Course Code	Course Title	L	T	P	C
OLMCA616	Exploratory Data Analysis	3	1	0	4
Pre-requisite	NIL	Syllabus Version			
		1.0			

Course Objectives

1. To emphasize on the tools and techniques of visual business intelligence.
2. To explore trends, relationships, confirm hypothesis, communicate findings and gain insight about data.
3. To familiarize the techniques for creating interactive visual representations from raw data.

Course Outcomes

Upon completion of this course, learners should be able to

1. Explain Analytical methods for managing large quantities of the data.
2. Describe plot techniques for visualization.
3. Analyze Data sets using the principles of Exploratory Data Analysis.
4. Understand and interpret visual information generated by software.
5. Explain and present the findings in the data sets, after the analysis is completed.

Module:1 Introduction to R

Introduction- Packages and Library - Data types - Basic operators - R objects Vectors -
Lists- Arrays - Matrix- Factors - Data frame- R file formats- Importing and exporting files - Data Visualization in R: Lattice package- Box plot- bar chart - scatter plot- GGplot2

Module:2 Data Evolution

Introduction - Types of Data: Numeric - Categorical - Graphical - High Dimensional Data - Data Classification - Sources of Data: Time Series data- Transactional Data - Biological Data - Spatial Data - Social Network Data- Data Evolution - Data Sources.

Module:3 Data Pre-processing

Need for pre-processing - Data summarization - Data cleaning -Data integration – Data transformation - Data reduction - Data Discretization-Introduction to Missing data, Traditional methods for dealing with missing data, Missing data handling-Multiple Imputation.

Module:4 Data Analysis

Statistical summary measures, data elaboration, 1-D Statistical data analysis, 2-D Statistical data Analysis, n-D Statistical data analysis.

Module:5 Data Visualization

Basic charts and shapes - Histogram - Pie charts -Working with hierarchical data - Visualizing graphs- - Box plot- Multi-vari chart- Andrew Plots -Run chart-Pareto chart- Scatter plot-Stem-and-leaf plot, Working with geo data - Visualizing streaming data: Heat maps, Voronoi diagrams.

Module:6 Outlier Analysis

Outliers and Outlier Analysis - Outlier Detection Methods - Proximity-Based Approaches - distance metrics, Mahalanobis distance, Outlier Detection in High-Dimensional Data.

Module:7 Dimensionality reduction

Dimensionality reduction, unsupervised clustering algorithms, such ask-means clustering, and nearest neighbour algorithms; multidimensional scaling and principal component analysis.			
	Total Hours of Study:		**120 hours**
Text Book(s)			
1.	Martinez, W.L., Martinez, A.R., & Solka, J. (2017). Exploratory Data Analysis with MATLAB (3rd ed.). Chapman and Hall/CRC		
2.	Charu C. Aggarwal, Data Mining: The Text book, Springer, 2015.		
3.	Charu C. Aggarwal, Data Classification Algorithms and Applications, CRC press, 2015.		
Reference Books			
1.	Reiche, N al at, Data-driven storytelling, CRC Press, 2018.		
2.	Tukey, J W. Exploratory Data analysis, Addison-Wesley 1977.		
3.	Jambu M. Exploratory and Multivariate Data Analysis. Academic Press 1991.		
4.	Peng, Roger D. Exploratory Data Analysis with R, 2015.		
Mode of Evaluation: CAT, Digital assignments, Quiz, and FAT			
Recommended by Board of Studies	28-10-2024		
Approved by Academic Council	No. 76	Date	28-11-2024