

MAHATMA GANDHI INSTITUTE OF TECHNOLOGY (Autonomous)**B.Tech. Minor in Data Science****Scheme of Instruction and Examination****Applicable for the batches admitted from the academic year 2022-23****(MR22 - Regulations)**

S.No	Year/Semester	Course Code	Course Title	Instruction Hours Per Week			Credits
				L	T	P	
1	V -Semester	M155AC	Introduction to Data Science	3	0	0	3
2	V - Semester	M15503	R Programming Laboratory	0	0	3	1.5
3	VI - Semester	M156AC	Data Science Applications	4	0	0	4
4	VII - Semester	M157AC M157AD	Either Online through MOOCS or off-line Class: Data Wrangling and Visualization / Big Data Analytics	3	0	0	3
5	VII - Semester	M15703 M15704	The Corresponding lab: Data Wrangling and Visualization / Big Data Analytics Lab	0	0	3	1.5
6	VIII - Semester	M158AG M158AH M158AI M158AJ M158AK M158AL	Elective: Any ONE of the following subjects 1. Exploratory Data Analysis 2. Mining Massive Databases 3. Social Network Analysis 4. Predictive Analytics 5. Web and Social Media Analytics 6. Video Analytics	3	0	0	3
7	VIII - Semester	M15802	Mini Project	0	0	0	2
Total Credits							18

L: Lecture T: Tutorial P: Practical

B.Tech. Data Science (Minor) V - Semester

L	T	P	C
3	0	0	3

M155AC: INTRODUCTION TO DATA SCIENCE**Course objectives:**

1. Learn concepts, techniques and tools they need to deal with various facets of data science practice, including data collection and integration.
2. Understand the basic types of data and basic statistics
3. Identify the importance of data reduction and data visualization techniques

Course Outcomes: After completion of the course, the student should be able to

1. CO-1: Understand basic terms what Statistical Inference means. Identify probability distributions commonly used as foundations for statistical modeling. Fit a model to data
2. CO-2: describe the data using various statistical measures
3. CO-3: utilize R elements for data handling
4. CO-4: perform data reduction and apply visualization techniques.

UNIT-I: Introduction

What is Data Science? - Big Data and Data Science hype – and getting past the hype - Datafication - Current landscape of perspectives - Statistical Inference - Populations and samples - Statistical modeling, probability distributions, fitting a model – Over fitting.

Basics of R: Introduction, R-Environment Setup, Programming with R, Basic Data Types.

UNIT-II: Data Types & Statistical Description

Types of Data: Attributes and Measurement, What is an Attribute? The Type of an Attribute, The Different Types of Attributes, Describing Attributes by the Number of Values, Asymmetric Attributes, Binary Attribute, Nominal Attributes, Ordinal Attributes, Numeric Attributes, Discrete versus Continuous Attributes.

Basic Statistical Descriptions of Data: Measuring the Central Tendency: Mean, Median, and Mode, Measuring the Dispersion of Data: Range, Quartiles, Variance, Standard Deviation, and Inter-quartile Range, Graphic Displays of Basic Statistical Descriptions of Data.

UNIT-III

Vectors: Creating and Naming Vectors, Vector Arithmetic, Vector sub setting,

Matrices: Creating and Naming Matrices, Matrix Sub setting, Arrays, Class.

Factors and Data Frames: Introduction to Factors: Factor Levels, Summarizing a Factor, Ordered Factors, Comparing Ordered Factors, Introduction to Data Frame, sub setting of Data Frames, Extending Data Frames, Sorting Data Frames. **Lists:** Introduction, creating a List: Creating a Named List, Accessing List Elements, Manipulating List Elements, Merging Lists, Converting Lists to Vectors

UNIT-IV

Conditionals and Control Flow: Relational Operators, Relational Operators and Vectors, Logical Operators, Logical Operators and Vectors, Conditional Statements.

Iterative Programming in R: Introduction, While Loop, For Loop, Looping Over List.

Functions in R: Introduction, writing a Function in R, Nested Functions, Function Scoping, Recursion, Loading an R Package, Mathematical Functions in R.

UNIT-V:

Data Reduction: Overview of Data Reduction Strategies, Wavelet Transforms, Principal Components Analysis, Attribute Subset Selection, Regression and Log-Linear Models: Parametric Data Reduction, Histograms, Clustering, Sampling, Data Cube Aggregation.

