

DIPLOMA IN DATA ANALYTICS

Syllabus

(With effect from 2020-21)

Program Code:



DEPARTMENT OF ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING

Bharathiar University

**(A State University Accredited with "a" by NAAAC and
13th Rank among Indian Universities by MHRD-NIRF)**

Coimbatore 641046, INDIA

DIPLOMA IN DATA ANALYTICS

SCHEME OF EXAMINATION

Course Title	Instructional Hours /Week	Examination	
		Duration in Hours	Total Marks
SEMESTER – I			
1.1 Big Data Analytics	5	3	100
1.2.Data Visualization Techniques	5	3	100
1.3.R Programming	5	3	100
1.4. R Programming -Laboratory	5	3	100
TOTAL	-	-	400

BIG DATA ANALYTICS

UNIT I

Types of Digital Data, Introduction to Big Data, Big Data Analytics, History of Hadoop, Apache Hadoop, Analysing Data with Unix tools, Analyzing Data with Hadoop, Hadoop Streaming, Hadoop Echo System, IBM Big Data Strategy, Introduction to Info sphere Big Insights and Big Sheets.

UNIT II:

HDFS(Hadoop Distributed File System) 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.

UNIT III:

Map Reduce Anatomy of a Map Reduce Job Run, Failures, Job Scheduling, Shuffle and Sort, Task Execution, Map Reduce Types and Formats, Map Reduce Features.

Unit IV:

Hadoop Eco System Pig: Introduction to PIG, Execution Modes of Pig, Comparison of Pig with Databases, Grunt, Pig Latin, User Defined Functions, Data Processing operators. Hive : Hive Shell, Hive Services, Hive Megastores, Comparison with Traditional Databases, HiveQL, Tables, Querying Data and User Defined Functions. Hbase: HBasics, Concepts, Clients, Example, Hbase Versus RDBMS. Big SQL : Introduction

UNIT V:

Data Analytics with R Machine Learning: Introduction, Supervised Learning, Unsupervised Learning, Collaborative Filtering. Big Data Analytics with BigR.

TEXT BOOKS

- Tom White “ Hadoop: The Definitive Guide” Third Edit on, O’reily Media, 2012
- Seema Acharya, SubhasiniChellappan, "Big Data Analytics" Wiley 2015.

REFERENCES

- Michael Berthold, David J. Hand, "Intelligent Data Analysis”, Springer, 2007.
- Jay Liebowitz, “Big Data and Business Analytics” Auerbach Publications, CRC press (2013)

- Tom Plunkett, Mark Hornick, “Using R to Unlock the Value of Big Data: Big Data Analytics with Oracle R Enterprise and Oracle R Connector for Hadoop”, McGraw-Hill/Osborne Media (2013), Oracle press.
- AnandRajaraman and Jeffrey David Ulman, “Mining of Massive Datasets”, Cambridge University Press, 2012.
- Bill Franks, “Taming the Big Data Tidal Wave: Finding Opportunities in Huge Data Streams with Advanced Analytics”, John Wiley & sons, 2012.
- Glen J. Myat, “Making Sense of Data”, John Wiley & Sons, 2007
- Pete Warden, “Big Data Glossary”, O’Reily, 2011.
- Michael Mineli, Michele Chambers, AmbigaDhiraj, "Big Data, Big Analytics: Emerging Business Intelligence and Analytic Trends for Today's Businesses", Wiley Publications, 2013.
- ArvindSathi, “BigDataAnalytics: Disruptive Technologies for Changing the Game”, MC Press, 2012
- Paul Zikopoulos ,Dirk DeRoos , Krishnan Parasuraman , Thomas Deutsch , James Giles , David Corigan , "Harness the Power of Big Data The IBM Big Data Platform ", Tata McGraw Hill Publications, 2012

DATA VISUALIZATION TECHNIQUES

UNIT I CORE SKILLS FOR VISUAL ANALYSIS

Information visualization – effective data analysis – traits of meaningful data – visual perception – making abstract data visible – building blocks of information visualization – analytical interaction – analytical navigation – optimal quantitative scales – reference lines and regions – trellises and crosstabs – multiple concurrent views – focus and context – details on demand – over-plotting reduction – analytical patterns – pattern examples.

UNIT II TIME-SERIES, RANKING, AND DEVIATION ANALYSIS

Time-series analysis – time-series patterns – time-series displays – time-series best practices – part-to-whole and ranking patterns – part-to-whole and ranking displays – best practices – deviation analysis – deviation analysis displays – deviation analysis best practices.

