



**SIDDARTHA INSTITUTE OF SCIENCE AND TECHNOLOGY:: PUTTUR
(AUTONOMOUS)**

Siddharth Nagar, Narayanavanam Road-517583

Subject with Code: DATA WAREHOUSING & DATA MINING(23CS0534)

Course & Branch: CSIT

Year & Sem: III B.Tech & I Sem

Regulation: R23

UNIT I

1	a)	Define Data Warehouse.	[L1,CO1]	[2M]
	b)	List the phases of data warehouse development.	[L1,CO1]	[2M]
	c)	Name any two parallel DBMS vendors.	[L1,CO1]	[2M]
	d)	What is a Snowflake schema?	[L1,CO1]	[2M]
	e)	What are OLTP and OLAP?	[L1,CO1]	[2M]
2		Illustrate a simple architecture of a data warehouse and explain its components.	[L3,CO1]	[10M]
3		Describe the top-down and bottom-up approaches to building a data warehouse.	[L2,CO1]	[10M]
4		Design a parallel processing architecture for a large e-commerce company.	[L6,CO1]	[5M]
5	a)	Compare MOLAP, ROLAP, and HOLAP based on performance and storage.	[L4,CO1]	[5M]
	b)	Evaluate the multidimensional data model for decision support applications.	[L5,CO1]	[5M]
6	a)	Explain hierarchical levels in dimensions with examples.	[L2,CO1]	[5M]
	b)	Design a star schema for a university performance analysis system.	[L6,CO1]	[5M]
7		Design a star schema for a university performance analysis system.	[L6,CO1]	[10M]
8		Create concept hierarchies for time, product, and location dimensions.	[L6,CO1]	[10M]
9	a)	Explain the characteristics of OLAP systems.	[L2,CO1]	[5M]
	b)	Describe typical OLAP operations with examples.	[L4,CO1]	[5M]
10	a)	Describe typical OLAP operations with examples.	[L2,CO1]	[5M]
	b)	Analyze the differences between OLTP and OLAP systems.	[L4,CO1]	[5M]

UNIT II

1	a)	What do you mean by Knowledge Discovery in Databases (KDD)?	[L1,CO2]	[2M]
	b)	Define nominal and ordinal attributes.	[L1,CO2]	[2M]
	c)	What is discretization?	[L1,CO2]	[2M]
	d)	Name any two visualization tools.	[L1,CO2]	[2M]
	e)	List any three data reduction techniques.	[L1,CO2]	[2M]
2		Explain the process of KDD with a neat diagram.	[L2,CO2]	[10M]
3		Discuss the difference between data mining and data warehousing.	[L2,CO2]	[10M]
4		Analyze the challenges and issues in data mining system design.	[L2,CO2]	[10M]
5	a)	Describe the different types of data objects with examples.	[L2,CO2]	[5M]
	b)	Explain the types of attributes used in data analysis.	[L2,CO2]	[5M]
6	a)	Explain the importance of data preprocessing.	[L2,CO2]	[5M]
	b)	Describe data reduction techniques.	[L4,CO2]	[5M]
7		Evaluate whether visualization techniques improve decision making in business intelligence.	[L5,CO2]	[5M]
8		Explain the role of visualization in decision-making.	[L1,CO2]	[10M]
9		Explain various similarity measures for numerical, categorical, and binary data.	[L2,CO2]	[10M]
10		Differentiate similarity/dissimilarity metrics for clustering tasks.	[L2,CO2]	[10M]

