Department : Computer Science and Engineering

Year & Semester : III Year & V Semester

Sub Code & Sub Name : 23CSE354A & Data Warehousing & Data Mining

Unit-I

S.No	Part-A Questions
1.	What is a data warehouse?
2.	Define OLAP.
3.	List two differences between OLAP and OLTP.
4.	What is the purpose of metadata in a data warehouse?
5.	Name any two components of a data warehouse.
6.	Define a fact table.
7.	What is a dimension table?
8.	What is a star schema?
9.	Define concept hierarchy in the context of OLAP.
10.	What is a snowflake schema?
11.	List two typical OLAP operations.
12.	What is a roll-up operation in OLAP?
13.	What is the role of ETL in building a data warehouse?
14.	What is a parallel database system?
15.	Name any two vendors of parallel DBMS.

S.No	Part-B Questions
1.	Explain the basic concepts of data warehousing. Why is a data warehouse needed in modern enterprises?
2.	Describe the architecture of a data warehouse and explain the role of its components (ETL, Metadata, OLAP Engine, etc.).
3.	Explain the process of building a data warehouse. What are the major steps involved?
4.	Compare and contrast OLTP and OLAP systems with examples and characteristics.
5.	What are the major differences among star, snowflake, and fact constellation schemas? Illustrate with diagrams.
6.	Explain the multidimensional data model. How does it support decision-making?
7.	What are concept hierarchies? How are they used in data warehousing and OLAP?
8.	Describe in detail the characteristics of OLAP systems.
9.	What are the typical OLAP operations? Explain with examples of roll-up, drill-down, slice, and dice.
10.	Explain the different types of data warehouse schemas used for decision support.
11.	Discuss the components of a data warehouse with a neat diagram.
12.	Explain the database architectures for parallel processing. What are the advantages of parallel DBMS?

13.	Describe the types of parallelism used in parallel DBMS (inter-query, intra-query, etc.).
14.	Explain the role of parallel DBMS vendors in data warehousing. Mention and compare any two vendors.
15.	Discuss the challenges and best practices in designing and maintaining a data warehouse.

Unit-II

S.No	Part-A Questions
1.	What is data mining?
2.	Define the term 'Knowledge Discovery in Databases (KDD)'.
3.	List any two data mining applications.
4.	Mention two key challenges in data mining.
5.	What is the role of data preprocessing in data mining?
6.	Differentiate between nominal and ordinal attribute types.
7.	What is data cleaning? Give an example.
8.	What is the difference between data integration and data transformation?
9.	Define data discretization.
10.	Name two data reduction techniques.
11.	What is data normalization?
12.	What are outliers in data?
13.	Define similarity and dissimilarity with respect to data.
14.	What is meant by data visualization?
15.	List any two statistical measures used to describe data.

S.No	Part-B Questions
1.	Explain the architecture of a data mining system with a neat diagram.
2.	Describe the steps involved in the Knowledge Discovery Process.
3.	Discuss various data mining techniques with examples.
4.	Explain different issues in data mining. How do they affect the mining process?
5.	Discuss the major applications of data mining across different domains.
6.	Describe the types of data objects and attribute types used in data mining.
7.	Explain statistical description of data using measures of central tendency and dispersion.
8.	What are the different steps in data preprocessing? Explain each in detail.
9.	Discuss in detail about data cleaning techniques.
10.	Explain the process of data integration and the problems associated with it.
11.	Describe data reduction techniques with appropriate examples.
12.	What are data transformation techniques? Explain normalization and discretization.
13.	Explain data visualization techniques used for data analysis.

S.No	
14.	Describe different types of similarity and dissimilarity measures used in data mining.
15.	Why is data preprocessing considered essential in data mining? Illustrate with examples.

Unit-III

S.No	Part-A Questions
1.	What is a frequent pattern in data mining?
2.	Define support and confidence in association rule mining.
3.	What is the Apriori property?
4.	Name two methods for mining frequent itemsets.
5.	What is the difference between frequent, closed, and maximal itemsets?
6.	What is meant by correlation in pattern mining?
7.	List two pattern evaluation measures apart from support and confidence.
8.	What is lift in association rule mining?
9.	What is the main goal of pattern evaluation in frequent pattern mining?
10.	Define multilevel frequent pattern mining with an example.
11.	What is multidimensional pattern mining?
12.	What is constraint-based pattern mining? Give one example of a constraint.
13.	How does classification using frequent patterns work?
14.	What is the role of the FP-Growth algorithm in pattern mining?
15.	Differentiate between single-level and multilevel association rules.

S.No	Part-B Questions
1.	Explain the Apriori algorithm for mining frequent itemsets. How does it reduce the search space using the Apriori property?
2.	Discuss the FP-Growth algorithm. How is it different from the Apriori algorithm in terms of performance and methodology?
3.	Define and differentiate between frequent patterns, closed patterns, and maximal patterns with examples.
4.	Describe the key steps involved in association rule mining. How are support and confidence thresholds used?
5.	What is pattern evaluation in association mining? Discuss the role of interestingness measures like lift, leverage, and conviction.
6.	Explain how frequent pattern mining can be extended to multilevel association rules. Provide an example to support your answer.
7.	Discuss multidimensional association rule mining with suitable examples. How does it differ from single-dimensional mining?
8.	What is constraint-based frequent pattern mining? Explain the types of constraints used with examples.
9.	Describe the challenges and methods of mining frequent patterns in large databases.

