Course code	20CSEAC 08	**BIG DATA ANALYTICS**		**L**	**T**	**P**	**C**
Core/Elective/Supportive		Core		**3**	**0**	**1**	**4**
Pre-requisite		Knowledge on Data and data types		**Syllabus Version**		**2020-2021**	

Course Objectives:

The main objectives of this course are to:

1. To understand Data source evolution, data Characteristics and Big data processing models.
2. To understand and apply Data Analytics Techniques on Datasets
3. To analyze and Build Data Analytics use cases for specific domain and applications.

Expected Course Outcomes:

On the successful completion of the course, student will be able to:

1	Understand Data sources, generations, data formats, Data Evolution, Data from various domains	K1, K2
2	Understand Big Data Characteristics , Frameworks , components and Limitation of traditional approaches and map Big Vs to Data Domains	K3
3	Understand the Concepts of Data Analytics Phases and Techniques	K2
4	Apply Data Analytics Techniques practically using R environment	K2-K5
5	Analyze various domains of Data Characteristics, Platform, Programming Model and Design Data Analytic ecosystem, and data processing framework	K4-K5

K1 - Remember; **K2** - Understand; **K3** - Apply; **K4** - Analyze; **K5** - Evaluate; **K6** - Create

Unit:1	**Big Data Landscape**	**18-- hours**

Data Evolution: Data Development Time Line – ICT Advancement-a Perspective – Data Growth-a Perspective – IT Components-Business Process – Landscape-Data to Data Science – Understanding data: Data Classification – Hot Data – Cold Data – Warm Data – Thick Data – Thin Data - Classification of digital Data: Structured, Semi-Structured and Un-Structured. Data Sources - Data Science-Components – Data Science vs Statistics – Mathematics - Programming Language - Database, - Machine Learning. Data Analytics Relation: Data Science, Analytics, Big Data Analytics

Unit:2	**Big Data Components**	**18-- hours**

Big Data: Introduction To Big Data: - Evolution What is Big Data – Sources of Big Data. Characteristics of Big Data 6Vs – Big data-Challenges of Conventional Systems- -– Data Processing Models – Limitation of Conventional Data Processing Approaches – Big Data Myths - Data Discovery-Traditional Approach, Big Data Technology: Big Data Exploration - Data Augmentation – Operational Analysis – 360 View of Customers – Security and Intelligence – Hadoop: Basic Concepts-An Overview of Hadoop-The Hadoop Distributed File System-Anatomy of a Hadoop Cluster-Hadoop Ecosystem Components –SPARK architecture: RDD – Transformation - SPARK Vs Hadoop - NoSQL Database: Types

Unit:3	Data Analytics using R	18-- hours
Data Analytics Classification – Descriptive – Diagnostic – Predictive – Diagnostic – Data Analytics - Case Studies – Data mining in Big Data –Big Data Roles Data Scientist, Data Architect, Data Analyst – Skills - R Basics Data Structures – Vectors – Lists – Tuples – Data Frames – Packages - Visualization plots: Data Analytics: Histogram – Boxplot – Scatter Plot – Bar Chart- Pier Chart – Mosaic Plot – MASS – R Reporting – Markdown – Flex Dashboard		
Unit:4	Data Analytics Techniques	18-- hours
Overview of Data mining: Data Mining Vs. Data Analytics - Data Preprocessing – Unsupervised approaches - –Clustering techniques: Clustering paradigms – Partition algorithm-K- Medeoid algorithms – CLARA- CLARANS –Hierarchical DBSCAN- BIRCH -Categorical clustering algorithms - Introduction to neural network - learning in NN- Genetic algorithm - Classification Technique: Introduction – Decision Trees: Tree Construction Principle - Decision Tree construction Algorithm – CART – ID3 – Random Forest – Rule Based Approaches – Association Rule Mining – Recommender Algorithms		
Unit:5	Data Science Usecases	18-- hours
Data Science - Big Data– Big Data and AI - Use cases – Discussion – Data Sources Identification – Data Types –Data Classification – Data Characteristics of Big V's – Data Science P's – Big Data Frameworks – Data Analytics Classification Domains : Customer Insights – Behavioural Analysis – Marketing – Retails – Insurance – Risk and Security –Health care – Supply Chain Logistics		
Unit:6	Contemporary Issues	2 hours
Addressing Controversy Views of social media – Big Data Source – Data Science Technology - Expert lectures, online seminars – webinars		
	Total Lecture hours	92-- hours