



Course Title : Data Analytics			
Course Code : P18CS63	Semester : 6	L :T:P : 4:0:0	Credits: 4
Contact Period: Lecture: 52 Hrs, Exam: 3 Hrs		Weightage: CIE:50%, SEE:50%	

### Course Content

#### Unit-1

**Introduction To Data Analytics:** What is Data? A Short Taxonomy of Data Analytics, Examples of Data Use, Breast Cancer in Wisconsin, Polish Company Insolvency Data, A Project on Data Analytics, The KDD Process, The CRISP-DM Methodology **Descriptive Statistics:** Scale Types, Descriptive Univariate Analysis, Descriptive bivariate Analysis. **Multivariate Analysis:** Multivariate Frequencies, Multivariate Data Visualization, Multivariate Statistics.  
Self study component : Operationalize, Case study : Global innovation network and analysis.

10 Hours

#### Unit-2

**Data Quality and Pre-processing:** Data Quality, Missing Values, Redundant Data, Inconsistent Data, Noisy Data Outliers, Converting to a Different Scale Type, Converting to a Different Scale, Data Transformation, Dimensionality Reduction: Attribute Aggregation: Principal Component Analysis. Attribute selection: filters, wrappers. **Review of Basic Data Analytic Methods using R:** Introduction to R: Data types , Data import and Export, Descriptive statistics. **Exploratory Data Analysis:** Visualization tools for single and multivariable: Bar chart, Histogram, line graph, box plot and scatter plot and scatter plot matrix.  
Self study component : Search strategies.

10 Hours

#### Unit-3

**Clustering :** Distance Measures, Difference between Values of Common Attribute Types, Distance Measures for Objects with Quantitative Attributes, Distance Measures for Non-conventional Attributes, Clustering Validation, Clustering Techniques, K-means, Centroids and Distance Measures, How K-means Works, Density-based spatial clustering of applications with noise (DBSCAN), **Frequent Pattern Mining:** Frequent Item sets, Setting the min\_sup Threshold, Apriori – a Join-based Method, Eclat, FP-Growth, Maximal and Closed Frequent Item sets, Association Rules.  
Self study component : Other types of pattern.

11 Hours

#### Unit-4

**Regression:** Predictive Performance Estimation, Generalization, Model Validation, Predictive Performance Measures for Regression, Finding the Parameters of the Model, Linear Regression, **Classification :** Binary Classification , Predictive Performance Measures for Classification, Distance-based Learning Algorithms ,K-nearest Neighbor Algorithms, Case-based Reasoning, Logistic Regression Algorithm, Naive Bayes Algorithm.  
Self study component : Empirical error.

11 Hours

#### Unit-5

**Additional Predictive Methods:** Search-based Algorithms, Decision Tree Induction Algorithms, Decision Trees for Regression **Optimization-based Algorithms:** Support



**Vector Machines Applications for Text, Web and Social Media:** Working with Texts, Recommender Systems.

Self study component : Social network analysis.

10 Hours

**Textbook:**

1. A General Introduction to Data Analytics, João Mendes Moreira, André C.P.L.F. deCarvalho, © 2019 John Wiley & Sons, Inc
2. Data Science & Big Data Analytics , Discovering, Analyzing, Visualizing and Presenting Data, Published by John Wiley & Sons, Inc.

**Reference Books:**

1. Big Data and Data Analytics by Seema Acharya & Subhashini Chellappan by Wiley India Pvt Ltd.

**Course Outcomes:**

1. Students can able to Understanding the Data Analytics Life Cycle & Taxonomy.
2. Students can Analyze Preprocessing of the Data & Analytical Methods with data system With R Tool.
3. Students can able to build and Perform Clustering & Frequent Pattern Mining.
4. Student can able to Implement Additional Predictive Methods including Applications of Text, Web & Social Media.
5. Design and Implement real time applications in data analytics

**CO-PO Mapping**

Semester : 6		Course code : P18CS63					Title : Data Analytics								
CO	Statement	PO 1	PO 2	PO 3	PO 4	PO 5	PO 6	PO 7	PO 8	PO 9	PO 10	PO 11	PO 12	PS O1	PS O2
CO -1	Students can able to Understanding the Data Analytics Life Cycle & Taxonomy	2	2												2
CO-2	Students can Analyze Preprocessing of the Data & Analytical Methods with data system With R Tool.	1	1	1	1	2									2
CO-3	Students can able to build and Perform Clustering & Frequent Pattern Mining.	1	1	1	1	1	1								2
CO -4	Student can able to Implement Additional Predictive Methods including Applications of Text, Web & Social Media	1	1	1	1	1	1								2
CO-5	Design and Implement real time applications in data analytics	2	2	2	2	2	2								2