

Syllabus for  
**Master of Science (M.Sc.) in Data Science**

For the students admitted from the academic year 2020-2021  
**(BASED ON CHOICE BASED CREDIT SYSTEM (CBCS))**

2020-2021

**DEPARTMENT OF DATA SCIENCE**

**ST. JOSEPH'S COLLEGE, (AUTONOMOUS)**

*Accredited at A++ Grade (4<sup>th</sup> Cycle) Special Heritage Status*

*Awarded by UGC, College with Potential for Excellence by*

*UGC DBT-STAR & DST-FIST Sponsored College*

*Tiruchirappalli - 620 002*

**M. Sc. Data Science  
Course Pattern – 2020**

Sem.	Code	Course	Hr.	Cr.
<b>I</b>	20PDS1101	Data Science and Big Data Analytics	5	4
	20PDS1102	Statistical Computing	5	4
	20PDS1103	Advanced Database Management Systems	4	4
	20PDS1104	Data Mining	4	4
	20PDS1105	Software Lab-I (R Programming)	4	2
	20PDS1106	Software Lab-II (Databases)	4	2
	20PDS1201A	Core Elective I: Design and Analysis of Algorithms (OR)	4	4
	20PDS1201B	Core Elective I: Applied Statistics		
	20PDS1401	Extra Credit Course-I: MOOC	-	(2)
<b>Total for Semester I</b>			<b>30</b>	<b>24 + (2)</b>
<b>II</b>	20PDS2107	Exploratory and Descriptive Data Analytics	5	4
	20PDS2108	Cloud Computing	5	4
	20PDS2109	Advanced Web Technology	4	4
	20PDS2110	Software Lab-III (Advanced Web Technology)	4	2
	20PDS2202A	Core Elective II: Cryptography and Network Security (OR)	4	4
	20PDS2202B	Core Elective II: WAP and XML		
	20PDS2203A	Core Elective III: Internet of Things (OR)	4	4
	20PDS2203B	Core Elective III: Object Oriented System Development		
	20PDS2111	Self- paced Learning: Advanced Python Programming	-	2
	20PDS2301	IDC: Multivariate Techniques for Data Analytics	4	4
<b>Total for Semester II</b>			<b>30</b>	<b>28</b>
<b>III</b>	20PDS3112	Digital Image Processing	5	4
	20PDS3113	Machine Learning	5	4
	20PDS3114	Time Series and Sampling Theory	4	4
	20PDS3115	Software Lab-IV (Image Processing)	4	2
	20PDS3116	Software Lab-V (Machine Learning)	4	2
	20PDS3117	Internship	-	2
	20PDS3204A	Core Elective IV: Software Project Management (OR)	4	4
	20PDS3204B	Core Elective IV: Natural Language Processing		
	20PDS3302	IDC: Sentiment Analysis	4	4
	20PDS3402	Extra Credit Course-II: MOOC	-	(2)
<b>Total for Semester III</b>			<b>30</b>	<b>26 +(2)</b>
<b>IV</b>	20PDS4118	Project Work	18	13
	20PDS4119	Comprehensive Examination	-	2
	20PDS4205A	Core Elective V: Soft Computing (OR)	4	4
	20PDS4205B	Core Elective V: Deep Learning		
	20PDS4206A	Core Elective VI: Data Science for Business (OR)	4	4
	20PDS4206B	Core Elective VI: Business Data Analytics		
	20PDS4303	IDC: Soft Skills	4	4
<b>Total for Semester IV</b>			<b>30</b>	<b>27</b>
<b>I-IV</b>	20PCW4501	Outreach Programme (SHEPHERD)	-	5
<b>Total for All Semesters</b>			<b>120</b>	<b>110 + (4)</b>

## DATA SCIENCE AND BIG DATA ANALYTICS

### Course Outcomes

On successful completion of the course, the students will be able to

1. Recall the basic and advanced methods of big data technology
2. Define the theories and methods used in data analytics
3. Demonstrate the tools, including Map Reduce and Hadoop and its ecosystem
4. Explain the core concepts of data science, data analytics algorithms and their usage
5. Apply the theories and methods to analyse data
6. Take part in data science and big data analytics projects

### Unit - I

11 hrs

**Introduction to Big Data Analytics:** Big Data Overview – Data Structures – Analyst Perspective on Data Repositories - State of the Practice in Analytics – BI Versus Data Science - Current Analytical Architecture – Drivers of Big Data – Big Data Ecosystem - Data Analytics Lifecycle – Data Discovery – Data Preparation – Model Planning – Model Building – Communicate Results – Operationalize.

### Unit - II

11 hrs

**Basic Data Analytic Methods Using R:** Introduction to R programming – R Graphical User Interfaces – Data Import and Export - Attribute and Data Types – Descriptive Statistics  
**Exploratory Data Analysis:** Visualization Before Analysis – Dirty Data – Visualizing a Single Variable – Examining Multiple Variables Data Exploration Versus Presentation.

### Unit – III

11 hrs

**Advanced Analytical Theory and Methods using R Programming:** Unsupervised Learning Algorithms – Use Cases – Overview – Determining number of groups – Diagnostics- Reasons to choose and cautions  
**Market Basket Analysis:** Algorithm for Rules Generation – Evaluation of Rules – Validation and Testing – Diagnostics.  
**Regression:** Use cases for regression analysis – Model Description – Diagnostics - Additional Regression Models

### Unit - IV

11 hrs

**Learning Analytics Using R Programming:** Supervised Learning – Learning Algorithms – Evaluating Learning algorithms – Implementation of Learning Algorithm– Naive Bayes – Bayes Theorem – Naïve Bayes Classifier – Smoothing – Diagnostics – Naïve Bayes in R – Diagnostics of Classifiers – Additional supervised learning Methods –  
**Text Analysis :** Text Analysis Steps – Example – Collecting – Representing Term Frequency – Categorizing – Determining Sentiments – Gaining Insights.

### Unit – V

11 hrs

**Advanced Analytics-Technology and Tools:** Map Reduce and Hadoop - Analytics for Unstructured Data. - Use Cases - MapReduce - Apache Hadoop – The Hadoop Ecosystem – pig – Hive – Hbase – Mahout – NoSQL. MADiib – Analytics Reports Consolidation – Communicating and operationalizing and Analytics Project – Creating the Final Deliverables -Developing Core Material for Multiple Audiences – Project Goals – Main Findings – Approach Model Description – Key points support with Data - Model details – Recommendations – Data Visualization.

**Text Book(s)**

1. EMC Education Services Data Science & Big Data Analytics: Discovering, Analysing, Visualizing and Presenting Data”, Wiley & Sons, Inc Publications, 2016. **(Unit 1,2)**
2. Raj Kamal, Preeti Saxena · BIG DATA ANALYTICS: Introduction to Hadoop, Spark, and Machine-Learning, McGraw-Hill Education, 2019. **(Unit 3,4,5)**

**Books for Reference**

1. Richard Hurley, "Data Science A Comprehensive Guide to Data Science, Data Analytics, Data Mining, Artificial Intelligence, Machine Learning, and Big Data", Atona Publications, 2020.
2. Thomas A. Runkler, "Data Analytics Models and Algorithms for Intelligent Data Analysis", Springer Vieweg, 2020.

## STATISTICAL COMPUTING

### Course Outcomes

On successful completion of the course, the students will be able to

1. Recall the basics of statistics and its usage in computing data
2. Define various Hypotheses tests and their implications
3. Demonstrate the applications of various correlation, regression methods and Probability concepts
4. Classify and model the sampling concepts
5. Apply correlation and regression techniques on datasets
6. Examine the probability distribution and analyze variables

#### Unit - I 11 hrs

**Correlation:** Definition of Correlation- Scatter Diagram- Karl Pearson's Coefficient of Linear Correlation- Coefficient of Correlation and Probable Error of  $r$ - Coefficient of Determination - Merits and Limitations of Coefficient of Correlation- Spearman's Rank Correlation.

#### Unit - II 11 hrs

**Regression Analysis:** Regression and Correlation (Intro) - Difference between Correlation and Regression Analysis- Linear Regression Equations -Least Square Method- Regression Lines- Properties of Regression Coefficients- Standard Error of Estimate

#### Unit - III 11 hrs

**Probability Distribution:** Probability Distribution and mathematical Expectation- Random Variable- Defined - Probability Distribution of a Random Variable- Expectation of Random Variable- Properties of Expected Value and Variance.

#### Unit - IV 11 hrs

**Sampling:** Sampling and Sampling Distributions - Data Collection- Sampling and Non-Sampling Errors – Principles of Sampling- Merits and Limitations of Sampling- Methods of Sampling- Parameter and Statistic- Sampling Distribution of a Statistic- Examples of Sampling Distributions- Standard Normal, Student's  $t$ , Chi-Square ( $\chi^2$ ) and Snedecor's  $F$ -Distribution.

#### Unit – V 11 hrs

**Statistical Inference:** Estimation and Testing of Hypothesis - Statistical Inference- Estimation- Point and interval- Confidence interval using normal,  $t$  and  $\chi^2$  Distributions- Testing of Hypothesis- Significance of a mean - Using  $t$  Distribution.

### Text Book

1. Cheryl Ann Willard, "Statistical Methods An Introduction to Basic Statistical Concepts and Analysis", Taylor & Francis Publication, 2020, ISBN:9780429523151, 0429523157

### Books for Reference

1. Debbie L. Hahs-Vaughn, Richard G. Lomax, "Statistical Concepts - A First Course", Taylor & Francis Publication, 2020.
2. Jay Lehmann, "A Pathway to Introductory Statistics", Pearson Publications, 2020.

## ADVANCED DATABASE MANAGEMENT SYSTEMS

### Course Outcomes

On successful completion of the course, the students will be able to

1. Recall how a database is designed and differentiated
2. Define Database Models and Database Management Skills
3. Explain the applications of Database Models and Emerging Trends
4. Infer ideas on how to design databases and classify them
5. Experiment with various Data models and Database Architectures
6. Distinguish Object Oriented Databases and Advanced databases

### Unit - I

11 hrs

**Relational and parallel Database Design:** Basics - Entity Types - Relationship Types - ER Model - ER-to- Relational Mapping algorithm. **Normalization:** Functional Dependency - 1NF, 2NF, 3NF, BCNF, 4NF and 5NF. Architecture - I/O Parallelism - Interquery Parallelism - Intraquery Parallelism - Intraoperation Parallelism -Interoperation Parallelism.

### Unit - II

11 hrs

**Distributed and Object based Database:** Architecture - distributed data storage - Distributed transactions - Commit protocols - Concurrency control - Query Processing - Complex Data Types - Structured Types and Inheritance - Table Inheritance - array and Multiset - Object Identity and Reference Types - Object Oriented versus Object Relational.

### Unit - III

11 hrs

**Spatial Database:** Spatial Database Characteristics - Spatial Data Model - Spatial Database Queries - Techniques of Spatial Database Query. **Logic based Database:** Overview - Propositional Calculus - Predicate Calculus - Deductive Database Systems - Recursive Query Processing.

### Unit - IV

11 hrs

**XML Database:** XML Hierarchical data model - XML Documents – DTD - XML Schema - XML Querying – XHTML - Illustrative Experiments.

### Unit – V

11 hrs

**Temporal Database:** Introduction – Intervals - Packing and Unpacking Relations - Generalizing the relational Operators - Database Design, Integrity Constraints. **Multimedia Databases:** Multimedia Sources - Multimedia Database Queries - Multimedia Database Applications.

### Text Book(s)

1. Abraham Silberschatz, Henry F Korth , S Sudarshan, “Database System Concepts”, 6th edition , McGraw-Hill International Edition , 2015

## **Books for Reference**

1. Chhanda Ray, "Advanced Database System", Independently Published, 2020
2. Wilfried Lemahieu, Seppe vanden Broucke, Bart Baesens, "Principles of Database Management: The Practical Guide to Storing, Managing and Analyzing Big and Small Data", Cambridge University Press, 2018.

## DATA MINING

### Course Outcomes

On successful completion of the course, the students will be able to

1. Recall the fundamental concepts and techniques of data mining
2. Define the Concepts and design of data warehousing
3. Demonstrate the functionality of data warehouse architecture
4. Explain the ideology behind data warehouse development
5. Solve real world problems by deploying data mining concepts
6. Analyse data mining concepts, functionalities in the problem-solving process

### Unit - I

11 hrs

**Introduction:** Data mining concepts - Data Warehouses - Data Mining Functionalities - Basic Data mining tasks - Data Mining Issues - Social Implications of Data Mining - Applications and Trends in Data Mining.

### Unit - II

11 hrs

**Data Pre-processing:** Need for pre-processing -Data Cleaning - Data Integration and Transformation - Data Reduction - Data Cube Aggregation - Attribute Subset Selection

### Unit - III

11 hrs

**Data Mining Techniques:** Association Rule Mining – The Apriori Algorithm -  
**Classification and Prediction:** Issues regarding Classification and Prediction – Decision Tree Induction-**Clustering Techniques:** cluster Analysis – Clustering Methods- K-means Clustering

### Unit - IV

11 hrs

**Data Warehousing:** An introduction - characteristics of a data warehouse - Data marts - other aspects of data mart. Online analytical processing: OLTP & OLAP systems.

### Unit – V

11 hrs

**Developing a Data Warehouse:** Need and process of building a data warehouse - Data warehouse architectural strategies and organizational issues - Design consideration - Data content - meta data - distribution of data - tools for data warehousing - Performance Considerations

### Text Book(s)

1. Tan, Pang-Ning. “Introduction to Data Mining”, Pearson Publication, 2018.

### Books for Reference

1. Meira, Jr, Wagner, and Zaki, Mohammed J. “Data Mining and Machine Learning: Fundamental Concepts and Algorithms”,Cambridge University Press, 2020.
2. Jianxin Li, Sen Wang, Shaowen Qin, Shuliang Wang, Xue Li, “Advanced Data Mining and Applications”, Springer International Publishing, 2019.



**Software Lab-I**  
**R PROGRAMMING**

**Course Outcomes**

On successful completion of the course, the students will be able to

1. Show how to transform the raw data into cleansed dataset by employing pre-processing techniques
2. Illustrate how to link data, statistical methods, and actionable questions
3. Demonstrate the features and constructs of R language
4. Interpret the visualization results of analytics effectively
5. Apply using different analytical approaches and techniques in R to disentangle real time analytics problems
6. Analyse the Machine Learning methods and visualize data

**Lab Exercises:**

1. Program for basics and Looping in R
  - a) Data types and data structures
  - b) Flow control and looping
2. Functions in R
  - a) Writing and calling functions
  - b) Top-down design, testing
  - c) Debugging and functions as objects
3. Pre-processing and preparing data
  - a) Raw data
  - b) Clean data
  - c) Tidy data
4. Exploratory data analytics
  - a) Descriptive Statistics
  - b) Hypothesis testing
5. Statistical and Machine learning methods
  - a) Linear Regression
  - b) Logistic Regression
6. Data Visualization in R
  - a) Graphs
  - b) Charts
7. Advance Data Visualization in R
  - a) Lattice
  - b) ggplot2
8. Case Studies for Industrial Domain
  - a) Tax data
  - b) Automotive data
9. Case Study for Time Series Analysis
  - a) Stock market data
10. Case Studies for generating graphical report
  - a) Employment data
  - b) Sports data, etc.

**Software Lab-II**  
**DATABASES**

**Course Outcomes**

On successful completion of the course, the students will be able to

1. Recall advanced commands and handle bulk load of data
2. Distinguish the SQL and NoSQL databases and to perform database operations
3. Demonstrate advanced Database tools and techniques
4. Show how to write Code using tools like Neo4J and Cypher
5. Develop efficient programs to perform database operations
6. Apply various DML and DDL statements on Databases

**Lab Exercises:**

1. Normalization: 1NF, 2NF, 3NF
2. DDL Statements
3. Basic DML: Selection, Filtering and Join
4. Sorting and Handling Null Values
5. Built-in Functions
6. Advanced Insert Commands and Bulk Load of Data
7. Creating Stored Procedures
8. Creating Triggers
9. Neo4J and Basics of CYPHER
10. CYPHER for Pattern Matching
11. Simple nosql program to be included

**Core Elective-I**  
**DESIGN AND ANALYSIS OF ALGORITHMS**

**Course Outcomes**

On successful completion of the course, the students will be able to

1. Choose the algorithmic procedure to determine the computational complexity of algorithms
2. Define the stepwise procedure to tackle problems
3. Demonstrate effective analytical skills in computing applications using algorithms
4. Show how to break down problems into small pieces for program development
5. Develop efficient approaches for Computational Modeling
6. Analyze the complexity of algorithms

**Unit - I** **11 hrs**

**Introduction:** Algorithm Definition – Algorithm Specification –Asymptotic Notations. Elementary Data Structures: Stacks and Queues – Trees – Graphs

**Unit - II** **11 hrs**

**Divide and Conquer:** Binary Search – Finding the Maximum And Minimum – Merge Sort – Quick Sort-Selection.

**Unit - III** **11 hrs**

**The Greedy Method:** Knapsack Problem - Job Sequencing with Deadlines - Minimum Cost Spanning Trees - Optimal Storage on Tapes - Single Source Shortest Paths.

**Unit - IV** **11 hrs**

**Dynamic Programming:** Multistage Graphs – All-Pairs Shortest Paths – Optimal Binary Search Trees - Reliability Design - The Traveling Salesperson Problem - Flow Shop Scheduling.

**Unit – V** **11 hrs**

**Basic Traversal and Search Techniques:** Techniques for Binary Trees – Techniques for Graphs – Connected Components and Spanning Trees. **Backtracking:** The 8-Queens Problem – Sum of Subsets – Graph Coloring

**Text Book(s)**

1. Robert Karamagi, "Design and Analysis of Algorithms",Lulu.com Publisher,2020

**Books for Reference**

1. Tim Roughgarden, "Beyond the Worst-Case Analysis of Algorithms", Cambridge University Press, 2021
2. Sandeep Sen, Amit Kumar, "Design and Analysis of Algorithms A Contemporary Perspective", Cambridge University Press, 2019

Core Elective-I  
APPLIED STATISTICS

Course Outcomes

On successful completion of the course, the students will be able to

1. Recall the core ideologies of statistics and its methodologies
2. Define the hypothesis of events and results
3. Explain the sampling theories and tests of significance
4. Extend knowledge on the Classification and correlations strategies
5. Identify data correlation and apply mathematical approaches on them
6. Classify data and identify the distribution pattern

Unit - I

11 hrs

**Introduction:** Meaning - Definition - Statistics as a data - Statistics as a Method. Importance - Functions - and Limitations of Statistics in Data Science. Finite and Infinite population - Hypothetical and existent population - census method - sample method - Random sampling - Statistical Sampling - Systematic Sampling - Clustering Sampling - Judgment Sampling - Quota Sampling - Convenience or Churk Sampling - Statistical Errors - Absolute Error - Relative error - Reducing Sample Error - Test of Reliability Error.

Unit - II

11 hrs

**Classification and Tabulation:** Overview of Classification - Statistical Series - Types of Series - Frequency Distribution - Continuous or Grouped Frequency Distribution. Magnitude of Class intervals - Cumulative Frequency Distribution - Two Way Frequency Distribution. **Measures of Central Tendency:** Arithmetic Mean - Geometric Mean - Harmonic Mean - Median - Mode. **Dispersion:** Overview - Mean Deviation - Standard Deviation - Combined Standard Deviation.

Unit - III

11 hrs

**Correlation and Association:** Introduction - Classification - Correlation and Association - Types of Association - Comparison of Observed and Expected Frequencies - Yule's Coefficient of Association - Yule's Coefficient of Colligation - Pearson's Coefficient of Contingency Partial Association.

Unit - IV

11 hrs

**Probability Distribution:** Discrete - Binominal Distribution - Poison Distribution - Negative binomial - geometric - Hyper geometric. Continuous - Uniform - Normal Distribution- Skewness and Kurtosis-central limit theorem - t-distribution - Gamma Distribution - Exponential Distribution - Laplace distribution - Hazard rate function - Chi-Square Distribution - Weibull Distribution - Beta Distribution - Dirichlet - F-distribution

Unit – V

11 hrs

**Sampling Theory and test of significance:** Errors and significance-Null Hypothesis- Type I / Type II errors- The meaning of p value – power - Confidence intervals- One-tailed vs. two-tailed tests- Parametric vs. Non-parametric methods- Tests of Variances and Goodness-of-fit- One-sample Variance Test- Two-sample Variance Test

**Text Book(s)**

1. R.S.N. Pillai, Bagavathi, “Statistics Theory and Practice”, S.Chand& Company,2013. (Unit 1,2)
2. Douglas C. Montgomery, George C. Runger., “Applied Statistics for Engineers”, John Wiley & Sons. Inc,2003. (Unit 3)
3. Maureen Hillenmeyer , “Applied Statistics”, Stanford University, June 2005.(Unit 4,5)

**Books for Reference**

1. Sorana D. Bolboacă, Applied and Computational Statistics, MDPI AG Publisher, 2020
2. Dieter Rasch, Rob Verdooren, Jürgen Pilz, "Applied Statistics: Theory and Problem Solutions with R", Wiley Publisher, 2019.

## EXPLORATORY AND DESCRIPTIVE DATA ANALYTICS

### Course Outcomes

On successful completion of the course, the students will be able to

1. Define exploratory analytics and descriptive analytics;
2. Recall various data types and data standardization
3. Explain the nuances in summarizing the results obtained from data
4. Show the relationships between variables
5. Apply the knowledge on exploratory and descriptive analytics to few cases
6. Analyse various application domain data through the knowledge obtained

### Unit - I

11 hrs

Introduction - Data Analytics – Exploratory Data Analytics (EDA) - Need for EDA - EDA - Objectives - Google Trend analysis - Explore Trends R Visualization - Packages - Lattice - understanding plots - aesthetics -- Statistical function – Histogram Box Plot - Density Plot - Scatter Plots - Summarizing Data in R

### Unit - II

11 hrs

Variable Analysis - One variable - Understanding outliers through - histogram, boxplot, density plot - dataset - pseudo dataset of facebook - Exploring two variables - Understanding Variables and relationships - correlations - condition means - Explore multivariate variables - Explore Diamond dataset for price prediction

### Unit - III

11 hrs

Data types - Categorical - Binary - ordinal - Nominal - Continuous - Discrete - Data dimensions - Univariate - bivariate - multivariate - Numerical Measures - Central Tendency Mean - Median - Mode - Understanding data using central tendency - plotting histogram density plots and inference of plot - Variability Measure - Variance -Range - IQC -and standard Deviation - Sum of squares - Squared Deviations - Absolute Deviations - Identify outlier using Inter Quartile Range - Visualization using boxplot

### Unit - IV

11 hrs

Data standardization -Z Score - Negative Z Score - Continuous Distributions - Compute proportions - Relative Frequency histogram -Normalized Distribution using Ztable - Probability Distributions - Probability of mean - location of mean distribution - Sampling distributions - Klout Sampling Distribution -Understanding Shape of Distribution - Standard Error -Standard Deviation of sampling distribution -Ratio of Sampling Distribution - Central Limit Theorem R -Mean of sample means Advanced Analytics

### Unit – V

11 hrs

Case Study -EDA analytics on dataset Movies -Social network using R -Prediction of Movie ratings - Descriptive Analytics on Movie Dataset

### Text Book(s)

1. Ahmed, Usman, and Mukhiya, Suresh Kumar. Hands-On Exploratory Data Analysis with Python: Perform EDA Techniques to Understand, Summarize, and Investigate Your Data. United Kingdom, Packt Publishing, 2020. (Unit 1,5)

2. Pearson, Ronald K.. Exploratory Data Analysis Using R. United States, CRC Press, 2018. (Unit 2,3,4)

**Books for Reference**

1. Salmaso, Luigi, et al. End-to-end Data Analytics for Product Development: A Practical Guide for Fast Consumer Goods Companies, Chemical Industry and Processing Tools Manufacturers. United Kingdom, Wiley, 2020.
2. Cirillo, Andrea. Exploratory Data Analysis with R. United States, Packt Publishing, 2020.

## CLOUD COMPUTING

### Course Outcomes

On successful completion of the course, the students will be able to

1. Recall the basic concepts of cloud computing technology
2. Show a comprehensive understanding of cloud technologies and architectures
3. Summarize micro services and develop server less architectures
4. Explain various infrastructure services, micro services offered in cloud
5. Identify cloud security considerations and models
6. Categorize the level of security requirements in the cloud

#### **Unit - I** **11 hrs**

Evolution of Cloud Computing -Essential Characteristics of cloud computing - Operational models such as private, dedicated, virtual private, community, hybrid and public cloud - Service models such as IaaS, PaaS and SaaS - Governance and Change Management - Business drivers, metrics and typical use cases. Example cloud vendors - Google cloud platform, Amazon AWS, Microsoft Azure, Pivotal cloud foundry and Open Stack.

#### **Unit - II** **11 hrs**

Basics of Virtual Machines - Taxonomy of Virtual Machines. Virtualization Architectures. Challenges with Dynamic Infrastructure - Principles of Infrastructure as Code - Considerations for Infrastructure Services and Tools - Monitoring: Alerting, Metrics, and Logging - Service Discovery - Server Provisioning via Templates - Patterns and Practices for Continuous Deployment - Organizing Infrastructure and Testing Infrastructure - Change Management Pipelines for Infrastructure.

#### **Unit - III** **11 hrs**

Cloud Native Design and Microservices- Containerized - Dynamically orchestrated design - Continuous delivery - Support for a variety of client devices - Monolithic vs Microservices Architecture - Characteristics of microservice architecture - 12 factor application design - Considering service granularity - Scalable Services - Sharing dependencies between microservices - Stateless versus Stateful microservices - Service discovery - Service Registry- Performance Considerations.

#### **Unit - IV** **11 hrs**

Function as a Service (FaaS) - Backend as a Service (BaaS) - Advantages of serverless architectures - Taking a hybrid approach to serverless architecture - Function deployment and Function invocation. Introduction to DevOps - The Deployment Pipeline - The Overall Architecture - Building and Testing - Deployment - Crosscutting Concerns such as Monitoring, Scalability, Repeatability, Reliability, Recoverability, Interoperability, Testability, and Modifiability

#### **Unit - V** **11 hrs**

Security Considerations - STRIDE Threat Model - Cloud Security Challenges - Cloud specific Cryptographic Techniques - CIA Triad - Security by Design - Common Security Risks - Risk Management Security Monitoring - Security Architecture Design - Data Security - Application Security Virtual Machine Security.



**Text Book(s)**

1. Buyya, Vecciola and Selvi, “Mastering Cloud Computing: Foundations and Applications Programming”, Tata McGraw Hill, 2013.(Unit 1,2)
2. ArshdeepBahga, Vijay Madiseti, “Cloud Computing: A Hands – On Approach”, Universities press (India) Pvt. limited 2016.(Unit 3,4,5)

**Books for References**

1. Dinesh G. Harkut, Kashmira Kasat, Saurabh Shah "Cloud Computing: Technology and Practices",IntechOpen Publisher, 2019.
2. Scott Goessling , Kevin L. Jackson, “Architecting Cloud Computing Solutions”, Packt Publishing, 2018.

**ADVANCED WEB TECHNOLOGY**

**Course Outcomes**

On successful completion of the course, the students will be able to

1. Define Django framework
2. Relate the forms with files and databases
3. Demonstrate Django Handlers, CSS and JavaScript integration
4. Explain the process of Web Application development
5. Develop secured Web Application
6. Build web forms for websites

**Unit - I** **11 hrs**

Introduction to Django Framework - Django Framework components - Installing Python and Django - Setting Up Django Projects - Working with Django Models - Models and fields - Import CSV Data - Django Admin - Query with Django ORM.

**Unit - II** **11 hrs**

Building URL Handlers and Views - URL Patterns - Implement Views - Build Django Templates - Structure Templates - Integrate CSS and Javascript.

**Unit - III** **11 hrs**

Using Django Forms - Making forms - Form Fields - Submitting forms - Using submitted data - Advanced Django Forms - models - widgets - forms and files - using multiple forms on a page

**Unit - IV** **11 hrs**

Styling and customizing Form Appearance - validation and errors - server based errors - form rendering - using CSS - form styling - Homepage styling - Building a REST API - creating class based views - creating switchboard views - using Django REST Framework - Authentication with DRF - Optimizing the pip environment

**Unit – V** **11 hrs**

Unit Testing - Testing models - testing views - Securing Django APIs - securing the Django Admin - planning for failure - ScraPy for Scrapping Websites

**Text Book(s)**

1. Nigel George , “Build Your First Website with Django 2.1: Master the Basics of Django While Building a Fully-Functioning Website”, Paperback, 2018.

**Books for Reference**

1. Antonio Mele , “Django 2 by Example”, Paperback, 2018.
2. Harry J.w Percival , “Test–Driven Development with Python 2e”, Paperback, 2017.

**Software Lab-III**  
**ADVANCED WEB TECHNOLOGY**

**Course Outcomes**

On successful completion of this course, the students will be able to

1. Show the usage of Django frameworks
2. Illustrate a clear picture of advanced web technology and how to deploy it in the design and development of web applications
3. Interpret the advanced web technologies using Python and Django
4. Outline the new ways to design using Django templates and forms
5. Apply the advanced and customized form designs
6. Analyse the tools and techniques for web scrapping in Django framework using ScraPy

**Lab Exercises:**

1. Program for Setting Up Django Projects
2. Program for working with Django Models and Import Data
3. Program for building URL Handlers and Views
4. Program for building Django Templates
5. Program for using Django Forms
6. Program for implementing Advanced Django Forms
7. Program to work with styling and customizing of Form Appearance
8. Program for building a REST API
9. Program for Unit Testing and Securing Django APIs
10. Program for Web Scrapping using ScraPy

**Core Elective II**  
**CRYPTOGRAPHY AND NETWORK SECURITY**

**Course Outcomes**

On successful completion of the course the students will be able to

1. Recall the fundamentals of networks security, security architecture, threats and vulnerabilities
2. Define Cryptography Theories, Algorithms and Systems
3. Explain various Security practices and System security standards
4. Demonstrate various Authentication schemes to simulate different applications
5. Apply diverse cryptographic operations of cryptographic algorithms and identify the malicious software & firewalls
6. Analyze essential approaches and techniques to build protection mechanisms to secure computer networks

**Unit - I** **11 hrs**

Introduction - Security trends – Legal, Ethical and Professional Aspects of Security, Need for Security at Multiple levels, Security Policies – Model of network security – Security attacks, services and mechanisms OSI security architecture – Classical encryption techniques: substitution techniques, transposition techniques, steganography- Foundations of modern cryptography: perfect security – information theory – product cryptosystem – cryptanalysis.

**Unit - II** **11 hrs**

Symmetric Encryption and Message Confidentiality - Symmetric Encryption Principles, Symmetric Block Encryption Algorithms, Stream Ciphers and RC4, Cipher Block Modes of Operation, Location of Encryption Devices, Key Distribution. Public-key Cryptography and Message Authentication: Approaches to Message Authentication, Secure Hash Functions and HMAC, Public-Key Cryptography Principles, Public-Key Cryptography Algorithms, Digital Signatures, Key Management.

**Unit - III** **11 hrs**

Authentication Applications - Kerberos, x.509 Authentication Service, Public-Key Infrastructure. Electronic Mail Security: Pretty Good Privacy (PGP), S/MIME.

**Unit - IV** **11 hrs**

IP Security - IP Security Over view, IP Security Architecture, Authentication Header, Encapsulating Security Payload, Combining Security Associations. Web Security: Web Security Considerations, Secure Socket Layer (SSL) and Transport Layer Security (TLS), Secure Electronic Transaction (SET). Network Management Security: Basic Concepts of SNMP, SNMPv1 Community Facility, SNMPv3.

**Unit - V** **11 hrs**

Intruders - Intruders, Intrusion Detection, Password Management. Malicious Software: Virus and Related Threats, Virus Countermeasures, Distributed Denial of Service Attacks. Firewalls: Firewall Design Principles, Trusted Systems, Common Criteria for Information Technology Security Evaluation.

**Text Book(s)**

1. Stallings William, “Cryptography and Network Security - Principles and Practice”, 2017.

**Books for Reference**

1. Bhavesh B. Prajapati, Nirbhay Kumar Chaubey,"Quantum Cryptography and the Future of Cyber Security", IGI Global, 2020.
2. Atul Kahate, "Cryptography and Network Security",4e, McGraw-Hill Education, 2019.

**Core Elective II  
WAP and XML**

**Course Outcomes**

On successful completion of the course, the students will be able to

1. Recall the concepts and procedures of WAP and XML
2. Define eXtensible Markup Language (XML)
3. Explain secure messaging through web services
4. Build efficient documents and display them in an organized way
5. Make use of XML concepts to develop Web applications
6. Develop SOA application by deploying XML and Web Services

**Unit - I** **11 hrs**

**Overview of WAP:** WAP and the wireless world – WAP application architecture – WAP internal structure-WAP versus the Web – WAP 1.2 – WTA and push features. Setting up WAP: Available software products – WAP resources – The Development Toolkits.

**Unit - II** **11 hrs**

**WAP gateways:** Definition – Functionality of a WAP gateway – The Web model versus the WAP model Positioning of a WAP gateway in the network – Selecting a WAP gateway Basic WML: Extensible markup language – WML structure – A basic WML card – Text formatting – navigation – Advanced display features.

**Unit - III** **11 hrs**

**Interacting with the user:** Making a selection – Events – Variables – Input and parameter passing. **WML Script:** Need for WML script – Lexical Structure – Variables and literals – Operators – Automatic data type conversion – Control Constructs Functions – Using the standard libraries – programs – Dealing with Errors.

**Unit - IV** **11 hrs**

**XML: Introduction XML:** An Eagle's Eye view of XML – XML Definition - List of an XML Document – Related Technologies – An introduction to XML Applications – XML Applications – XML for XML – First XML Documents Structuring Data: Examining the Data XMLizing the data The advantages of the XML format – Preparing a style sheet for Document Display.

**Unit – V** **11 hrs**

**Attributes, Empty Tags and XSL:** Attributes – Attributes Versus Elements – Empty Tags – XSL – Well-formed XML documents – Foreign Languages and Non-Roman Text – Non Roman Scripts on the Web Scripts, Character sets, