II Semester M. Tech. CSE (BDA)

Sr No	Course Code	Course Title	TOTAL Hours			Credits	Maximum Marks		Total	Minimum Passing Marks	No. of Hrs for ESE
			L	T	P		Continual Assessment	End Sem Examination			
1	25BD201T	Data Driven Storytelling and Visualization	3	-	-	3	40	60	100	50	3
2	25BD202T	Programming for Data Analytics	3	-	-	3	40	60	100	50	3
3	25BD203T	Social Media and Big Data Analytics	3	-	-	3	40	60	100	50	3
4	25BD204T	Research Methodology	2	-	-	2	20	30	50	25	1.5
5	25BD205T	Program Elective – II	3	-	-	3	40	60	100	50	3
6	25BD206P	Technical Seminar – II	-	-	6	3	50	-	50	25	-
7	25BD207P	Mini Project-II	-	-	6	3	50	-	50	25	-
Total			14	-	12	20	280	270	550	-	-

Professional Elective Basket

Program Elective – II	
25BD205T (i)	Exploratory Data Analysis for Data Science with R Software
25BD205T (ii)	Scalable Data Science

SECOND SEMESTER

Course Code	Course Name	Th	Tu	Pr	Credits	Evaluation		
						CA	ESE	Total
25BD201T	Data Driven Storytelling and Visualization	3	-	-	3	40	60	100

Course Objectives	Course Outcomes
This course is intended to 1. Understand about Tableau and Power BI's foundational principles for data visualization and analysis. 2. Create diverse visualizations, including charts, matrices, and maps 3. Acquire skills in integrating multiple data sources, utilizing relationships, joins, unions, and advanced blending techniques in Power BI. 4. Develop an interactive dashboard for diverse analyses. 5. Utilize advanced Power BI and Tableau features for insightful data analysis.	**At the end of the Course, the Student will be able to:** 1. Understand and demonstrate the skills acquired in Tableau and PowerBI. 2. Apply the knowledge of varied visualizations such as charts, matrices, and maps by leveraging the functionalities of Tableau and Power BI. 3. Create the real-time applications using the blending techniques in Power 4. Evaluate the interactive dashboard for diverse analyses. 5. Integrate with advanced Power BI and Tableau features for insightful data analysis.

Unit I **[10Hrs]**

Introduction to Tableau for Data Visualization: Introduction to Tableau, Tableau Installation, Tableau Core Concepts, Dimensions and Measures, Aggregation and Granularity, Discrete Date Parts and Continuous Date Values, Chart Types – Bar Chart and Stacked Bar Chart - Line Chart, Sorting – Computed and Manual, Filters – Dimension and Measure, Reference Line and Trend Lines, Scatter Plot – Hierarchies – Cross Tabs and Highlight Tables, Histograms

Unit II **[9Hrs]**

Data Visualizations and Dashboards in Tableau: Advanced Charts in Tableau, Pie Charts, Tree Maps, Word Cloud, Packed Bubbles, Viz in Tooltip - Dual Axis - Lollipop Chart, Add Custom Shapes – Animations, Dashboard Introduction - Create A Profit Analysis Dashboard, Use Dashboard Objects (Navigation and Download), Create Course Comparison Dashboard

Unit III **[7Hrs]**

Data Integration Technique in Tableau: Table Calculations, Create Custom Dates, Scope and Direction- Percent Total, Cohort Analysis, Data Model, Relationships with Calculations, Data Blending, Use Relationships Instead of Blend, Summarize Relationships, Joins, and Blends, Understand Unions

Unit IV **[9Hrs]**

Power BI Visualization Techniques: PowerBI, Introduction -Downloading, Installing, and Activating the New Features in Power BI, Basic Charts In Power BI - Maps In Power BI- Tables and Matrix In Power BI, Formatting a Table, Matrix, Formatting Numbers in a Table, Area Chart, Scatter plot, Waterfall Chart, Tree Map, Gauge Chart, Cards and Filters In Power BI

Unit V **[10Hrs]**

Advanced Power BI Techniques: Advanced Charts In Power BI, Power BI Objects, Power BI Services - Text Functions In Power Query, Merge Queries - Conditional Column - Important Topics In Power BI, working with the Fill up and Fill Down Options, Grouping in Power BI – Transpose – Unpivot - Data Types, Replacing Errors and Values - Keeping and Removing, Rows - Adding, Removing, and Going to Specific Columns

Text Books

S. No	Title	Authors	Edition	Edition
1	Communicating Data with Tableau	Ben Jones	3rd	O'Reilly
2	Power BI Data Analysis and Visualization	Suren Machiraju, Suraj Gaurav	4th	De Gruyter

