Code No: R2022422

R20

SET - 1

II B. Tech II Semester Regular/Supplementary Examinations, April/May-2024

DATA WAREHOUSING AND MINING

(Common to CSE(AIML),CSE(AI),CSE(DS),CSE(AIDS), CSD, AIDS&AIML)

Time: 3 hours

Max. Marks: 70

Answer any **FIVE** Questions each Question from each unit

All Questions carry **Equal** Marks

~~~~~~~~~~~~~~~~~~~~~~~~~

UNIT-I

1	a)	What are some common challenges faced during the implementation of a data warehouse, and how can they be addressed?	[7M]
	b)	How does a data warehouse facilitate decision-making processes within an organization?	[7M]

Or

2	a)	What role do ETL (Extract, Transform, Load) processes play in data warehouse implementation?	[7M]
	b)	What is the relationship between data warehousing and data mining?	[7M]

UNIT-II

3	a)	Define data mining and explain its significance in the field of information technology.	[7M]
	b)	How do similarity and dissimilarity measures contribute to the effectiveness and accuracy of data mining algorithms?	[7M]

Or

4	a)	Discuss the challenges associated with analyzing unstructured data and explain how data mining techniques can be applied to extract meaningful insights from such data sources.	[7M]
	b)	Explore different types of variable transformation techniques, such as normalization and standardization.	[7M]

UNIT-III

5	a)	Differentiate between classification and regression tasks, providing examples of each.	[7M]
	b)	Discuss the basic structure of a decision tree and how it represents decision rules.	[7M]

Or

6	a)	What are the causes for Model Overfitting and underfitting? Discuss various approaches to resolve them.	[7M]
	b)	Define Bayes' theorem and explain the classification model based on this.	[7M]

UNIT-IV

7	a)	Define association analysis and explain its applications in real-time.	[7M]
	b)	Discuss the Apriori principle and how it is used to efficiently generate frequent itemsets.	[7M]

Or

8	a)	Explain the FP-Growth algorithm with an example.	[7M
	b)	How does the FP-Growth algorithm differ from the Apriori algorithm in terms of efficiency and scalability?	[7M
~~~~~~~~~~~~~~~~~~~~~~~~~

Code No:R2022422

R20

SET - 1

UNIT-V

9	a)	Provide examples of scenarios where each type of clustering algorithm is most suitable.	[7M]
	b)	Define partitioning clustering and provide examples of partitioning algorithms.	[7M]

Or

10	a)	Discuss how the K-means algorithm initializes cluster centroids and assigns data points to clusters.	[7M]
	b)	Explain how DBSCAN handles noise points and outlier detection.	[7M]

Code No: R2022422 R20 SET - 2

II B. Tech II Semester Regular/Supplementary Examinations, April-2024
DATA WAREHOUSING AND MINING
(Common to CSE(AIML),CSE(AI),CSE(DS),CSE(AIDS), CSD, AIDS&AIML)

Time: 3 hours **Max. Marks: 70**

Answer any **FIVE** Questions each Question from each unit
All Questions carry **Equal** Marks

UNIT-I

1 a) Describe the typical components of a data warehouse architecture. [7M]
b) How does data integration play a role in the architecture of a data warehouse? [7M]

Or

2 a) What are some common data mining techniques used to extract insights from a data warehouse? [7M]
b) Can you provide examples of how data mining has been successfully applied in real-world business scenarios? [7M]

UNIT-II

3 a) Identify and explain the various dimensions of data quality, such as accuracy, completeness, consistency, and timeliness. [7M]
b) How do similarity and dissimilarity measures contribute to the effectiveness and accuracy of data mining algorithms? [7M]

Or

4 a) What is meant by Dimensionality reduction? Explain with example. [7M]
b) Define data mining and explain its significance in the field of information technology. [7M]

UNIT-III

5 a) Define classification. Discuss its applications and algorithms. [7M]
b) Discuss strategies for handling missing values and categorical attributes in decision tree induction. [7M]

Or

6 a) Describe the random sub-sampling technique for performance evaluation and discuss its use cases. [7M]
b) Discuss the advantages and limitations of the Naïve Bayes classifier in real-world applications. [7M]

UNIT-IV

7 a) Describe the process of candidate generation and pruning in Apriori-based algorithms. [7M]
b) Provide examples of scenarios where frequent itemsets are useful for discovering patterns in data. [7M]

Or

Code No:R2022422 **R20** **SET - 2**

8 a) Explain the process of rule generation from frequent itemsets. [7M]

b) Discuss the main steps involved in the FP-Growth algorithm for frequent itemset generation. [7M]

UNIT-V

9 a) Explain the concept of clustering and its goal in grouping similar objects together. [7M]

b) Discuss the importance of data pre-processing in cluster analysis and its impact on clustering outcomes. [7M]

Or

10 a) Explain hierarchical clustering and discuss the different linkage criteria used in hierarchical clustering. [7M]

b) Discuss how the K-means algorithm initializes cluster centroids and assigns data points to clusters. [7M]

23ftCSA4206.