Data Visualization: Pixel-Oriented Visualization Techniques, Geometric Projection Visualization Techniques, Icon-Based Visualization Techniques, Hierarchical Visualization Techniques, Visualizing Complex Data and Relations.

TEXT BOOKS:

1. Doing Data Science, Straight Talk from The Frontline. Cathy O’Neil and Rachel Schutt, O’Reilly, 2014
2. **Jiawei Han**, Micheline Kamber and Jian Pei. Data Mining: Concepts and Techniques, 3rd ed. The Morgan Kaufmann Series in Data Management Systems.
3. K G Srinivas, G M Siddesh, “Statistical programming in R”, Oxford Publications.

REFERENCE BOOKS:

1. Introduction to Data Mining, Pang-Ning Tan, Vipin Kumar, Michael Steinbach, Pearson Education.
2. Brain S. Everitt, “A Handbook of Statistical Analysis Using R”, Second Edition, 4 LLC, 2014.
3. Dalgaard, Peter, “Introductory statistics with R”, Springer Science & Business Media, 2008.
4. Paul Teetor, “R Cookbook”, O’Reilly, 2011.

B.Tech. Data Science (Minor) V - Semester

L	T	P	C
0	0	3	1.5

M15503: R PROGRAMMING LABORATORY

1. R Environment setup: Installation of R and RStudio in Windows
2. Write R commands for
 - i. Variable declaration and retrieving the value of the stored variables,
 - ii. Write an R script with comments,
 - iii. Type of a variable using class () Function.
3. Write R command to illustrate summation, subtraction, multiplication, and division operations on vectors using vectors. Enumerate multiplication and division operations between matrices and vectors in R console
4. Write R command to
 - i. Illustrate the usage of Vector sub setting& Matrix sub setting
 - ii. Write a program to create an array of 3×3 matrixes with 3 rows and 3 columns.
 - iii. Write a program to create a class, object, and function
5. Write a command in R console
 - i. to create a tshirt_factor, which is ordered with levels 'S', 'M', and 'L'. Is it possible to identify from the examples discussed earlier, if blood type 'O' is greater or less than blood type 'A'?
 - ii. Write the command in R console to create a new data frame containing the 'age' parameter from the existing data frame. Check if the result is a data frame or not. Also R commands for data frame functions *cbind()*, *rbind()*, *sort()*
6. Write R command for
 - i. Create a list containing strings, numbers, vectors and logical values
 - ii. To create a list containing a vector, a matrix, and a list. Also give names to the elements in the list and display the list also access the list elements
 - iii. To add a new element at the end of the list and delete the element from the middle display the same
 - iv. To create two lists, merge two lists. Convert the lists into vectors and perform addition on the two vectors. Display the resultant vector.
7. Write R command for
 - i. logical operators—AND (&), OR (|) and NOT (!). ii. Conditional Statements
 - ii. Create four vectors namely patientid, age, diabetes, and status. Put these four vectors into a data frame patientdata and print the values using a for loop& While loop
 - iii. Create a user-defined function to compute the square of an integer in R
 - iv. Create a user-defined function to compute the square of an integer in R
 - v. Recursion function for a) factorial of a number b) find nth Fibonacci number
8. Write R code for i) Illustrate Quick Sort ii) Illustrate Binary Search Tree
9. Write R command to
 - i. illustrate Mathematical functions & I/O functions
 - ii. Illustrate Naming of functions and *sapply()*, *lapply()*, *tapply()* & *mapply()*

10. Write R command for

- i. Pie chart& 3D Pie Chart, Bar Chart to demonstrate the percentage conveyance of various ways for traveling to office such as walking, car, bus, cycle, and train
- ii. Using a chart legend, show the percentage conveyance of various ways for traveling to office such as walking, car, bus, cycle, and train.

Walking is assigned red color, car – blue color, bus – yellow color, cycle – green color, and train – white color; all these values are assigned through *cols and lbls variables and the legend function*.

- a. *The fill parameter is used to assign colors to the legend.*
- b. Legend is added to the top-right side of the chart, by assigning
 - i. Using box plots, Histogram, Line Graph, Multiple line graphs and scatter plot to demonstrate the relation between the cars speed and the distance taken to stop, Consider the parameters data and *x* Display the *speed* and *dist* parameter of Cars data set using *x* and data parameters

TEXT BOOK:

1. K G Srinivas, G M Siddesh, “Statistical programming in R”, Oxford Publications.

B.Tech. Data Science (Minor) VI - Semester

L	T	P	C
4	0	0	4

M156AC: DATA SCIENCE APPLICATIONS

Course Objective: To give deep knowledge of data science and how it can be applied in various fields to make the life easy.

