

# Govt. College of Engineering & Technology, Bikaner

Department of Computer Science & Engineering

Lecture Plan: 7CS3 Data Mining and WareHousing

Name of Faculty: Kartar Singh Siddharth

Unit	Topics	No. of Lectures
I	<b>Overview, Motivation(for Data Mining),Data Mining-Definition &amp; Functionalities, Data Processing, Form of Data Preprocessing, Data Cleaning: Missing Values, Noisy Data, (Binning, Clustering, Regression, Computer and Human inspection), Inconsistent Data, Data Integration and Transformation. Data Reduction:-Data Cube Aggregation, Dimensionality reduction, Data Compression, Numerosity Reduction, Clustering, Discretization and Concept hierarchy generation*.</b>	5
IV	<b>Data Warehousing: Overview, Definition, Delivery Process, Difference between Database System and Data Warehouse, Multi Dimensional Data Model, Data Cubes, Stars, Snow Flakes, Fact Constellations, Concept hierarchy, Process Architecture, 3 Tier Architecture, Data Mining*.</b>	4
II	<b>Concept Description: Definition, Data Generalization, Analytical Characterization, Analysis of attribute relevance, Mining Class comparisons, Statistical measures in large Databases. Measuring Central Tendency, Measuring Dispersion of Data, Graph Displays of Basic Statistical class Description, Mining Association Rules in Large Databases, Association rule mining, mining Single-Dimensional Boolean Association rules from Transactional Databases– Apriori Algorithm, Mining Multilevel Association rules from Transaction Databases and Mining Multi-Dimensional Association rules from Relational Databases*.</b>	6
III	<b>What is Classification &amp; Prediction, Issues regarding Classification and prediction, Decision tree, Bayesian Classification, Classification by Back propagation, Multilayer feed-forward Neural Network, Back propagation Algorithm, Classification methods K-nearest neighbour classifiers, Genetic Algorithm. Cluster Analysis: Data types in cluster analysis, Categories of clustering methods, Partitioning methods. Hierarchical Clustering- CURE and Chameleon*. Density Based Methods-DBSCAN, OPTICS. Grid Based Methods- STING, CLIQUE. Model Based Method – Statistical Approach, Neural Network approach, Outlier Analysis</b>	7
V	Aggregation, Historical information, Query Facility, OLAP function and Tools.OLAP Servers, ROLAP, MOLAP, HOLAP, Data Mining interface, Security, Backup and Recovery, Tuning Data Warehouse, Testing Data Warehouse.	4
	Total	26

\*Topics mentioned in bold are covered so far.