Course Code	Course Title	L	T	P	C
OLMDS508	Data Mining	3	1	0	4
Pre-requisite	NIL	Syllabus version			
		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 businessintelligence

Course Outcomes:

1. Understand the distribution of data and its type to proceed the data pre-processing andmining.
2. Apply data summarization and appropriate pre-processing techniques as per therequirement 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 businessapplication.
6. Implement the clustering techniques and apply in real time business applications.
7. Use previously observed values to evaluate and interpret the future results.

Module:1 Introduction

Data Mining(DM)– Data Mining Functionalities - Rapid Growth - Core Ideas in Data Mining-Supervised and Unsupervised Learning - Steps in Data Mining – Data Warehousing - Business Intelligence(BI) – Online Analytical Processing, Data warehouse Architecture - Business Intelligent Architecture, Development of business intelligent System.

Module:2 Dimension Reduction

Data Summaries, Measuring Similarity and Dissimilarity between data objects, Correlation Analysis, Reducing Categories in Categorical Variables- Converting a Categorical Variable to a Numerical Variable – Feature Subset Selection - Principal Components Analysis.

Module:3 Performance Evaluation and prediction

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

Module:4 Classifications

Classification methods**-** Naïve Bayes- K-Nearest-Neighbors- classification and regression trees using ID3 and CART – Support Vector Machine -Rule-Based Classification- Associative Classification - Evaluating classification performance- Evaluating Goodness of fit - logistic regression.

Module:5	**Discriminant Analysis and Association Rules**
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 **-** FP-Growth Algorithm, FP-Tree Representation	
Module:6	**Cluster Analysis**
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**
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**
	Total Lecture hours: **120 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.	Jiawei Han, 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.

Mode of Evaluation: Quizzes, Assignments, Projects			
Recommended by Board of Studies	27-07-2022		
Approved by Academic Council	No. 67	Date	08-08-2022