Course Outcomes: After completion of course, students would:

1. To correlate the data science and solutions to modern problem.
2. To decide when to use which type of technique in data science.

UNIT - I

Data Science Applications in various domains, Challenges and opportunities, tools for data scientists Recommender systems – Introduction, methods, application, challenges.

UNIT - II

Time series data – stock market index movement forecasting. Supply Chain Management – Real world case study in logistics

UNIT - III

Data Science in Education, Social media

UNIT - IV

Data Science in Healthcare, Bioinformatics

UNIT - V

Case studies in data optimization using Python.

TEXT BOOKS:

1. Aakanksha Sharaff, G.K.Sinha , “Data Science and its applications “, CRC Press, 2021.
2. Q.A.Menon, S.A.Khoja, “Data Science: Theory, Analysis and Applications”, CRC Press, 2020.

B.Tech. Data Science (Minor) VII – Semester

L	T	P	C
3	0	0	3

M157AC: DATA WRANGLING AND DATA VISUALIZATION**Course Outcomes:**

1. To learn data wrangling techniques
2. To introduce visual perception and core skills for visual analysis

Course Outcomes: Upon completion of the course, the students will be able to:

1. Perform data wrangling
2. Explain principles of visual perception
3. Apply core skills for visual analysis
4. Apply visualization techniques for various data analysis tasks
5. Evaluate visualization techniques

UNIT - I:

Data Wrangling: Need of data cleanup, data clean up basics – formatting, outliers, duplicates, Normalizing and standardizing data.

UNIT - II:

Introduction of visual perception, visual representation of data, Gestalt principles, information overloads. Creating visual representations, visualization reference model, visual mapping, visual analytics, Design of visualization applications.

UNIT - III:

Classification of visualization systems, Interaction and visualization techniques misleading, Visualization of one, two and multi-dimensional data, text and text documents.

UNIT - IV:

Visualization of groups, trees, graphs, clusters, networks, software, Metaphorical visualization

UNIT - V:

Visualization of volumetric data, vector fields, processes and simulations, Visualization of maps, geographic information, GIS systems, collaborative visualizations, evaluating visualizations

TEXT BOOKS:

1. Jacqueline Kazil and Katharine Jarmul, Data Wrangling with Python: Tips and Tools to Make Your Life Easier, O'Reilly.
2. Ward, Grinstein Keim, Interactive Data Visualization: Foundations, Techniques, and Applications. Natick A K Peters, Ltd.

REFERENCE BOOK:

1. E. Tufte, The Visual Display of Quantitative Information, Graphics Press.

B.Tech. Data Science (Minor) VII – Semester

L	T	P	C
3	0	0	3

M157AD: BIG DATA ANALYTICS**Course Objectives:**

1. The purpose of this course is to provide the students with the knowledge of Big data analytics principles and techniques.
2. This course is also designed to give an exposure of the frontiers of Big data Analytics

Courses Outcomes:

1. Ability to explain the foundations, definitions, and challenges of Big Data and various Analytical tools.
2. Ability to program using HADOOP and Map reduce, NOSQL
3. Ability to understand the importance of Big Data in Social Media and Mining.

UNIT - I

Introduction to Big Data: Big Data and its Importance – Four V's of Big Data – Drivers for Big Data – Introduction to Big Data Analytics – Big Data Analytics applications.

UNIT - II

Big Data Technologies: Hadoop's Parallel World – Data discovery – Open source technology for Big Data Analytics – cloud and Big Data – Predictive Analytics – Mobile Business Intelligence and Big Data

UNIT - III

Introduction Hadoop: Big Data – Apache Hadoop & Hadoop EcoSystem – Moving Data in and out of Hadoop – Understanding inputs and outputs of MapReduce - Data Serialization.

UNIT - IV

Hadoop Architecture: Hadoop: RDBMS Vs Hadoop, Hadoop Overview, Hadoop distributors, HDFS, HDFS Daemons, Anatomy of File Write and Read., NameNode, Secondary NameNode, and DataNode, HDFS Architecture, Hadoop Configuration, Map Reduce Framework, Role of HBase in Big Data processing, HIVE, PIG.

