A

B.Tech. Degree VI Regular/Supplementary Examination June 2023

CS 19-202-0604 DATA MINING
(2019 Scheme)

Time: 3 Hours   Maximum Marks: 60

Course Outcome
On successful completion of the course, the students will be able to:

CO1:   Analyse various types of data, its collection and cleaning.
CO2:   Illustrate and analyse various applications of data mining.
CO3:   Analyse and compare various classification models in data mining.
CO4:   Understand developments in big data technologies.
CO5:   Familiarize the concepts of machine learning using R/Python.
CO6:   Analyse and make use of deep learning using R/Python.

Bloom's Taxonomy Levels (BL): L1 – Remember, L2 – Understand, L3 – Apply, L4 –Analyze, L5 – Evaluate, L6 – Create
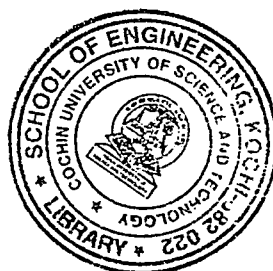PO – Programme Outcome

PART A
(Answer *ALL* questions)

			(8 × 3 = 24)	Marks	BL	CO	PO
I.	(a)	Demonstrate the working of different types of binning in data preprocessing working for the following set of data (2, 4, 9, 11, 13, 15, 22, 24, 26, 35, 37, 39)		3	L3	1	4
	(b)	Consider the following data mining tasks. Classify them as one among classification, clustering and preprocessing. Justify your reasonings for each data mining task		3	L2	2	2
		(i) Remove all outliers from the data and use only data within a range of [0,1].					
		(ii) Group the data with a trained model.					
		(iii) Group the data without a model.					
	(c)	Elucidate on the basic idea behind apriori algorithm.		3	L1	2	1
	(d)	Compare Neural Networks and SVMs.		3	L2	3	2
	(e)	Explain time-series mining.		3	L1	5	1
	(f)	Compare clustering and classification.		3	L2	5	2
	(g)	List the three V's of big data.		3	L1	4	1
	(h)	CNNs are preferred over ANNs for image data. Justify.		3	L2	6	2

PART B
(4 × 12 = 48)

			Marks	BL	CO	PO
II.	(a)	Discuss the different schemas for multidimensional databases with examples.	6	L2	1	1
	(b)	Discuss the major issues to consider during data integration.	6	L2	1	2

OR

			Marks	BL	CO	PO
III.	(a)	Discuss the different stages in data mining.	6	L2	1	1
	(b)	Explain the techniques in data cleaning.	6	L2	1	1

(P.T.O.)

		Marks	BL	CO	PO
IV.	Explain Naïve Bayes Classifier Algorithm. Consider the following dataset. Apply Naïve Bayes classifier on the dataset to predict whether the person plays cricket or not for the tuple X = (outlook: sunny, temperature: mild, humidity: normal, windy: false)	12	L3	3	3

Outlook	Temperature	Humidity	Windy	Play?
Sunny	Hot	High	False	No
Sunny	Hot	High	True	No
Overcast	Hot	High	False	Yes
Rainy	Mild	High	False	Yes
Rainy	Cool	Normal	False	Yes
Rainy	Cool	Normal	True	No
Overcast	Cool	Normal	True	Yes
Sunny	Mild	High	False	No
Sunny	Cool	Normal	False	Yes
Rainy	Mild	Normal	False	Yes
Sunny	Mild	Normal	True	Yes
Overcast	Mild	High	True	Yes
Overcast	Hot	Normal	False	Yes
Rainy	Mild	High	True	No

OR

		Marks	BL	CO	PO
V.	Explain the decision tree algorithm with an example dataset. Make your own assumptions.	12	L1	3	1
VI.	Assume points P1(2, 10), P2(2, 5), P3(8, 4), P4(5, 8), P5(7, 5), P6(6, 4), P7(1, 2), P8(4, 9). Cluster the following eight points into 3 clusters using K-Means Clustering. Explain the working of the algorithm and enumerate its limitations too.	12	L3	5	3

OR

		Marks	BL	CO	PO
VII.	Consider the following two scenarios. Justify which type of clustering is used in each case by defining and reasoning the type of clustering with detailed examples. (i) There are 30 data points about warehouse locations. We have to find a tree of city warehouses such as to find shortest paths for transportation of goods. (ii) We have a set of 30 points with few thick regions and few outliers. We have to group them according to minimum number of points in a group and minimum distance within a group.	12	L3	5	2
VIII. (a)	Explain Hadoop architecture and ecosystem.	8	L1	4	1
(b)	Compare Hadoop and Spark with examples and diagrams.	4	L2	4	2

OR

		Marks	BL	CO	PO
IX. (a)	Discuss on map-reduce framework. Design a solution using this framework to write summary of a text book with many chapters.	6	L2	4	3
(b)	Explain cloud-computing and its services.	6	L2	4	1

Blooms's Taxonomy Levels
L1 – 25%, L2 – 55%, L3 – 20%.

***