PAPER ID-311549

Roll No: | | | | | | | | | | | | | |

MTECH
(SEM I) THEORY EXAMINATION 2024-25
DATA WAREHOUSING & DATA MINING

TIME: 3 HRS **M.MARKS: 70**

Note: Attempt all Sections. In case of any missing data; choose suitably.

SECTION A

1. Attempt *all* questions in brief. **2 x 07 = 14**

Q no.	Question	CO	Level
a.	Define data extraction in the context of data warehousing.	1	K1
b.	Differentiate between data warehousing and data mining.	1	K2
c.	Explain the importance of data preprocessing in the context of data mining.	2	K2
d.	Define "cross-validation" and explain its purpose in model evaluation.	3	K1
e.	Define outlier analysis in clustering. Why is it important?	4	K1
f.	Explain the role of preprocessing in multimedia data mining and givean example of a preprocessing technique	5	K2
g.	Define web content mining.	5	K1

SECTION B

2. Attempt any *three* of the following: **07 x 3 = 21**

Q no.	Question	CO	Level
a.	Explain 3-tier data warehouse architecture with a help of diagram.	1	K2
b.	Explain the concept of constraint-based association mining.	2	K2
c.	Explain the working of Support Vector Machines in detail. How does it find the optimal hyperplane? Also, discuss its applications with a real-world example.	3	K2
d.	Describe partitioning clustering methods in detail. Explain how K-means clustering works, its strengths and weaknesses, and how it is applied to a real-world problem.	4	K2
e.	Explain multidimensional analysis and its role in mining complex data objects.	5	K4

SECTION C

3. Attempt any *one* part of the following: **07 x 1 = 07**

Q no.	Question	CO	Level
a.	Analyze the importance of data cleanup and transformation tools in the ETL process and how they contribute to the overall quality of the data in a data warehouse.	1	K4
b.	Discuss the differences between OLAP tools and traditional query tools. How do OLAP tools enhance business decision-making through multidimensional data analysis?	1	K2

4. Attempt any *one* part of the following: **07 x 1 = 07**

Q no.	Question	CO	Level
a.	Explain the concept of data integration and transformation.	2	K2

PAPER ID-311549

Roll No: | | | | | | | | | | | | | |

MTECH
(SEM I) THEORY EXAMINATION 2024-25
DATA WAREHOUSING & DATA MINING

TIME: 3 HRS **M.MARKS: 70**

b.	Consider a dataset with the following transactions: Transaction ID | Items Purchased T1 | {A, B, C} T2 | {A, C} T3 | {A, B} T4 | {B, C} T5 | {A, B, C} Given the following constraints: Minimum support threshold: 40% Minimum confidence threshold: 70% (a) Find the frequent itemsets that meet the minimum support threshold. (b) Calculate the association rules that meet the minimum confidence threshold and explain how these rules are filtered based on the constraints.	2	K3

5. Attempt any *one* part of the following: 07 x 1 = 07

Q no.	Question	CO	Level
a.	In a Bayesian classification scenario, suppose you are classifying an email as either "Spam" or "Not Spam". Given the following information: P(Spam) = 0.4, P(Not Spam) = 0.6 P(Keyword | Spam) = 0.9, P(Keyword | Not Spam) = 0.2 P(Email | Spam) = 0.6, P(Email | Not Spam) = 0.8 Using Bayes' Theorem, calculate the posterior probability of the email being spam, given that the email contains the keyword.	3	K3
b.	Explain the concept of back propagation in neural networks. Describe the forward propagation and error back propagation steps, and explain how weights are updated during training.	3	K4

6. Attempt any *one* part of the following: 07 x 1 = 07

Q no.	Question	CO	Level
a.	Explain the concept of density-based clustering methods. Discuss the DBSCAN (Density-Based Spatial Clustering of Applications with Noise) algorithm in detail.	4	K4
b.	Explain constraint-based clustering methods. Discuss the types of constraints used in clustering algorithms.	4	K2

7. Attempt any *one* part of the following: 07 x 1 = 07

Q no.	Question	CO	Level
a.	Explain the role of text mining in analyzing unstructured text data. Discuss the key techniques involved.	5	K2
b.	Explain the difference between web content mining and web structure mining. Discuss how content mining is used for data extraction and how structure mining can be used to understand the relationships between web pages and improve search engine ranking algorithms.	5	K4