UNIT - V

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

TEXT BOOKS:

1. Big Data Analytics, Seema Acharya, Subhasini Chellappan, Wiley 2015.
2. Big Data, Big Analytics: Emerging Business Intelligence and Analytic Trends for Today's Business, Michael Minelli, Michele Chambers, 1st Edition, Ambiga Dhiraj, Wiely CIO Series, 2013.
3. Hadoop: The Definitive Guide, Tom White, 3rd Edition, O'Reilly Media, 2012.
4. Big Data Analytics: Disruptive Technologies for Changing the Game, Arvind Sathi, 1st Edition, IBM Corporation, 2012.

REFERENCE BOOKS:

1. Big Data and Business Analytics, Jay Liebowitz, Auerbach Publications, CRC press (2013)
2. Using R to Unlock the Value of Big Data: Big Data Analytics with Oracle R Enterprise and Oracle R Connector for Hadoop, Tom Plunkett, Mark Hornick, McGraw-Hill/Osborne Media (2013), Oracle press.
3. Professional Hadoop Solutions, Boris lublinsky, Kevin t. Smith, Alexey Yakubovich, Wiley, ISBN: 9788126551071, 2015.
4. Understanding Big data, Chris Eaton, Dirk deroos et al., McGraw Hill, 2012.
5. Intelligent Data Analysis, Michael Berthold, David J. Hand, Springer, 2007.
6. Taming the Big Data Tidal Wave: Finding Opportunities in Huge Data Streams with Advanced Analytics, Bill Franks, 1st Edition, Wiley and SAS Business Series, 2012.

B.Tech. Data Science (Minor) VII – Semester

L	T	P	C
0	0	3	1.5

M15703: DATA WRANGLING AND VISUALIZATION LAB**Course Objectives:**

1. To learn data wrangling techniques
2. To introduce visual perception and core skills for visual analysis

Course Outcomes: Upon completion of the course, the students will be able to:

1. Perform data imputation
2. Find and handle outliers and duplicates
3. Perform normalization on data
4. Visualization using charts (2-D and multiple dimension)
5. Data visualization using trees and networks

List of Experiments:

Implement the following experiments using Python

1. Find missing values and perform data imputation.
2. Find outliers in a chosen dataset.
3. Methods to handle duplicate data.
4. Perform data normalization
5. Explore 2-D charts such as Clustered bar charts, connected dot plots, pictograms, bubble charts, radar charts, polar charts, Range chart, Box-and-whisker plots, univariate scatter plots, histograms word cloud, pie chart, waffle chart, stacked bar chart, tree map.
6. Multi-dimensional data visualization
7. Graph data visualization

TEXT BOOKS:

1. Jacqueline Kazil and Katharine Jarmul, Data Wrangling with Python: Tips and Tools to Make Your Life Easier, O'Reilly
2. Andy Kirk, Data Visualization A Handbook for Data Driven Design, Sage Publications, 2016
3. Philipp K. Janert, Gnuplot in Action, Understanding Data with Graphs, Manning Publications, 2010.

B.Tech. Data Science (Minor) VII – Semester

L	T	P	C
0	0	3	1.5

M15704: BIG DATA ANALYTICS LAB**Course Objectives:**

1. The purpose of this course is to provide the students with the knowledge of Big data Analytics principles and techniques.
2. This course is also designed to give an exposure of the frontiers of Big data Analytics

Course Outcomes:

1. Use Excel as a Analytical tool and visualization tool.
2. Ability to program using HADOOP and Map reduce
3. Ability to perform data analytics using ML in R.
4. Use cassandra to perform social media analytics

List of Experiments

1. Implement a simple map-reduce job that builds an inverted index on the set of input documents (Hadoop)
2. Process big data in HBase
3. Store and retrieve data in Pig
4. Perform Social media analysis using cassandra
5. Buyer event analytics using Cassandra on suitable product sales data.
6. Using Power Pivot (Excel) Perform the following on any dataset
 - a) Big Data Analytics
 - b) Big Data Charting
7. Use R-Project to carry out statistical analysis of big data
8. Use R-Project for data visualization of social media data

TEXT BOOKS:

1. Big Data Analytics, Seema Acharya, Subhasini Chellappan, Wiley 2015.
2. Big Data, Big Analytics: Emerging Business Intelligence and Analytic Trends for Today's
3. Business, Michael Minelli, Michehe Chambers, 1st Edition, Ambiga Dhiraj, Wiely CIO Series, 2013.
4. Hadoop: The Definitive Guide, Tom White, 3rd Edition, O'Reilly Media, 2012.
5. Big Data Analytics: Disruptive Technologies for Changing the Game, Arvind Sathi, 1st Edition, IBM Corporation, 2012.