UNIT III

1	a)	List any two limitations of Apriori.	[L1,CO3]	[2M]
	b)	What is support threshold?	[L1,CO3]	[2M]
	c)	Mention the need for multi-level mining.	[L1,CO3]	[2M]
	d)	Define data constraint.	[L1,CO4]	[2M]
	e)	What is predictive accuracy?	[L1,CO4]	[2M]
2	a)	Explain correlation analysis with an example.	[L4,CO3]	[5M]
	b)	Compare Apriori and FP-Growth.	[L4,CO3]	[5M]
3		Analyze the limitations of using only objective measures for rule evaluation	[L4,CO3]	[10M]
4		Explain multi-level frequent pattern mining with an example.	[L3,CO3]	[10M]
5		Describe the challenges in multi-dimensional pattern mining.	[L4,CO3]	[10M]
6		Analyze the efficiency issues in multi-dimensional frequent pattern mining.	[L4,CO3]	[10M]
7	a)	Explain constraint-based pattern generation.	[L3,CO3]	[5M]
	b)	Compare strong and weak constraints.	[L4,CO3]	[5M]
8	a)	Explain the process of associative classification with examples.	[L3,CO3]	[5M]
	b)	Discuss rule generation and rule selection.	[L4,CO3]	[5M]
9		Analyze the strengths and limitations of classification using frequent patterns.	[L4,CO3]	[10M]
10		Construct a classification model using frequent patterns and compare it with traditional classifiers.	[L6,CO4]	[10M]

UNIT IV

Logic concepts

1	a)	Define decision tree.	[L1,CO5]	[2M]
	b)	What are lazy learners?	[L1,CO5]	[2M]
	c)	What is confusion matrix?	[L1,CO5]	[2M]
	d)	Define grid-based clustering.	[L1,CO5]	[2M]
	e)	Mention two uses of outlier detection.	[L2,CO5]	[2M]
2		Explain the process of decision tree induction.	[L3,CO5]	[10M]
3	a)	Discuss the strengths and limitations of naïve Bayes classification.	[L4,CO5]	[5M]
	b)	Explain attribute selection measures used in decision trees.	[L4,CO5]	[5M]
4		Analyze the advantages and limitations of rule-based classifiers.	[L4,CO5]	[10M]
5	a)	Explain the working of SVM with a neat diagram.	[L3,CO5]	[5M]
	b)	Describe methods to improve classification accuracy.	[L3,CO5]	[5M]
6		Explain cross-validation and hold-out method.	[L3,CO5]	[10M]
7		Compare various model selection techniques.	[L4,CO5]	[5M]
			[L2,CO5]	[5M]
8	a)	Explain k-means clustering with an example.	[L3,CO5]	[5M]
	b)	Explain the concept of clustering high-dimensional data.	[L3,CO5]	[5M]
9		Analyze the performance of k-means and DBSCAN on noisy data.	[L4,CO5]	[10M]
10		Explain internal and external clustering validation indices.	[L3,CO5]	[10M]
11		Discuss data distribution-based outlier detection.	[L3,CO5]	[10M]

UNIT V

1	a)	Define supervised and unsupervised datasets.	[L1,CO5]	[2M]
	b)	State any two challenges in Auto Imports dataset.	[L2,CO5]	[2M]
	c)	What is the role of filters in WEKA?	[L2,CO5]	[2M]
	d)	Mention any two model visualization methods.	[L1,CO5]	[2M]
	e)	List two clustering evaluation measures.	[L2,CO5]	[2M]
2		Discuss various types of datasets used in ML experiments.	[L4,CO5]	[10M]
3		Describe the role of datasets in model training and evaluation.	[L3,CO5]	[10M]
4	a)	Compare the classes of Breast Cancer dataset	[L4,CO5]	[5M]
	b)	Explain how the Iris dataset is used for classification experiments.	[L3,CO5]	[5M]
5	a)	Explain the architecture of WEKA.	[L3,CO5]	[5M]
	b)	Discuss preprocessing and filters in WEKA.	[L4,CO5]	[5M]
6	a)	Explain dataset splitting techniques (Train/Test, CV, etc.).	[L3,CO5]	[5M]
	b)	Compare different classifiers for high-dimensional datasets.	[L4,CO5]	[5M]
7		Discuss association rule generation and evaluation.	[L4,CO5]	[10M]
8		Analyze Apriori results using support and confidence thresholds.	[L4,CO5]	[10M]
9		Evaluate clustering performance using different distance metrics.	[L5,CO5]	[10M]
10		Compare Apriori with FP-Growth in terms of performance.	[L4,CO5]	[10M]

Prepared By:
N. Anitha, MTech SIETK

