.	What is exploratory data analysis?	2	K1	CO5

PART – B

Q.No.	Questions	Marks	KL	CO
11. a)	Outline the various commonly used modern data analytic tools.	13	K1	CO1
	(OR)			
b)	Explain in detail about various phases of data analytics life cycle.	13	K1	CO1
12. a)	Justify why SVM is effective on high dimensional data and discuss the polynomial kernel function for multiple classes.	13	K4	CO2
	(OR)			
b)	Explain in detail about the fuzzy decision tree algorithm and stopping criterion for decision making using an example.	13	K2	CO2
13. a) i.	Explain and analyze the algorithm which eliminates most of the tuples that do not meet the criteria during filtering of streams.	8	K2	CO3
ii.	Explain the algorithm to determine distinct elements appearing in a stream.	5	K2	CO3
	(OR)			
b) i.	Discriminate the concept of sampling data in a stream.	9	K2	CO3
ii.	Illustrate the various Real Time Analytic Platform (RTAP) applications.	4	K1	CO3
14. a)	Explain the concept of Apriori algorithm with an illustrative example.	13	K1	CO4
	(OR)			
b)	Use the K-means algorithm and Euclidean distance to cluster the following 8 examples into 3 clusters: A1 = (2,10), A2 =(2,5), A3=(8,4), A4=(5,8), A5=(7,5) A6=(6,4), A7=(1,2), A8=(4,9). Suppose that the initial seeds (Contents of each cluster) are A1, A4 and A7. Run the k-means algorithm for 1 epoch only. At the end of this epoch show			
i.	The new clusters	5	K2	CO4
ii.	The centres of the new clusters	3	K2	CO4
iii.	How many more iterations are needed to converge? Draw the result for each epoch.	5	K2	CO4

			Marks	KL	CO
15.	a)	Explain and illustrate the various core components of Hadoop framework.	13	K1	CO5
		(OR)			
	b)	Summarize the various visual data analysis techniques.	13	K1	CO5

PART – C

(1 x 15 = 15 Marks)

Q.No.		Questions	Marks	KL	CO
16.	a)	Design a real-time sentiment analysis framework for online movie review application and explain the various steps involved to understand the opinion of the viewer.	15	K4	CO3
		(OR)			
	b)	Suppose there are 100 items, numbered 1 to 100, and also 100 baskets, numbered 1 to 100. Item I is in basket b if and only if I divides b with no reminder. Thus, item 1 is in all baskets, Item 2 is in all fifty of the even numbered baskets, and so on. Basket 12 consists of items {1,2,3,4,6,12}. Since these are all the integers that divide 12. Answer the following questions:			
		i. If the support threshold is 5. Which items are frequent?	5	K2	CO4
		ii. What is the confidence of the following association rules? $\{5,7\} \rightarrow 2$, $\{2,3,4\} \rightarrow 5$.	5+5	K2	CO4