Reference Books

S. No	Title	Authors	Edition	Publisher
1	Learning Tableau 2020	Joshua N. Milligan	4th	Packt Publishing
2	Power BI Cookbook: Creating Business Intelligence Solutions of Analytical Data Models,Reports, and Dashboards	Brett Powell	2nd	Packt Publishing

		July 2025	1.0	Applicable for 2025-26
Chairman - BoS	Dean – Academics	Date of Release	Version	

SECOND SEMESTER

Course Code	Course Name	Th	Tu	Pr	Credits	Evaluation		
						CA	ESE	Total
25BD202T	Programming for Data Analytics	3	-	-	3	40	60	100

Course Objectives	Course Outcomes
This course is intended to 1. Understand the Fundamental concepts of Data Analytics. 2. Acquire proficiency in Python programming essentials and Numerical data processing. 3. Explore Data manipulation and advanced Data visualization techniques using Python. 4. Obtain an understanding of R programming fundamentals, covering basic operations and manipulation. 5. Develop a practical skill through real-world case studies and projects using Python and R.	**At the end of the Course, the Student will be able to:** 1. Demonstrated the knowledge of foundational concepts in data analytics 2. Attain competence in Python programming essentials for numerical data processing. 3. Utilized Advanced data manipulation and diverse libraries for Dynamic Visualizations in Python. 4. Illustrated the core proficiency in R programming. 5. Create practical expertise via real-world projects in Python and R.

Unit I **[9Hrs]**

Introduction to Data Analytics: Foundation of Data Analytics, use cases of Data Analytic, Describing the Data, Types of Data, Types of Variables, Data Exploration, Data Cleaning and Preprocessing, Feature Engineering, Statistical Analytics, Data Visualizations

[10Hrs]

Unit II

Data Analytics and Data processing using Python: Programming for Data Analytics using Python, Overview, Control flow statements, Strings, Data Structures, Numerical Data Processing, NumPy Dimensions, NumPy Shapes, Size and Bytes, NumPy Arange and Random Package, NumPy and Reshape, NumPy Slicing and Masking, NumPy Broadcasting

[9Hrs]

Unit III

Advanced Data Handling and Visualization in Python: Introduction to Pandas, Pandas Series, Data Frame, Data Manipulation using Pandas, Data Visualization with Matplotlib, Data Visualization with Matplotlib with Seaborn, Interactive Plotting with Bokeh, 3D Plotting with Plotly **,** Geo plotting with Folium, Pandas Data plotting

Unit IV **[8Hrs]**

Introduction to R and Fundamentals: Fundamentals of R, Coercion Rules, Vectors, Vector Operations, Matrices, Conditional and Iterative Statements, Data Frames, Data Manipulation, Dplyr Package, Tidying Data in R

Unit V **[9Hrs]**

Advanced Data Handling and Analysis with R: Data Visualization using ggplot2, Histogram, Bar Chart – Building a Box, Whiskers Plot and Scatterplot, Exploratory Data Analysis, Hypothesis Testing – Type I and Type II Errors, Test for the Mean - Population Variance Known – The P-Value, Test for the Mean - Population Variance Unknown, Linear Regression Analysis

Text Books

S. No	Title	Authors	Edition	Publisher
1	R Programming for Dummies	Vries, Andrie De Meys, Joris		Wiley India (P) Ltd.
2	Data Analytics using R	Seema Acharya		Mcgraw Hill
3	Python for Programmers	Deitel, Paul J Deitel, Harvey M		Pearson Education

Reference Books

S. No	Title	Authors	Edition	Publisher
1	Python for Data Analysis: Dive into the World of Data Science	Wes McKinney	3rd	O'Reilly Media
2	R for Data Science: Statistical Computing and Statistical Thinking with R	Hadley	2nd	O'Reilly Media

		July 2025	1.0	Applicable for 2025-26
Chairman - BoS	Dean – Academics	Date of Release	Version	

SECOND SEMESTER

Course Code	Course Name	Th	Tu	Pr	Credits	Evaluation		
						CA	ESE	Total
25BD203T	Social Media and Big Data Analytics	3	-	-	3	40	60	100

Course Objectives	Course Outcomes
This course is intended to 1. Understand the basic concepts and techniques of social media and Big Data Analytics 2. Develop skills in data acquisition, cleaning, and analysis using big data tools and techniques. 3. Comprehend the principles and methods of predictive modelling. 4. Explore the applications of social media and Big Data Analytics in real-world scenarios. 5. Discuss the ethical considerations in Big Data Analytics.	**At the end of the Course, the Student will be able to:** 1. Remember various NLP techniques and applications such as Sentiment Analyzer and Classifiers. 2. Understand the storage and processing aspects of the Hadoop Eco System. 3. Apply various MapReduce Techniques for data processing. 4. Analyse ML models and various best practices to get optimized results. 5. Create solutions using NLP techniques, Hadoop Eco System, MapReduce Techniques, ML models.