UNIT III DISTRIBUTION, CORRELATION, AND MULTIVARIATE ANALYSIS

Distribution analysis – describing distributions – distribution patterns – distribution displays – distribution analysis best practices – correlation analysis – describing correlations – correlation patterns – correlation displays – correlation analysis techniques and best practices – multivariate analysis – multivariate patterns – multivariate displays – multivariate analysis techniques and best practices.

UNIT IV INFORMATION DASHBOARD DESIGN

Information dashboard – Introduction– dashboard design issues and assessment of needs – Considerations for designing dashboard-visual perception – Achieving eloquence.

UNIT V INFORMATION DASHBOARD DESIGN

Advantages of Graphics _Library of Graphs – Designing Bullet Graphs – Designing Sparklines – Dashboard Display Media –Critical Design Practices – Putting it all together- Unveiling the dashboard.

REFERENCES:

1. Ben Fry, "Visualizing data: Exploring and explaining data with the processing environment", O'Reilly, 2008.
2. Edward R. Tufte, "The visual display of quantitative information", Second Edition, Graphics Press, 2001.

3. Evan Stubbs, "The value of business analytics: Identifying the path to profitability", Wiley, 2011.
4. Gert H. N. Laursen and Jesper Thorlund, "Business Analytics for Managers: Taking business intelligence beyond reporting", Wiley, 2010.
5. Nathan Yau, "Data Points: Visualization that means something", Wiley, 2013.
6. Stephen Few, "Information dashboard design: Displaying data for at-a-glance monitoring", second edition, Analytics Press, 2013.
7. Stephen Few, "Now you see it: Simple Visualization techniques for quantitative analysis", Analytics Press, 2009.
8. Tamara Munzner, Visualization Analysis and Design, AK Peters Visualization Series, CRC Press, Nov. 20

R PROGRAMMING

UNIT – I

Introducing to R – R Data Structures – Help Functions in R – Vectors – Scalars – Declarations – Recycling – Common Vector Operations – Using all and any – Vectorized operations – NA and NULL values – Filtering – Vectorised if-then else – Vector Element names. (9).

UNIT – II

Creating matrices – Matrix Operations – Applying Functions to Matrix Rows and Columns – Adding and deleting rows and columns - Vector/Matrix Distinction – Avoiding Dimension Reduction – Higher Dimensional arrays – lists – Creating lists – General list operations – Accessing list components and values – applying functions to lists – recursive lists.

UNIT – III

Creating Data Frames – Matrix-like operations in frames – merging Data frames – Applying functions to Data Frames – Factors and Tables – Factors and levels – Common Functions used with factors – Working with tables – Other factors and table related functions – Control statements – Arithmetic and Boolean operators and values – Default Values for arguments – Returning Boolean Values – Functions are objects – Environment and scope issues – Writing Upstairs – Recursion – Replacement functions – Tools for Composing function code – Math and Simulation in R.

UNIT – IV

S3 Classes – S4 Classes – Managing your objects – Input/output – accessing keyboard and monitor – reading and writing files – accessing the internet – String Manipulation – Graphics – Creating Graphs – Customizing Graphs – Saving Graphs to files – Creating Three-Dimensional plots.

UNIT – V

Interfacing R to other languages – Parallel R – Basic Statistics – Linear Model – Generalized Linear models – Non-linear Models – Time Series and Auto-Correlation – Clustering.

TEXT BOOKS

1. Norman Matloff, “The Art of R Programming: A Tour of Statistical Software Design”, No Starch Press, 2011.
2. Jared P. Lander, “R for Everyone: Advanced Analytics and Graphics”, Addison-Wesley Data & Analytics Series, 2013.

REFERENCE BOOKS

1. Mark Gardner, “Beginning R – The Statistical Programming Language”, Wiley, 2013.
2. Robert Knell, “Introductory R: A Beginner’s Guide to Data Visualisation, Statistical Analysis and programming in R”, Amazon Digital South Asia Services Inc, 2013. Richard Cotton(2013). Learning R, O’Reilly Media.
3. Garret Grolemond (2014). Hands-on Programming with R. O’Reilly Media, Inc.
4. Roger D.Peng (2018). R Programming for Data Science. Lean Publishing.

R PROGRAMMING – LABORATORY

1. R Expressions and Data Structures
2. Manipulation of vectors and matrix
3. Operators on Factors in R
4. Data Frames in R
5. Lists and Operators
6. Working with looping statements.
7. Graphs in R
8. 3D plots in R