10.	How is correlation analysis useful in association rule mining? Explain how it can help identify strong or weak associations.
11.	Explain classification using frequent patterns. How does it differ from traditional classification methods?
12.	Compare and contrast rule-based classification and classification using frequent patterns.
13.	What are the advantages and limitations of using the FP-Growth algorithm for frequent pattern mining?
14.	Explain how interestingness measures help filter unimportant association rules. Give examples.
15.	Illustrate with examples how frequent pattern mining can be applied in real-world scenarios like market basket analysis or web usage mining.

Unit-IV

S.No	Part-A Questions
1.	What is a decision tree?
2.	Define Bayesian classification.
3.	What is a rule-based classifier?
4.	What is backpropagation in neural networks?
5.	Mention one real-world application of Support Vector Machines
6.	Define lazy learners with an example.
7.	Define partitioning methods in clustering
8.	What is cluster analysis?.
9.	Name any two techniques to improve classification accuracy.
10.	What is the purpose of model evaluation in classification?
11.	What is the main idea behind hierarchical clustering?
12.	Name two outlier detection techniques.
13.	Define outlier in data mining.
14.	What is grid-based clustering?
15.	Give one example of a density-based clustering method.

S.No	Part-B Questions
1.	Compare and contrast Partitioning and Hierarchical Clustering methods with suitable examples.
2.	What is Cluster Analysis? Describe the general steps involved in a clustering process.
3.	Explain Density-Based Clustering methods with an example. How does DBSCAN work?
4.	Describe Grid-Based Clustering methods. How do they differ from other clustering approaches?
5.	What is Outlier Analysis? Explain different methods used for outlier detection in large datasets.
6.	What are the challenges in Clustering High Dimensional Data? Discuss methods to

		handle high dimensionality.
7.		How is clustering evaluated? Discuss internal and external cluster evaluation measures.
8.		Explain the process of Decision Tree Induction. How does pruning help improve the model performance?
9.		Discuss Bayesian Classification. How is Naïve Bayes classifier different from general Bayesian classification? Provide examples.
10.		Describe Rule-Based Classification. How are rules generated and evaluated in this method?
11.		Explain the concept of Backpropagation in Neural Networks. How is it used for classification tasks?
12.		What are Support Vector Machines (SVM)? Explain how SVM finds the optimal hyperplane for classification.
13.		Differentiate between Lazy Learners and Eager Learners. Explain with examples of algorithms that follow each approach.
14.		What is Model Evaluation in classification? Discuss the different metrics used for evaluating classification models.
15.		List and explain at least four techniques used to improve classification accuracy.

Unit-V

S.No	Part-A Questions
1.	What is WEKA?
2.	Name any two datasets commonly used in WEKA.
3.	What is the purpose of the Iris dataset?
4.	Define the term "attribute" in the context of datasets.
5.	What is ARFF in WEKA?
6.	What does the Explorer interface in WEKA do?
7.	Mention any two types of learning algorithms in WEKA.
8.	What is the use of the "Preprocess" tab in WEKA Explorer?
9.	What is clustering in machine learning?
10.	Name any two clustering algorithms available in WEKA.
11.	Define association rule mining.
12.	Give an example of an association rule.
13.	What is the output format of the Apriori algorithm?
14.	What is the significance of the "Classify" tab in WEKA?
15.	What type of data is used in the Breast Cancer dataset?

S.No	Part-B Questions
1.	Explain the architecture and components of the WEKA Explorer interface.
2.	Describe the process of loading and preprocessing a dataset in WEKA.
3.	Explain the Iris Plants database. Mention its attributes and typical use cases.
4.	Discuss the Breast Cancer dataset. How is it used in classification?

5.	Write a detailed note on the Auto Imports dataset and its applications in WEKA.
6.	Explain how clustering is performed in WEKA with an example.
7.	What is the Apriori algorithm? How is it implemented in WEKA?
8.	Write the step-by-step procedure to install and run WEKA.
9.	Explain how to visualize dataset attributes and results using WEKA's Explorer.
10.	Describe how cross-validation is implemented in WEKA and why it is important.
11.	How do you evaluate the performance of a machine learning model in WEKA?
12.	Discuss any three learning algorithms in WEKA with their advantages.
13.	Describe the steps to apply a classification algorithm (e.g., J48 or Naive Bayes) using WEKA.
14.	Compare classification and clustering techniques in WEKA with examples.
15.	Explain the process of performing data clustering using the K-Means algorithm in WEKA. Include steps, example dataset, and interpretation of results.

By

V Shaik Mohammad Shahil