GANDHINAGAR INSTITUTE OF TECHNOLOGY

Information Technology & Computer Engineering Department

Lesson Planning

Semester: 7

Subject: Data Mining & Business Intelligence
Subject Code: 2170715
Subject Coordinator: Prof. Svapnil Vakharia (SMV) - IT
Department Representative: Prof. Sudha Patel (SKP) - CE
Faculty Members:

- Prof. Rahul Vaghela (RAV) - IT
- F2

Text Books:

1. J. Han, M. Kamber, "Data Mining Concepts and Techniques", Morgan Kaufmann
2. M. Kantardzic, "Data mining: Concepts, models, methods and algorithms, John Wiley &Sons Inc.

Unit	Lec. No.	Content	Faculty	Refer Book Chapter
1	**Introduction to Data Warehousing and Business Intelligence (5 Hour)**			
	1	Why reporting and Analysing data, Raw data to valuable information-Lifecycle of Data	RAV,F2	
	2	What is Business Intelligence BI and DW in today's perspective		
	3	What is data warehousing The building Blocks: Defining Features Data warehouses and data 1marts		
	4	Overview of the components Metadata in the data warehouse Need for data warehousing		
	5	Basic elements of data warehousing trends in data warehousing.		
2	**The Architecture of BI and DW (7 Hour)**			
	6	BI and DW architectures and its types Relation between BI and DW	RAV,F2	
	7	OLAP (Online analytical processing) definitions		
	8	Difference between OLAP and OLTP		
	9	Dimensional analysis		
	10	What are cubes? Drill-down and roll-up - slice and dice or rotation		
	11	OLAP models, ROLAP versus MOLAP		
	12	defining schemas: Stars, snowflakes and fact constellations		
	Revision and Tutorial -1			
3	**Introduction to data mining (DM) (4 Hour)**			
	13	Motivation for Data Mining Data Mining-Definition and Functionalities		
	14	Classification of DM Systems		

		DM task primitives	SMV	
	15	Integration of a Data Mining system with a Database or a Data Warehouse		
	16	Issues in DM KDD Process		
4	**Data Pre-processing (7 Hour)**			
	17	Why to pre-process data? Descriptive data summarization	RAV,SMV,F2	
	18	Data cleaning - Missing Values - Noisy Data - Data Cleaning as a Process		
	19	Data Integration and transformation		
	20	Data Reduction		
	21	Data Compression		
	22	Data Mining Primitives Languages and System Architectures -Task relevant data		
	23	Kind of Knowledge to be mined Discretization and Concept Hierarchy		
	Revision and Tutorial -2			
5	**Concept Description & Association Rule Mining (7 Hour)**			
	24	What is concept description?	RAV,SMV,F2	
	25	Data Generalization and summarization-based characterization		
	26	Attribute relevance class comparisons		
	27	Association Rule Mining: - Market basket analysis - basic concepts		
	28	Finding frequent item sets: - Classification of frequent pattern mining - Apriori algorithm - generating rules from frequent itemset		
	29	Improved Apriori algorithm - Mining frequent itemsets without candidate generation - Incremental ARM		
	30	Associative Classification, Rule Mining		
	Revision and Tutorial -3			
6	**Classification and Prediction (7 Hour)**			
	31	What is classification and prediction?	RAV,F2	
	32	Issues regarding Classification and prediction		
	33	Classification methods - Decision tree - Bayesian Classification		
	34	- Rule based Classification - CART, Neural Network		
	35,36	- Linear and non-linear regression - Logistic Regression		
	37	Introduction of tools such as DB Miner /WEKA/DTREG DM Tools		
7	**Data Mining for Business Intelligence Applications (4 Hour)**			
	38,39	Data mining for business Applications like Balanced Scorecard, Fraud Detection, Clickstream Mining,		

		Market Segmentation, retail industry, telecommunications industry, banking & finance and CRM etc.,	RAV,F2	
	40	Data Analytics Life Cycle: Introduction to Big data Business Analytics - State of the practice in analytics role of data scientists		
	41	Key roles for successful analytic project - Main phases of life cycle - Developing core deliverables for stakeholders.		
	Revision and Tutorial -4			
8	**Advance topics (4 Hour)**			
	42	Introduction and basic concepts of following topics: - Clustering, - Spatial mining - web mining - text mining	SMV	
	43	Big Data: Introduction to big data: distributed file system – Big Data and its importance, Four Vs, Drivers for Big data, Big data analytics, Big data applications. Algorithms using map reduce,		
	44	Matrix-Vector Multiplication by Map Reduce. Introduction to Hadoop architecture: Hadoop Architecture, Hadoop Storage: HDFS, Common Hadoop Shell commands		
	45	Anatomy of File Write and Read., NameNode, Secondary NameNode, and DataNode, Hadoop MapReduce paradigm, Map and Reduce tasks, Job, Task trackers – Cluster Setup – SSH & Hadoop Configuration – HDFS Administering –Monitoring & Maintenance.		
	Revision and Tutorial – 5			

Reference Book:

1. Paulraj Ponnian, "Data Warehousing Fundamentals", John Willey.
2. M. Dunham, "Data Mining: Introductory and Advanced Topics", Pearson Education.
3. G. Shmueli, N.R. Patel, P.C. Bruce, "Data Mining for Business Intelligence: Concepts, Techniques, and Applications in Microsoft Office Excel with XLMiner", Wiley India.