REFERENCE BOOKS:

1. Big Data and Business Analytics, Jay Liebowitz, Auerbach Publications, CRC press (2013)
2. Using R to Unlock the Value of Big Data: Big Data Analytics with Oracle R Enterprise and Oracle R Connector for Hadoop, Tom Plunkett, Mark Hornick, McGraw-Hill/Osborne Media (2013), Oracle press.
3. Professional Hadoop Solutions, Boris lublinsky, Kevin t. Smith, Alexey Yakubovich, Wiley, ISBN: 9788126551071, 2015.
4. Understanding Big data, Chris Eaton, Dirk deroos et al., McGraw Hill, 2012.
5. Intelligent Data Analysis, Michael Berthold, David J. Hand, Springer, 2007.
6. Taming the Big Data Tidal Wave: Finding Opportunities in Huge Data Streams with Advanced Analytics, Bill Franks, 1st Edition, Wiley and SAS Business Series, 2012.

B.Tech. Data Science (Minor) VIII – Semester

L	T	P	C
3	0	0	3

M158AG: EXPLORATORY DATA ANALYSIS**Course Objectives:**

1. This course introduces the methods for data preparation and data understanding.
2. It covers essential exploratory techniques for understanding multivariate data by summarizing it through statistical methods and graphical methods.
3. Supports to Summarize the insurers use of predictive analytics, data science and Data Visualization.

Course Outcomes:

1. Handle missing data in the real-world data sets by choosing appropriate methods.
2. Summarize the data using basic statistics. Visualize the data using basic graphs and plots.
3. Identify the outliers if any in the data set.
4. Choose appropriate feature selection and dimensionality reduction
5. Techniques for handling multi-dimensional data

UNIT - I:

Introduction To Exploratory Data Analysis: Data Analytics lifecycle, Exploratory Data Analysis (EDA) – Definition, Motivation, Steps in data exploration, The basic data types Data Type Portability.

UNIT - II:

Preprocessing - Traditional Methods and Maximum Likelihood Estimation: Introduction to Missing data, Traditional methods for dealing with missing data, Maximum Likelihood Estimation – Basics, Missing data handling, Improving the accuracy of analysis. Preprocessing Bayesian Estimation: Introduction to Bayesian Estimation, Multiple Imputation- Imputation Phase, Analysis and Pooling Phase, Practical Issues in Multiple Imputation, Models for Missing Notation Random Data.

UNIT - III:

Data Summarization & Visualization: Statistical data elaboration, 1-D Statistical data analysis, 2-D Statistical data Analysis, N-D Statistical data analysis.

UNIT - IV:

Outlier Analysis: Introduction, Extreme Value Analysis, Clustering based, Distance Based and Density Based outlier analysis, Outlier Detection in Categorical Data.

Feature Subset Selection: Feature selection algorithms: filter methods, wrapper methods and embedded methods, Forward selection backward elimination, Relief, greedy selection, genetic algorithms for features selection.

UNIT - V

Dimensionality Reduction: Introduction, Principal Component Analysis (PCA), Kernel PCA, Canonical Correlation Analysis, Factor Analysis, Multidimensional scaling, Correspondence Analysis.

TEXT BOOK:

1. Making sense of Data: A practical Guide to Exploratory Data Analysis and Data Mining, by Glenn J. Myatt

REFERENCE BOOKS:

1. Charu C. Aggarwal, “Data Mining the Text book”, Springer, 2015.
2. Craig K. Enders, “Applied Missing Data Analysis”, The Guilford Press, 2010.
3. Inge Koch, “Analysis of Multivariate and High dimensional data”, Cambridge University Press, 2014.
4. Michael Jambu, “Exploratory and multivariate data analysis”, Academic Press Inc., 1990.
5. Charu C. Aggarwal, “Data Classification Algorithms and Applications”, CRC press, 2015.

B.Tech. Data Science (Minor) VIII – Semester

L	T	P	C
3	0	0	3

M158AH: MINING MASSIVE DATABASES

Prerequisites: Students should be familiar with Data mining, algorithms, basic probability theory and Discrete math.

Course Objectives:

1. This course will cover practical algorithms for solving key problems in mining of massive datasets.
2. This course focuses on parallel algorithmic techniques that are used for large datasets.
3. This course will cover stream processing algorithms for data streams that arrive constantly, page ranking algorithms for web search, and online advertisement systems that are studied in detail.

Course Outcomes:

1. Handle massive data using MapReduce.
2. Develop and implement algorithms for massive data sets and methodologies in the context of data mining.
3. Understand the algorithms for extracting models and information from large datasets.
4. Develop recommendation systems.
5. Gain experience in matching various algorithms for particular classes of problems.

UNIT - I:

Data Mining-Introduction-Definition of Data Mining-Statistical Limits on Data Mining. MapReduce and the New Software Stack-Distributed File Systems, MapReduce, Algorithms Using MapReduce.

UNIT - II:

Similarity Search: Finding Similar Items-Applications of Near-Neighbor Search, Shingling of Documents, Similarity-Preserving Summaries of Sets, Distance Measures. Streaming Data: Mining Data Streams-The Stream Data Model, Sampling Data in a Stream, Filtering Streams.

UNIT - III:

Link Analysis-PageRank, Efficient Computation of PageRank, Link Spam. Frequent Itemsets-Handling Larger Datasets in Main Memory, Limited-Pass Algorithms, Counting Frequent Items in a Stream. Clustering-The CURE Algorithm, Clustering in Non-Euclidean Spaces, Clustering for Streams and Parallelism.

UNIT IV:

Advertising on the Web-Issues in On-Line Advertising, On-Line Algorithms, The Matching Problem, The Adwords Problem, Adwords Implementation. Recommendation Systems-A Model for Recommendation Systems, Content-Based Recommendations, Collaborative Filtering, Dimensionality Reduction, The Netflix Challenge.

UNIT - V:

Mining Social-Network Graphs-Social Networks as Graphs, Clustering of Social-Network Graphs, Partitioning of Graphs, Simrank, Counting Triangles

TEXT BOOKS:

1. Jure Leskovec, Anand Rajaraman, Jeff Ullman, Mining of Massive Datasets, 3rd Ed.

REFERENCE BOOKS:

1. Jiawei Han & Micheline Kamber, Data Mining – Concepts and Techniques, 3rd Edition Elsevier.
2. Margaret H Dunham, Data Mining Introductory and Advanced topics, PEA.
3. Ian H. Witten and Eibe Frank, Data Mining: Practical Machine Learning Tools and Techniques, Morgan Kaufmann.

B.Tech. Data Science (Minor) VIII – Semester

L	T	P	C
3	0	0	3

M158AI: SOCIAL NETWORK ANALYSIS**Pre-requisites:**

1. A course on “Web Technologies”;
2. A course on “Computer Networks”;
3. A course on “Data Warehousing and Data Mining”.

Course Objectives:

1. It introduces the concepts of social media
2. It provides the mechanisms for social network analysis
3. Includes the concepts that allow for better visualization and analysis of widely used services such as email, Wikis, Twitter, flickr, YouTube, etc.

Course Outcomes:

1. Ability to construct social network maps easily.
2. Gain skills in tracking the content flow through the social media.
3. Use NodeXL to perform social network analysis.

UNIT - I:

Introduction: Social Media and Social Networks.

Social Media: New Technologies of Collaboration.

Social Network Analysis: Measuring, Mapping, and Modelling collections of Connections.

UNIT - II:

NodeXL, Layout, Visual Design, and Labelling, Calculating and Visualizing Network Metrics, Preparing Data and Filtering, Clustering and Grouping.

UNIT - III:

CASE STUDIES - I:

Email: The lifeblood of Modern Communication.

Thread Networks: Mapping Message Boards and Email Lists.

Twitter: Conversation, Entertainment and Information.

UNIT - IV:

CASE STUDIES - II:

Visualizing and Interpreting Facebook Networks, WWW Hyperlink Networks.

UNIT-V:

CASE STUDIES - III:

You Tube: Contrasting Patterns of Content Interaction, and Prominence.

Wiki Networks: Connections of Creativity and Collaboration.

TEXT BOOKS:

1. Hansen, Derek, Ben Sheiderman, Marc Smith, Analysing Social Media Networks with NodeXL: Insights from a Connected World, Morgan Kaufmann, 2011.
2. Avinash Kaushik, Web Analytics 2.0: The Art of Online Accountability, Sybex, 2009.

REFERENCE BOOK:

1. Marshall Sponder, Social Media Analytics: Effective Tools for Building, Interpreting and Using Metrics, 1st Edition, McGraw Hill, 2011.

B.Tech. Data Science (Minor) VIII – Semester

L	T	P	C
3	0	0	3

M158AJ: PREDICTIVE ANALYTICS

Course Objectives: The course serves to advance and refine expertise on theories, approaches and techniques related to prediction and forecasting.

Course Outcomes

1. Understand prediction-related principles, theories and approaches.
2. Learn model assessment and validation.
3. Understand the basics of predictive techniques and statistical approaches.
4. Analyze supervised and unsupervised algorithms

UNIT - I

Linear Methods for Regression and Classification: Overview of supervised learning, Linear regression models and least squares, Multiple regression, Multiple outputs, Subset selection, Ridge regression, Lasso regression, Linear Discriminant Analysis, Logistic regression, Perceptron learning algorithm.

UNIT - II

Model Assessment and Selection: Bias, Variance, and model complexity, Bias-variance trade off, Optimism of the training error rate, Estimate of In-sample prediction error, Effective number of parameters, Bayesian approach and BIC, Cross-validation, Bootstrap methods, conditional or expected test error.

UNIT - III

Additive Models, Trees, and Boosting: Generalized additive models, Regression and classification trees, Boosting methods-exponential loss and AdaBoost, Numerical Optimization via gradient boosting, Examples (Spam data, California housing, New Zealand fish, Demographic data).

UNIT - IV

Neural Networks (NN), Support Vector Machines (SVM), and K-nearest Neighbor: Fitting neural networks, Back propagation, Issues in training NN, SVM for classification, Reproducing Kernels, SVM for regression, K-nearest –Neighbor classifiers (Image Scene Classification).

UNIT - V

Unsupervised Learning and Random forests: Association rules, Cluster analysis, Principal Components, Random forests and analysis.

TEXT BOOK:

1. Trevor Hastie, Robert Tibshirani, Jerome Friedman, The Elements of Statistical Learning- Data Mining, Inference, and Prediction, Second Edition, Springer Verlag, 2009

REFERENCE BOOKS:

1. C. M. Bishop –Pattern Recognition and Machine Learning, Springer, 2006
2. L. Wasserman-All of statistics
3. Gareth James. Daniela Witten. Trevor Hastie Robert Tibshirani. An Introduction to Statistical Learning with Applications in R

B.Tech. Data Science (Minor) VIII – Semester

L	T	P	C
3	0	0	3

M158AK: WEB AND SOCIAL MEDIA ANALYTICS

Course Objectives: Exposure to various web and social media analytic techniques.

Course Outcomes:

1. Knowledge on decision support systems.
2. Apply natural language processing concepts on text analytics.
3. Understand sentiment analysis.
4. Knowledge on search engine optimization and web analytics.

UNIT – I:

An Overview of Business Intelligence, Analytics, and Decision Support: Analytics to Manage a Vaccine Supply Chain Effectively and Safely, Changing Business Environments and Computerized Decision Support, Information Systems Support for Decision Making, The Concept of Decision Support Systems (DSS), Business Analytics Overview, Brief Introduction to Big Data Analytics.

UNIT – II:

Text Analytics and Text Mining: Machine Versus Men on Jeopardy! The Story of Watson, Text Analytics and Text Mining Concepts and Definitions, Natural Language Processing, Text Mining Applications, Text Mining Process, Text Mining Tools,

UNIT – III:

Sentiment Analysis: Sentiment Analysis Overview, Sentiment Analysis Applications, Sentiment Analysis Process, Sentiment Analysis and Speech Analytics.

UNIT – IV:

Web Analytics, Web Mining: Security First Insurance Deepens Connection with Policyholders, Web Mining Overview, Web Content and Web Structure Mining, Search Engines, Search Engine Optimization, Web Usage Mining (Web Analytics), Web Analytics Maturity Model and Web Analytics Tools.

UNIT – V:

Social Analytics and Social Network Analysis: Social Analytics and Social Network Analysis, Social Media Definitions and Concepts, Social Media Analytics.