Unit I **[9Hrs]**

Introduction to Natural Language Processing: Introduction to NLP, Setup and Installation, Understanding NLP and Its Benefits, Exploring NLP Tools and Libraries, Tokenization, Stemming and Lemmatization, Named Entity Recognition, Introduction to Sentiment Analysis, Pre-Processing the Dataset, Word Embeddings, Build the Network - Train the Model Building a Chatbot

Unit II **[9Hrs]**

Web Scrapping and Python Interactive APIs: Markup Languages, The Structure of HTML Code, More Complex HTML Parsing, Structuring Parsing Program Better, Splitting HTML Locators out of the Python Class, Understanding HTML with the Browser, Scraping the First Website with Python, signing up for Open exchange rates, Getting All Exchange Rates from the API, Creating a Currency Exchange Library

Unit III **[8Hrs]**

Hadoop Eco System Basics: Introduction and Installation of Hadoop, The Hortonworks and Cloudera Merger and its Effects on the Course, Hadoop Overview and History, Overview of the Hadoop Ecosystem, Hadoop Distributed File System (HDFS), MapReduce, Distributes Processing, MapReduce Example, Ranking Movies by Their Popularity

Unit IV **[10Hrs]**

Hadoop Advanced and Spark Basics: Introducing Ambari, Introducing the Pig, Finding the Oldest Movie with Five-Star Rating Using the Pig, Finding the Old Five-Star Movies with Pig. More Pig Latin - Pig Challenge - Comparing Results, Introduction to Spark - The Resilient Distributed Datasets (RDD) - Datasets and Spark 2.0 -Spark's Machine Learning Library (MLLib), Introduction to Hive - Hive Works, Integrating MySQL with Hadoop, Using Sqoop to Import Data from MySQL to HFDS/Hive, NoSQL – Hbase, Writing Spark Output into Cassandra, Choosing a Database for a Given Problem

Unit V **[9Hrs]**

YARN and Spark Further: Yet Another Resource Negotiator (YARN), Mesos, ZooKeeper, Oozie, Feeding Data to Cluster, Kafka - Analyzing Streams of Data, Spark Streaming- Introduction to Apache Spark – RDD - Map and Filter Transformation, Spark Architecture and Components - Pair RDD - Create Pair RDDs - Reduce by Key Aggregation, Spark SQL - Spark SQL Joins- Dataframe and RDD Conversion - Running Spark in a Cluster.

Text Books

S. No	Title	Authors	Edition	Publisher
1	Natural Language Processing with Python	Steven Bird, Euwan Klein	1st	O'Reilly
2	Big Data for Dummies	Dr. Fern Hapler	1st	Wiley

Reference Books

S. No	Title	Authors	Edition	Publisher
1	Natural Language Processing with Python	Steven Bird, Ewan Klein, and Edward Loper	1st	O'Reilly
2	Natural Language Processing Fundamentals	Ghosh, S., & Gunning, D.	1st	Packt Publishing

		July 2025	1.0	Applicable for 2025-26
Chairman - BoS	Dean – Academics	Date of Release	Version	

SECOND SEMESTER

Course Code	Course Name	Th	Tu	Pr	Credits	Evaluation		
						CA	ESE	Total
25BD204T	Research Methodology	2	-	-	2	20	30	50

Course Objectives	Course Outcomes
This course is intended to 1. Define the concept, objectives, and types of research, and formulate well-structured research problems. 2. Design appropriate research methodologies by selecting suitable designs, methods, and conducting thorough literature reviews. 3. Collect and analyse data using appropriate sampling techniques, ethical guidelines, statistical tools, and data processing software for effective interpretation. 4. Communicate research outcomes through well-structured technical reports, proper referencing, and an understanding of intellectual property, plagiarism.	**At the end of the Course, the Student will be able to:** 1. Understand research fundamentals, problem definition, research design, and literature survey methods. 2. Apply data collection, sampling, ethical practices, statistical analysis, and hypothesis testing techniques. 3. Prepare research reports using proper technical writing, referencing styles, and plagiarism guidelines.

