ITA5007	Data Mining and Business Intelligence	L	T	P	J	C
		3	0	0	4	4
Pre-requisite	**Nil**	**Syllabus version**				
		v. 1.0				

Course Objectives:

1. To learn and apply appropriate data pre-processing techniques.
2. To learn data mining algorithms and significance.
3. To learn to apply appropriate predictive and descriptive mining algorithms for business intelligence

Expected Course Outcomes:

1. Understand the distribution of data and its type to proceed the data pre -processing and mining.
2. Apply data summarization and appropriate pre-processing techniques as per the requirement of the data mining task.
3. Understand and incorporate the statistical models behind prediction process.
4. Apply various representations of classification models and evaluate the performance.
5. Identify the appropriate association data mining techniques to improvise business application.
6. Implement the clustering techniques and apply in real time business applications.
7. Use previously observed values to evaluate and interpret the future results.

Student Learning Outcomes (SLO)	**2,7, 14**

Module:1	**Introduction**	**6 hours**

Data Mining(DM)–origin–rapid growth--Core Ideas in Data Mining-Supervised and Unsupervised Learning - Steps in Data Mining – Data Warehousing -Business Intelligence(BI)-Role of mathematical model, Business Intelligent Architecture, Development of business intelligent system.

Module:2	**Dimension Reduction**	**6 hours**

Data Summaries, Correlation Analysis,Reducing the Number of Categories in Categorical Variables- Converting a Categorical Variable to a Numerical Variable - Principal Components Analysis.

Module:3	**Performance Evaluation and prediction**	**7 hours**

Evaluating Classification and Predictive Performance - Introduction - Judging Classification Performance - Evaluating Predictive Performance –Prediction - Multiple linear regression- Explanatory vs predictive modelling – Estimating the regression equation and prediction variable selection in linear regression.

Module:4	**Classifications**	**6 hours**

Classification methods- Naïve Bayes- K-Neares-Neighbors- classification and regression trees – logistic regression models-Evaluatingclassification performance- Evaluating Goodness of fit - logistic regression for more than two classes

Module:5	**Discriminant Analysis and Association Rules**	**6 hours**
Discriminant analysis-classification performance of discriminant -prior probabilities-unequal classification costs- classifying more than two classes. Association Rules: Introduction - Discovering Association Rules in Transaction Databases - Generating Candidate Rules - Selecting Strong Rules**.**		
Module:6	**Cluster Analysis**	**6 hours**
Cluster analysis –Introduction –distance between two records- measuring distance between two clusters-Hierarchical clustering-Non-hierarchical clustering –k-means algorithm		
Module:7	**Forecasting Time Series**	**6 hours**
Introduction to time series - Explanatory versus Predictive Modelling - Popular Forecasting Methods in Business - Time Series Components - Data Partitioning -Regression-Based Forecasting - Model with Trend - Model with Seasonality - Model with Trend and Seasonality - Autocorrelation and ARIMA Models -Smoothing Methods .		
Module:8	**Contemporary issues**	**2 hours**
Expert Talk		
	Total Lecture hours:	**45 hours**

Text Book(s)

1.	GalitShmueli, Peter C. Bruce, Nitin R. Patel. Data Mining for Business Analytics: Concepts, Techniques and Applications in XL Miner, 2010, 2nd Edition,Wiley Publications
2	Carlo Vercellis, Business Intelligence: Data Mining and optimization for Decision Making, 2009,1st Edition, Wiley Publications..

Reference Books

1.	JiaweiHan, Micheline, Jian Pei. Data Mining: Concepts and Techniques, 2011, 3rd Edition, The Morgan Kaufmann Series.
2	Margaret. H. Dunham, Data Mining: Introductory and Advanced Topics, 2006, 1st Edition, Pearson Education.

Recommended by Board of Studies	05-03-2016		
Approved by Academic Council	40th	Date	18-03-2016