SOMAIYA
VIDYAVIHAR

K J Somaiya Institute of Engineering and Information Technology
An Autonomous Institute permanently affiliated to University of Mumbai.
Accredited with A grade by NAAC, approved by AICTE, new Delhi.

Date : 25-05-22

End Semester Exam
April - May 2022/ Nov - Dec 2021

(B. Tech / M. Tech.)   **Electronics and Telecommunication Engineering**

Examination: FY/SY/TY/LY   Semester: I/II/III/IV/V/VI/VII/**VIII**

Course Code: **1UEXDLC8034**   Course Name: **Fundamental of Data Science**

Duration: **03 Hours**   Max. Marks: **60**

Instructions:
1. All questions are compulsory.
2. Draw neat diagrams wherever applicable.
3. Assume suitable data, if necessary.

Question No.	Statement of Question	Marks	CO	BT Level
Q 1	Solve any six questions out of eight:	12		
i)	What is Extended version of ARIMA models ?	2	6	U
ii)	Define outlier	2	3	U
iii)	Define data Warehouse	2	5	U
iv)	List types of OLAP (Online Analytical Process)	2	3	U
v)	List Tasks involved in Data Mining	2	4	U
vi)	List any 6 features of pandas library	2	1	U
vii)	List classification criteria for database mining	2	2	U
viii)	What are the strategies to handle missing terms?	2	4	U
Q.2	Solve any four questions out of six.	16		
i)	Explain Frequent Pattern Mining	4	4	U
ii)	Explain Mining Multilevel Association Rules	4	5	U
iii)	State Issues Regarding Classification and Prediction	4	3	U
iv)	With Suitable example illustrate Data cleansing workflow	4	2	U
v)	Discuss how the exceptions defined. Explain how it is handled with example	4	1	U
vi)	What are steps in designing the data warehouse? Explain in detail	4	3	U

SOMAIYA
VIDYAVIHAR

K J Somaiya Institute of Engineering and Information Technology
An Autonomous Institute permanently affiliated to University of Mumbai.
Accredited with A grade by NAAC, approved by AICTE, new Delhi.

Question No.	Statement of Question	Marks	CO	BT Level
Q.3	Solve any two questions out of three.	16		
i)	Explain decision tree induction algorithm for classifying data tuples and with suitable example	8	5	U
ii)	What is KDD? Explain about data mining as a step in the process of knowledge discovery.	8	3	U
iii)	Discuss about Data Mining Task primitives with examples?	8	4	U
Q.4	Solve any two questions out of three.	16		
i)	Consider the transaction database given below,	8	5, 6	AP

TID	Items
10	1, 3, 4
20	2, 3, 5
30	1, 2, 3, 5
40	2, 5
50	1, 3, 5

Use Apriori Algorithm with min-support count 2 and min-confidence 60% to find all frequent item sets and strong association rules.

		Marks	CO	BT Level
ii)	Suppose that the data for analysis includes the attribute salary. We have the following values of salary (in Thousands of INR), shown in increasing order: 30, 36, 47, 50, 52, 52, 56, 60, 63, 70, 70, 110.	8	1,2	AP

  i. What are the mean, median, mode and midrange of the data?
  ii. Find first quartile, second quartile and third quartile of the data.
  iii. Show the boxplot of the data

		Marks	CO	BT Level
iii)	Suppose that the data mining task is to cluster points (with (x, y) representing location) into three clusters, where the points are: A1 (2, 10), A2 (2, 5), A3 (8, 4), B1 (5, 8), B2 (7, 5), B3 (6, 4), C1 (1, 2), C2 (4, 9). The distance function is Euclidean distance. Suppose initially we assign A1, B1, and C1 as the center of each cluster, respectively. Use the k-means algorithm to show only	8	4	A

  (i) The three cluster centers after the first round of execution
  (ii) The final three clusters