Prescriptive Analytics - Optimization and Multi-Criteria Systems: Multiple Goals, Sensitivity Analysis, What-If Analysis, and Goal Seeking.

TEXT BOOK:

1. Ramesh Sharda, Dursun Delen, Efraim Turban, Business Intelligence and Analytics: Systems for Decision Support, Pearson Education.

REFERENCE BOOKS:

1. Rajiv Sabherwal, Irma Becerra - Fernandez, Business Intelligence Practice, Technologies and Management”, John Wiley 2011.
2. Lariss T. Moss, Shaku Atre, “Business Intelligence Roadmap”, Addison-Wesley It Service.
3. Yuli Vasiliev, “Oracle Business Intelligence: The Condensed Guide to Analysis and Reporting”, SPD Shroff, 2012.

B.Tech. Data Science (Minor) VIII – Semester

L	T	P	C
3	0	0	3

M158AL: VIDEO ANALYTICS

Course Objectives: To acquire the knowledge of extracting information from surveillance videos, understand the models used for recognition of objects, humans in videos and perform gait analysis.

Course Outcomes:

1. Understand the basics of video- signals and systems.
2. Able to estimate motion in a video
3. Able to detect the objects and track them
4. Recognize activity and analyze behaviour
5. Evaluate face recognition technologies

UNIT – I:

Introduction: Multidimensional signals and systems: signals, transforms, systems, sampling theorem. Digital Images and Video: human visual system and color, digital video, 3D video, digital-video applications, image and video quality.

UNIT – II:

Motion Estimation: Image formation, motion models, 2D apparent motion estimation, differential methods, matching methods, non-linear optimization methods, transform domain methods, 3D motion and structure estimation.

UNIT – III:

Video Analytics: Introduction - Video Basics - Fundamentals for Video Surveillance - Scene Artifacts - Object Detection and Tracking: Adaptive Background Modelling and Subtraction- Pedestrian Detection and Tracking Vehicle Detection and Tracking- Articulated Human Motion Tracking in Low-Dimensional Latent Spaces.

UNIT – IV:

Behavioural Analysis & Activity Recognition: Event Modelling- Behavioural Analysis- Human Activity Recognition-Complex Activity Recognition Activity modelling using 3D shape, Video summarization, shape-based activity models- Suspicious Activity Detection.

UNIT – V:

Human Face Recognition & Gait Analysis: Introduction: Overview of Recognition algorithms – Human Recognition using Face: Face Recognition from still images, Face Recognition from video, Evaluation of Face Recognition Technologies- Human Recognition using gait: HMM Framework for Gait Recognition, View Invariant Gait Recognition, Role of Shape and Dynamics in Gait Recognition

TEXT BOOKS:

1. A. Murat Tekalp, “Digital Video Processing”, second edition, Pearson, 2015.
2. Rama Chellappa, Amit K.Roy-Chowdhury, Kevin Zhou.S, “Recognition of Humans and their Activities using Video”, Morgan & Claypool Publishers, 2005.
3. Yunqian Ma, Gang Qian, “Intelligent Video Surveillance: Systems and Technology”, CRC Press (Taylor and Francis Group), 2009.

REFERENCE BOOKS:

1. Richard Szeliski, “Computer Vision: Algorithms and Applications”, Springer, 2011.
2. Yao Wang, Jorn Ostermann and Ya-Qin Zhang, “Video Processing and Communications”, Prentice Hall, 2001.
3. Thierry Bouwmans, Fatih Porikli, Benjamin Höferlin and Antoine Vacavant, “Background Modeling and Foreground Detection for Video Surveillance: Traditional and Recent Approaches, Implementations, Benchmarking and Evaluation”, CRC Press, Taylor and Francis Group, 2014.
4. Md. Atiqur Rahman Ahad, “Computer Vision and Action Recognition-A Guide for Image Processing and Computer Vision Community for Action Understanding”, Atlantis Press, 2011.

B.Tech. Data Science (Minor) VIII – Semester

L	T	P	C
0	0	0	2

M15802: MINI PROJECT

The Mini Project is in the collaboration with an industry of their specialization. Students will register for this immediately after VI - Semester examinations and pursue it during the summer vacation. The Mini Project shall be submitted in a report form and presented before the committee in VIII - Semester. It shall be evaluated for 100 external marks. The committee consists of an external examiner, Head of the Department, Supervisor of the Mini project and a senior faculty member of the department. There shall be no internal marks for the Mini Project.