

Branch Name:	Information and Communication Technology
Semester/Year:	Semester VII / Final Year
Subject Title:	Data Mining and Business Intelligence (Elective-III)
Subject Code:	1ET1070708
Pre-requisite:	Knowledge of Programming, Algorithms

Course Objective:

The aim of data mining is to discover structure inside unstructured data, extract meaning from noisy data, discover patterns in apparently random data, and use all this information to better understand trends, patterns, correlations, and ultimately predict customer behavior, market and competition trends, so that the company uses its own data more meaningfully to better position itself on the new waves.

Teaching Scheme (Hours per week)				Evaluation Scheme (Marks)				
Lecture (L)	Tutorial (T)	Practical (P)	Credit	Theory (Marks)		Practical (Marks)		Total (Marks)
				University Assessment	Continuous Assessment	University Assessment	Continuous Assessment	
04	-	02	05	70	30	30	20	150

Contents:

Sr. No	Topic	Total Hours	Weight age (%)
1.	Introduction: Why Data Mining? What Is Data Mining? What Kinds of Data Can Be Mined? What Kinds of Patterns Can Be Mined? Which Technologies Are Used? Which Kinds of Applications Are Targeted? Major Issues in Data Mining	04	10%
2.	Data & Exploring Data: Data Objects and Attribute Types, Basic Statistical Descriptions of Data, Data Visualization, Data Cube and OLAP, Data Cube Materialization,	04	10%
3.	Data Preprocessing: An Overview, Data Cleaning, Data Integration, Data Reduction, Data Transformation and Data Discretization	04	08%
4.	Mining Frequent Patterns, Associations, and Correlations: Basic Concepts, Frequent Itemset Mining Methods, Pattern Evaluation Methods	04	08%
5.	Classification: Basic Concepts, Decision Tree Classifier, Model Overfitting, Types of Classifiers, Rule-Based Classifier, Nearest Neighbor Classifiers, Naïve Bayes Classifier, Bayesian Networks, Logistic Regression, Artificial Neural Network, Deep Learning, Support Vector Machine	08	16%

6.	Association Analysis: Frequent Itemset Generation, Rule Generation, Compact Representation of Frequent Itemsets, FP-Growth Algorithm, Evaluation of Association Patterns	06	15%
7.	Cluster Analysis: Overview, K-means, Agglomerative Hierarchical Clustering, DBSCAN, Cluster Evaluation,	06	15%
8.	Anomaly Detection: Characteristics of Anomaly Detection Problems, Characteristics of Anomaly Detection Methods, Statistical Approaches, Proximity-based Approaches, Clustering-based Approaches, Reconstruction-based Approaches, One-class Classification, Information Theoretic Approaches, Evaluation of Anomaly Detection	06	10%
9.	Data Mining for Business Intelligence Applications: Data mining for business Applications like Balanced Scorecard, Fraud Detection, Clickstream Mining, Market Segmentation, retail industry, telecommunications industry, banking & finance and CRM etc., Data Analytics Life Cycle: Introduction to Big data Business Analytics - State of the practice in analytics role of data scientists Key roles for successful analytic project - Main phases of life cycle - Developing core deliverables for stakeholders.	04	08%

Course Outcome:

After learning the course, the students will be able to understand and perform the following task.

1. Pattern Discovery
2. Hidden Insights
3. Frequency Analysis
4. Rare Item Analysis
5. Generating Automatic Rules
6. Discovering Groups of Similar Objects.
7. Eliminating Unwanted or Noisy data.

List of Reference Books:

1. Jiawei Han and Micheline Kamber, “Data Mining: Concepts and Techniques”, Morgan KaufmannPublisher
2. Tan, Pang-Ning and other “Introduction to Data Mining” Pearson Education
3. Dunhum M.H. & Sridhar S. “Data Mining-Introductory and Advanced Topics”, Pearson Education
4. Paulraj & Ponniah, “Data Warehousing Fundamentals”, Oxford University Press
5. Arun K. Pujari, “Data Mining”, Universities Press

List of Suggested titles of Experiments:

1. Review and Comparison of Various Data mining tools.
2. Hands on WEKA data mining tools.
3. Hands on RapidMiner data mining tools.
4. Hands on Orange data mining tools.
5. Hands on R-Langauge in R-Studio.
6. Demonstration of handling missing value (e.g. categorical, numeric) in dataset
7. Creation of Cube and Operation of Cube like slice, dice, pivot, roll-up and Drill down on dataset.
8. Demonstration of basic classifier like decision tree, K-NN etc using data mining tools.
9. Demonstration of associate rule mining in data mining tools.

e-Resources:

1. https://onlinecourses.nptel.ac.in/noc19_cs15/preview
2. <https://spoken-tutorial.org/watch/R/Introduction+to+basics+of+R/English/>
3. <https://www.coursera.org/specializations/data-mining>
4. <https://www.r-project.org>
5. <https://www.edx.org/learn/r-programming>