Unit I [10Hrs]

Introduction to Research, Problem Definition and Research Design: Meaning, Objective and importance of research, Types of research, steps involved in research, defining research problem, Research design, Methods of research design, research process and steps involved, Literature, Survey

Unit II [10Hrs]

Data Collection, Data Analysis and interpretation: Classification of Data, Methods of Data Collection, Sampling, Sampling techniques procedure, and methods, Ethical considerations in research, Data analysis, Statistical techniques and choosing an appropriate statistical technique, Hypothesis, Hypothesis testing, Data processing software (e.g. SPSS etc.), statistical inference, Interpretation of results

Unit III [10Hrs]

Technical Writing and reporting of research: Types of research report: Dissertation and Thesis, research paper, review article, short communication, conference presentation etc., Referencing and referencing styles, Research Journals, Indexing and citation of Journals, Intellectual property, Plagiarism

Text Books

S. No	Title	Authors	Edition	Publisher
1	Research Methodology Methods and Techniques	C. R. Kothari, Gaurav Garg,	Third Edition.	New Age International publishers,
2	Business Research Methods	Donald Cooper & Pamela Schindler	9th edition	TMGH
3	Research Methodology: A Step-by-Step Guide for Beginners	Ranjit Kumar	2nd Edition,	SAGE, 2005

SWAYAM/ NPTEL Course

S. No	Course Name	SME Name	Offering Institute	Duration
1	Research Methodology	Prof. Edamana Prasad, Prof. Prathap Haridoss	IIT Madras	8 Weeks

SECOND SEMESTER

Course Code	Course Name	Th	Tu	Pr	Credits	Evaluation		
						CA	ESE	Total
25BD205T(i)	PE – II Exploratory Data Analysis for Data Science with R Software	3	-	-	3	40	60	100

Course Objectives	Course Outcomes
This course is intended to 1. Introduce students to the fundamentals of R programming and its working environment for data analysis. 2. Develop an understanding of data preparation and exploratory statistical data analysis techniques using R. 3. Enable students to apply graphical and visualization techniques 4. Familiarize students with descriptive statistical measures, moments, skewness, kurtosis, and association of variables. 5. Provide hands-on experience in correlation analysis, linear modeling, text data analysis, and sampling methods using R.	**At the end of the Course, the Student will be able to:** 1. Demonstrate proficiency in using R software commands and environment for statistical computing and data analysis. 2. Perform data preparation and exploratory statistical data analysis, including frequency distributions and cumulative distribution functions using R. 3. Apply graphical and visualization techniques using base R and ggplot2 to analyse univariate and multivariate data. 4. Compute and interpret descriptive statistical measures, moments, skewness, kurtosis, and measures of association using R. 5. Analyze real-world datasets using correlation techniques, linear models, text data analysis, sampling methods, and multivariate exploratory data analysis in R.

NPTEL Course: -Exploratory Data Analysis for Data Science with R Software by Prof. Shalabh, IIT Kanpur

Week 1: Introduction to various topics and commands in R software

Week 2: Data Preparation, Basic concepts of exploratory statistical data analysis, frequency and frequency distribution, cumulative distribution functions and their use with R software

Week 3: Graphical procedures with various graphs in one dimension

Week 4: Graphical procedures with various graphs using ggplot2 package

Week 5: Measures of central tendency and their use with R software

Week 6: Measures of variation and their use with R software

Week 7: Moments and their use with R software

Week 8: Skewness, Kurtosis, Scaling of data, and Graphs for visualising the association of variables

Week 9: Graphical procedures for association of variables, Analytical procedures for the association of continuous variables, correlation coefficients and their use with R software

Week 10: Rank correlation, Association of discrete variables and their use with R software

Week 11: Fitting of linear models, Handling text data and their use with R software

Week 12: Analysis of text data, Selection of samples and simple random sampling, Multivariate exploratory data analysis

		July 2025	1.0	Applicable for 2025-26
Chairman - BoS	Dean – Academics	Date of Release	Version	

SWAYAM/ NPTEL Course

S. No	Course Name	SME Name	Offering Institute	Duration
1	Exploratory Data Analysis for Data Science with R Software	Prof. Shalabh	IIT Kanpur	12 Weeks

Text Books

S. No	Title	Authors	Edition	Publisher
1	R for Data Science	Hadley Wickham & Garrett Grolemund	2nd	O'Reilly
2.	Data Science for Business with R	Jeffrey S. Saltz, Jeffrey M. Stanton	2nd	Wiley

Reference Books

S. No	Title	Authors	Edition	Publisher
1	Introduction to Statistics and Data Analysis- With Exercises, Solutions and Applications in R	Christian Heumann, Michael Schomaker and Shalabh	2nd	Springer,
2.	Modern Data Science with R	Benjamin S. Baumer, Daniel T. Kaplan, Nicholas J. Horton B	2nd	CRC Press

