Semester V

Course Code	Course Name	Course Type	Cd	L	T	P	Marks		
							Sessional	Final Exam	Total
COM-501	Introduction to Data Science	PCC	4	4	0	0	50	100	150

Course Outcomes:

At the end of the course, the student will be able to:	
CO1	Describe big data concepts and emerging trends in analytics.
CO2	Apply data analysis lifecycle phases to real-world case studies.
CO3	Utilize R for descriptive statistics and exploratory data analysis.
CO4	Analyze clustering algorithms and association rules for data patterns.
CO5	Evaluate the effectiveness of regression and classification methods.

Detailed Syllabus

Section-A

Unit 1: Introduction to Big Data Analytics: Big Data Overview: Data Structures, Analyst Perspective on Data Repositories, State of the Practice in Analytics: BI Versus Data Science, Current Analytical Architecture, Drivers of Big Data, Emerging Big Data Ecosystem and a New Approach to Analytics, Key Roles for the New Big Data Ecosystem, Examples of Big Data Analytics. **(8 Hrs.)**

Unit 2: Data Analysis Lifecycle: Lifecycle Overview, Phase 1: Discovery, Phase 2: Data Preparation, Phase 3: Model Planning, Phase 4: Model Building, Phase 5: Communicate Results, Phase 6: Operationalize, Case Study: Global Innovation Network and Analysis (GINA). **(12 Hrs.)**

Unit 3: **Basic Data Analytics Methods using R:** Introduction to R - R Graphical User Interfaces, Data import and export, Attributes and Data Types, Descriptive Statistics, Exploratory Data Analysis, Statistical Methods for Evaluation: Hypothesis Testing, Difference of Means, Wilcoxon Rank-Sum Test, Type I and Type II Errors, Power and Sample Size, ANOVA. **(8 Hrs.)**

Section-B

Unit 4: Clustering and Association Rules: Overview of Clustering, K-means: Use Cases , Overview of the Method , Determining the Number of Clusters , Diagnostics, Association Rules overview, Apriori Algorithm, Evaluation of Candidate Rules , Applications of Association Rules, Examples, Validation and Testing. **(12 Hrs.)**

Unit 5: Advanced Analytical Theory and Methods: Linear Regression and Logistic Regression with Use Cases and Model Description. Decision Trees, Naïve Bayes: Bayes' Theorem, Naïve Bayes Classifier, Smoothing and Diagnostics, Diagnostics of Classifiers. **(8 Hrs.)**

Text Books