Sardar Patel Institute of Technology

Bhavan's Campus, Munshi Nagar, Andheri (West), Mumbai-400058-India
(Autonomous Institute Affiliated to University of Mumbai)

Course Code	Course Name	Teaching Scheme (Hrs/week)			Credits Assigned			
		L	T	P	L	T	P	Total
TEITC604	Data Mining & Business Intelligence	4	-	-	4	-	-	4
		Examination Scheme						
		ISE		MSE	ESE			
		10		30	100 (60% Weightage)			

Pre-requisite Course Codes	IT44 (Database Management Systems) IT34 (Object Oriented Programming)	
After successful completion of the course, student will be able to:		
Course Outcomes	CO1	Demonstrate an understanding of the importance of data mining and the principles of business intelligence.
	CO2	Able to prepare the data needed for data mining algorithms in terms of attributes, class inputs, training, validating, and testing files.
	CO3	Implement classification on large data sets and apply metrics to measure the performance of algorithms.
	CO4	Apply Clustering on large data sets and measure the performance of algorithms.
	CO5	Apply Association mining on large data sets.
	CO6	Apply BI to solve practical problems : Analyze the problem domain, use the data collected in enterprise apply the appropriate data mining technique, interpret and visualize the results and provide decision support.

Module No.	Topics	Ref.	Hrs.
1	**Introduction to Data Mining** What is Data Mining; Kind of patterns to be mined; Technologies used; Major issues in Data Mining	1,5	2
2	**Data Exploration** Types of Attributes; Statistical Description of Data; Data Visualization; Measuring similarity and dissimilarity.	1,5	4
3	**Data Preprocessing** Why Preprocessing? Data Cleaning; Data Integration; Data Reduction: Attribute subset selection, Histograms, Clustering and Sampling; Data Transformation & Data Discretization: Normalization, Binning, Histogram Analysis and Concept hierarchy generation	1,5	4

4	**Classification** Basic Concepts; Classification methods: 1. Decision Tree Induction: Attribute Selection Measures, Tree pruning. 2. Bayesian Classification: Naïve Bayes‟ Classifier. Prediction: Structure of regression models; Simple Linear regression, Multiple linear regression. Model Evaluation & Selection: Accuracy and Error measures, Holdout, Random Sampling, Cross Validation, Bootstrap; Comparing Classifier Performance using ROC Curves. Combining Classifiers: Bagging, Boosting, Random forests	1,5	8
5	**Clustering** Cluster Analysis: Basic Concepts; Partitioning Methods: K-Means, K-Mediods; Hierarchical Methods: Agglomerative, Divisive, BIRCH; Density-Based Methods: DBSCAN, OPTICS	1,6	8
6	**Outlier Analysis** What are outliers? Types, Challenges; Outlier Detection Methods: Supervised, Semi-Supervised, Unsupervised, Proximity based, Clustering Based.	1,6	2
7	**Frequent Pattern Mining** Market Basket Analysis, Frequent Item sets, Closed Item sets, and Association Rules Market Basket Analysis, Frequent Item sets, Closed Item sets, and Association Rules; Frequent Pattern Mining, Efficient and Scalable Frequent Item set Mining Methods, The Apriori Algorithm for finding Frequent Item sets Using Candidate Generation, Generating Association Rules from Frequent Item sets, Improving the Efficiency of Apriori, A pattern growth approach for mining Frequent Item sets; Mining Frequent item sets using vertical data formats; Mining closed and maximal patterns; Introduction to Mining Multilevel Association Rules and Multidimensional Association Rules; From Association Mining to Correlation Analysis, Pattern Evaluation Measures; Introduction to Constraint-Based Association Mining.	1,6	8
8	**Business Intelligence** What is BI? Effective and timely decisions; Data, information and knowledge; The role of mathematical models; Business intelligence architectures; Enabling factors in business intelligence project; Development of a business intelligence system; Ethics and business Intelligence	2,3	3
9	**Decision Support System** Representation of the decision-making process; Evolution of information systems; Definition of decision support system;	2,3	3

	Development of a decision support system		
10	**BI Applications** Data mining for business Applications like Fraud Detection, Click stream Mining, Market Segmentation, retail industry, telecommunications industry, banking & finance CRM etc	2,3	6
Total hours of instructions			**48**

References:

1. Han, Kamber, "*Data Mining Concepts and Techniques*", Morgan Kaufmann 3nd Edition
2. G. Shmueli, N.R. Patel, P.C. Bruce, "*Data Mining for Business Intelligence: Concepts, Techniques, and Applications in Microsoft Office Excel with XLMiner*", 1st Edition, Wiley India.
3. Carlo Vercellis *"Business Intelligence: Data Mining and Optimization for Decision Making"*, Wiley India Publications.
4. P. N. Tan, M. Steinbach, Vipin Kumar, "*Introduction to Data Mining*", Pearson Education
5. Michael Berry and Gordon Lin off , "*Data Mining Techniques*", 2nd Edition Wiley Publications.
6. Michael Berry and Gordon Lin off, "*Mastering Data Mining- Art & science of CRM"*, Wiley Student Edition
7. Vikram Pudi & Radha Krishna, *"Data Mining*", Oxford Higher Education.