

Course Code	Course Name	Teaching Scheme			Credits Assigned			
		Theory	Practical	Tutorial	Theory	Practical/Oral	Tutorial	Total
TEITC604	Data Mining and Business Intelligence	04 Hr/Week	02 Hr/Week	---	04	01	---	05

Course Code	Course Name	Examination Scheme							
		Theory Marks				Term Work	Practical	Oral	Total
		Internal assessment			End Sem. Exam				
		Test1	Test 2	Avg. of 2 Tests					
TEITC604	Data Mining and Business Intelligence	20	20	20	80	25	---	25	150

Course Objectives:

1. To introduce the concept of data Mining as an important tool for enterprise data management and as a cutting edge technology for building competitive advantage.
2. To enable students to effectively identify sources of data and process it for data mining.
3. To make students well versed in all data mining algorithms, methods, and tools.
4. Learning how to gather and analyse large sets of data to gain useful business understanding.
5. To impart skills that can enable students to approach business problems analytically by identifying opportunities to derive business value from data.

Course Outcomes: On successful completion of this course students should be able:

1. Demonstrate an understanding of the importance of data mining and the principles of business intelligence
2. Able to prepare the data needed for data mining algorithms in terms of attributes and class inputs, training, validating, and testing files.
3. Implement the appropriate data mining methods like classification, clustering or association mining on large data sets.
4. Define and apply metrics to measure the performance of various data mining algorithms.
5. 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.

DETAILED SYLLABUS:

Sr. No.	Module	Detailed Content	Hours
1	Introduction to Data Mining	What is Data Mining; Kind of patterns to be mined; Technologies used; Major issues in Data Mining	02
2	Data Exploration	Types of Attributes; Statistical Description of Data; Data Visualization; Measuring similarity and dissimilarity.	04
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.	04
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	08

		Forests.	
5	Clustering	Cluster Analysis: Basic Concepts; Partitioning Methods: K-Means, K-Medoids; Hierarchical Methods: Agglomerative, Divisive, BIRCH; Density-Based Methods: DBSCAN, OPTICS	08
6	Outlier Analysis	What are outliers? Types, Challenges; Outlier Detection Methods: Supervised, Semi-Supervised, Unsupervised, Proximity based, Clustering Based.	02
7	Frequent Pattern Mining	Market Basket Analysis, Frequent Itemsets, Closed Itemsets, and Association Rules; Frequent Pattern Mining, Efficient and Scalable Frequent Itemset Mining Methods, The Apriori Algorithm for finding Frequent Itemsets Using Candidate Generation, Generating Association Rules from Frequent Itemsets, Improving the Efficiency of Apriori, A pattern growth approach for mining Frequent Itemsets; Mining Frequent itemsets 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.	08
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	03
9	Decision Support System	Representation of the decision-making process; Evolution of information systems; Definition of decision support system; Development of a decision support system.	03
10	BI Applications	Data mining for business Applications like Fraud Detection, Clickstream Mining, Market Segmentation, retail industry, telecommunications industry, banking & finance CRM etc	06

Text Books:

1. Han, Kamber, "Data Mining Concepts and Techniques", Morgan Kaufmann 3rd 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. Business Intelligence: Data Mining and Optimization for Decision Making by Carlo Verzellis, Wiley India Publications

Reference Books:

1. P. N. Tan, M. Steinbach, Vipin Kumar, "Introduction to Data Mining", Pearson Education
2. Michael Berry and Gordon Linoff "Data Mining Techniques", 2nd Edition Wiley Publications.
3. Michael Berry and Gordon Linoff "Mastering Data Mining- Art & science of CRM", Wiley Student Edition
4. Vikram Pudi & Radha Krishna, "Data Mining", Oxford Higher Education.

Oral Exam:

An oral exam will be held based on the above syllabus.

Term work:

Assign a case study for group of 2/3 students and each group to perform the following experiments on their case-study; Each group should perform the exercises on a large dataset created by them.

Suggested Practical List:

- 1) 2 tutorials
 - a) Solving exercises in Data Exploration
 - b) Solving exercises in Data preprocessing
- 2) Use WEKA to implement the following Classifiers - Decision tree, Naïve Bayes, Random Forest;
- 3) Implementation of any one classifier using languages like JAVA;
- 4) Use WEKA to implement the following Clustering Algorithms – K-means, Agglomerative, Divisive;
- 5) Implementation of any one clustering algorithm using languages like JAVA;

- 6) Use Weka to implement Association Mining using – Apriori, FPM;
- 7) Detailed study of any one BI tool like Oracle BI, SPSS, Clementine, and XLMiner etc. (paper Assignment)
- 8) Business Intelligence Mini Project: Each group assigned one new case study for this; A BI report must be prepared outlining the following steps:
 - a) Problem definition, Identifying which data mining task is needed
 - b) Identify and use a standard data mining dataset available for the problem. Some links for data mining datasets are: WEKA site, UCI Machine Learning Repository, KDD site, KDD Cup etc.
 - c) Implement the data mining algorithm of choice
 - d) Interpret and visualize the results
 - e) Provide clearly the BI decision that is to be taken as a result of mining.

Theory Examination:

1. Question paper will comprise of 6 questions, each carrying 20 marks.
2. Total 4 questions need to be solved.
3. Q.1 will be compulsory, based on entire syllabus wherein sub questions of 2 to 3 marks will be asked.
4. Remaining question will be randomly selected from all the modules.
5. Weightage of marks should be proportional to number of hours assigned to each module.