

COURSE HANDOUT

BIG DATA ANALYTICS USING R (MCA-304)

MCA-3RD SEMESTER

ACADEMIC YEAR (2024-25)

Ms. Amita Khanna

Assistant Professor

P.G Department of Computer Applications



P.G Department of Computer Applications

Model Institute of Engineering & Technology (Autonomous)

Kot Bhalwal, Jammu - 181122

www.mietjmu.in

Course Code	Course Name	Course Type	Cd	L	T	P	Marks		
							Sessional	Final Exam	Total
MCA-304	Big Data Analytics using R	PCC	4	4	0	0	20	60	100

COURSE OUTCOMES

At the end of the course the student will be able to:	
CO1	To
CO2	
CO3	
CO4	
CO5	

Unit-I

Introduction to Big Data and Hadoop: Introduction to Big Data and Big Data Analytics, Types of Digital Data, Big data characteristics and applications, Understanding Big Data Storage-A General Overview of High-Performance Architecture, HDFS, MapReduce and YARN, Map Reduce Programming Model, History of Hadoop, Analyzing Data with Hadoop, Hadoop Streaming, Hadoop Echo System, IBM Big Data Strategy, Introduction to Infosphere Big Insights and Big Sheets. (12 Hours)

Unit-II

Hadoop File Distributed System (HDFS) & Map Reduce: The Design of HDFS, HDFS Concepts, Command Line Interface, Hadoop file system interfaces, Data flow, Data Ingest with Flume and Scoop and Hadoop archives, Hadoop I/O: Compression, Serialization, Avro and File-Based Data structures, Anatomy of a Map Reduce Job Run, Failures, Job Scheduling, Shuffle and Sort, Task Execution, Map Reduce Types and Formats, Map Reduce Feature (8 Hours)

Unit-III

Introduction to R programming: Introduction and Overview, Installation of R studio, Data Types and Functions, Data Structures: Vectors and Lists, Matrices, Arrays, Data frames, Factors, Data Transformations and Tables, Control Structures. (10 Hours)

Unit-IV

Data Visualization using R : Reading and getting data into R (External Data): Using CSV files, XML files, Web Data, JSON files, Databases, Excel files, Data Cleaning operations like Slicing, filtering, Piping etc., Exploratory data analysis (8 Hours)

Unit-V

Statistical & Regression Models and Predictive Analysis : Mean, median, Standard deviation, variance, correlation and covariance Random Forest, Decision Trees, Normal and Binomial distributions, Time Series Analysis, Linear and Multiple Regression, Logistic Regression, Survival Analysis Creating data for analytics through designed experiments, creating data for analytics through active learning, Creating data for analytics through reinforcement learning. (8 Hours)

Textbooks

S.No	Name of the Books	Name of the Author	Publisher Name	Edition (Pub.Yr.)
1	Big Data Analytics	Raj Kamal and Preeti Saxena	Mc Graw Hill Education	2nd(2019)
2.	Gardener, Beginning R	Mark Gardener	John Wiley & Sons	2nd (2012)

Reference Books

S.No	Name of the Books	Name of the Author	Publisher Name	Edition (Pub.Yr.)
1	Big Data Analytics with R	Simon Walkowiak	Packt Publishing	(2016)

COURSE PLAN		
Unit-I Introduction to Big data and Hadoop		
S.No	Topics	Recommended Books
1	Primitive Introduction to Bigdata and Hadoop	Book 2, Ch.1
2	Big data characteristics and application	Book 2, Ch.1
3	Big data storage	Book 1, Ch.2
4	HDFS	Book 2, Ch.2
5	Map Reduce and YARN	Book 2, Ch.2
6	Map Reduce Programming Model	Book 2, Ch.2
7	IBM Big Data Strategy	Book 2, Ch.2
Unit-II Hadoop File Distributed System & Map Reduce		
8	Introduction to Infosphere Big Insights and Big Sheets	Book 1, Ch.2
9	Design of HDFS	Book 1, Ch.3
10	Command line interface	Book 2, Ch.3
11	Data Ingest with Flame and scoop& Hadoop archives	Book 2, Ch.4
12	Hadoop I/O	Book 2, Ch.4
13	Anatomy of a Map Reduce Job Run	Book 2, Ch.4
14	Map Reduce Types	Book 2, Ch.4
Unit-III Introduction to R Programming		
15	Map Reduce Features	Book 2, Ch.4
16	Introduction to R Programming	Book 2, Ch.4
17	Installation of R Studio	Book 2, Ch.4
18	Data Types and Functions	Book 2, Ch.5
19	Data Structures	Book 2, Ch.5
20	Data Transformation and Tables	Book 2, Ch.5
21	Control Structures	Book 1, Ch.11
Unit-IV Data Visualization using R		
22	Reading and Getting data into R	Book 2, Ch.12
23	Using CSV files	Book 1, Ch.8
24	XML Files	Book 1, Ch.8
25	Web data and JSON Files	Book 1, Ch.8
26	Databases and Excel Files	Book 1, Ch.8
27	Data Cleaning operations	Book 1, Ch.8
28	Exploratory data analysis	Book 1, Ch.8
Unit-V Statistical & Regression Models and Predictive analysis		
29	Mean and Median	Book 1, Ch.7
30	Standard Deviation and Variance	Book 1, Ch.7
31	Correlation	Book 2, Ch.7
32	Random Forest	Book 1, Ch.7
33	Normal and Binomial distributions	Book 1, Ch.7

34	Time series Analysis	Book1, Ch. 7
35	Linear and Multiple Regression	Book 2, Ch 7
36	Creating data for analytics through active learning	Book 2, Ch.7
37	Creating data for analytics through reinforcement learning	Book 2, Ch.7

ADDITIONAL WEB RESOURCES

1.	MOOC: Data Structures and Algorithms Specializations https://www.coursera.org/specializations/introduction-to-big-data-with-spark-hadoop
2.	NPTEL: Video lectures on Big Data Analytics using R Lecture Series by Dr. Rajiv Misra, Professor, IIT Patna https://onlinecourses.nptel.ac.in/noc24_cs130/preview

GRADING AND ASSESSMENT

- **Sessional Test:** 20 marks
- **Assignment:** 10 marks
- **Attendance:** 10marks
- **Final Examination:** 60 marks

COURSE POLICIES

- **Attendance:** Minimum 75% attendance is mandatory to appear in the final examination of the course.
- **Academic Integrity:** MIET's academic integrity policies apply. Plagiarism will not be tolerated.
- **Late Submissions:** Assignments and projects must be submitted by the specified timelines.

FACULTY INFORMATION

- **Office Hours**
 Monday (01:45 PM – 02:35 PM)
 Friday (01:45 PM – 02:35 PM)
- **Contact Information**
amita.bca@mietjammu.in