Code No: R2022422 **R20** **SET - 3**

II B. Tech II Semester Regular/Supplementary Examinations, April-2024
DATA WAREHOUSING AND MINING
(Common to CSE(AIML),CSE(AI),CSE(DS),CSE(AIDS), CSD, AIDS&AIML)

Time: 3 hours **Max. Marks: 70**

Answer any **FIVE** Questions each Question from each unit
All Questions carry **Equal** Marks

UNIT-I

1 a) What is a data warehouse, and how does it differ from a traditional database? [7M]
b) How does a multidimensional data model differ from a relational data model? [7M]

Or

2 a) How does data integration play a role in the architecture of a data warehouse? [7M]
b) How does the knowledge gained from data mining feedback into the data warehouse architecture and decision-making processes? [7M]

UNIT-II

3 a) How does data mining contribute to the process of knowledge discovery in databases (KDD)? [7M]
b) Define data quality and discuss its importance in the context of data mining and decision-making. [7M]

Or

4 a) Define variable transformation and discuss its role in preparing data for analysis. [7M]
b) How do similarity and dissimilarity measures contribute to the effectiveness and accuracy of data mining algorithms? [7M]

UNIT-III

5 a) Explain the difference between supervised and unsupervised classification algorithms. [7M]
b) Explain the concept of feature selection and its role in improving the performance of classification models. [7M]

Or

6 a) How does regularization help prevent overfitting in machine learning models? [7M]
b) Explain the concept of cross-validation and how it addresses the limitations of other evaluation methods. [7M]

UNIT-IV

7 a) Provide examples of real-world applications where association analysis is commonly used. [7M]
b) Provide examples of scenarios where frequent itemsets are useful for discovering patterns in data. [7M]

Or

Code No: R2022422

R20

SET - 3

8 a) How does the Apriori algorithm handle large datasets and high-dimensional data? [7M]

b) Compare and contrast different compact representation methods, such as the FP-tree and FP-list. [7M]

UNIT-V

9 a) Discuss the different types of clustering algorithms based on their approaches, such as partitioning, hierarchical, and density-based. [7M]

b) Discuss the computational complexity and scalability of the K-means algorithm. [7M]

Or

10 a) Compare and contrast the strengths and weaknesses of different types of clustering algorithms. [7M]

b) Describe the basic agglomerative hierarchical clustering algorithm. [7M]

Code No: R2022422 **R20** **SET - 4**

II B. Tech II Semester Regular/Supplementary Examinations, April-2024

DATA WAREHOUSING AND MINING

(Common to CSE(AIML),CSE(AI),CSE(DS),CSE(AIDS), CSD, AIDS&AIML)

Time: 3 hours **Max. Marks: 70**

Answer any **FIVE** Questions each Question from each unit
All Questions carry **Equal** Marks

UNIT-I

1 a) How does data warehousing provide the foundation for data mining activities? [7M]
b) How does a data warehouse facilitate decision-making processes within an organization? [7M]

Or

2 a) What are some key considerations when designing a multidimensional schema for a data warehouse? [7M]
b) How do organizations identify and extract relevant data for inclusion in the data warehouse? [7M]

UNIT-II

3 a) Discuss the key challenges that motivate the need for data mining techniques in modern data analysis. [7M]
b) What is meant by data pre-processing. Explain various techniques used for data Pre-Processing [7M]

Or

4 a) Discuss dimensionality reduction techniques and their importance in handling high-dimensional datasets. [7M]
b) Provide examples of applications where similarity measures are used to perform tasks such as clustering and classification. [7M]

UNIT-III

5 a) Discuss the basic structure of a decision tree and how it represents decision rules. [7M]
b) Provide examples of real-world applications where classification techniques are commonly used. [7M]

Or

6 a) Discuss various approaches to resolve model overfitting in decision trees. [7M]
b) Explain Naive Bayes classification and discuss its limitations. [7M]

UNIT-IV

7 a) Describe the process of candidate generation and pruning in Apriori-based algorithms. [7M]
b) Explain the steps to generate association rules from frequent itemsets. [7M]

Or

8 a) Compare and contrast different compact representation methods, such as the FP-tree and FP-list. [7M]

b) Discuss the main steps involved in the FP-Growth algorithm for frequent item set generation. [7M]

UNIT-V

9 a) Discuss the challenges associated with cluster analysis, such as determining the appropriate number of clusters and handling high-dimensional data. [7M]

b) Compare and contrast the strengths and weaknesses of different types of clustering algorithms. [7M]

Or

10 a) Explain how DBSCAN identifies Core, Border and Noise points. [7M]

b) How to measure inter and intra cluster similarity? Explain with neat diagrams. [7M]