SECOND SEMESTER

Course Code	Course Name	Th	Tu	Pr	Credits	Evaluation		
						CA	ESE	Total
25BD205T(ii)	PE – II Scalable Data Science	3	-	-	3	40	60	100

Course Objectives	Course Outcomes
This course is intended 1. Strengthen Proficiency in R Programming 2. Enable Data Handling and Manipulation at Scale 3. Empower Statistical and Business Analytics 4. Introduce Advanced Visualization Techniques 5. Develop Real-World Analytical Applications	**At the end of the Course, the Student will be able to:** 1. Apply advanced programming constructs in R for building scalable and reusable analytics functions. 2. Perform efficient data manipulation and transformation for large and complex business datasets. 3. Implement statistical and machine learning models to derive business insights and forecasts. 4. Create dynamic and interactive dashboards and data visualizations to support business decision-making. 5. Build and deploy complete data analytics applications in R for real-world business scenarios.

NPTEL Course: - Scalable Data Science
By Prof. Anirban Dasgupta, Prof. Sourangshu Bhattacharya | IIT Gadhinagar, IIT Kharagpur

Week 1: Background: Introduction, Probability: Concentration inequalities, Linear algebra: PCA, SVD Optimization: Basics, Convex, GD, Machine Learning: Supervised, generalization, feature learning, clustering

Week 2: Memory-efficient data structures: Hash functions, universal / perfect hash families, Bloom filters, Sketches for distinct count , Misra-Gries sketch

Week 3: Memory-efficient data structures: Count Sketch, Count-Min Sketch, Approximate near neighbors search: Introduction, kd-trees etc, LSH families, MinHash for Jaccard, SimHash for L2

Week 4: Approximate near neighbors search: Extensions e.g. multi-probe, b-bit hashing, Data dependent variants, Randomized Numerical Linear Algebra Random projection

Week 5: Randomized Numerical Linear Algebra CUR Decomposition, Sparse RP, Subspace RP, Kitchen Sink

Week 6: Map-reduce and related paradigms Map reduce - Programming examples - (page rank, k-means, matrix multiplication), Big data: computation goes to data,Hadoop ecosystem

Week 7: Map-reduce and related paradigms: Scala,Spark, Distributed Machine Learning and Optimization: Introduction, SGD, Proof

Week 8: Distributed Machine Learning and Optimization: ADMM + applications, Clustering , Conclusion

Week 9- Week 12: Project assessment on the above topics.

SWAYAM/ NPTEL Course

S. No	Course Name	SME Name	Offering Institute	Duration
1	Scalable Data Science	Prof. Anirban Dasgupta, Prof. Sourangshu Bhattacharya	IIT Gadhinagar, IIT Kharagpur	8 Weeks

Text Books

S. No	Title	Authors	Edition	Publisher
1	Mining of Massive Datasets	J. Leskovec, A. Rajaraman and JD Ullman	2nd Edition	Cambridge University Press,

SECOND SEMESTER

Course Code	Course Name	Th	Tu	Pr	Credits	Evaluation		
						CA	ESE	Total
25BD206P	Technical Seminar – II	-	-	6	3	50	-	50

Course Objectives	Course Outcomes
This course is intended to Provide an opportunity to the students to explore and deepen their knowledge in emerging technologies.	**Students will be able to:** 1. Enhance knowledge and critical thinking 2. Improve presentation skills 3. Learn new research skills and problem-solving technique. 4. Increase their confidence

Student need to present seminar on various emerging technologies to explore and deepen their knowledge in a specific technical field, trends, and industry advancements, enabling them to enhance their expertise, critical thinking, and problem-solving skills within the chosen domain.

SECOND SEMESTER

Course Code	Course Name	Th	Tu	Pr	Credits	Evaluation		
						CA	ESE	Total
25BD207P	Mini Project – II	-	-	6	3	50	-	50

Course Objectives	Course Outcomes
This course is intended to Provide an opportunity for the students to apply the knowledge, develop the skills and provide hands-on experience on a practical based project.	**Students will be able to:** 1. Apply theoretical knowledge to address real-world problems, showcasing research, problem-solving, and critical thinking skills. 2. Effectively communicate project findings and demonstrate proficiency in project management and interdisciplinary learning. 3. Develop practical experience, ethical considerations, and the ability to adapt to challenges in a hands-on learning environment. 4. Develop the ability to document the program, methodologies and results effectively in the form of project report.

Student need to build a project by successfully applying their academic knowledge to solve practical problems, demonstrating research, critical thinking, communication skills in a real-world context and prepare the project report.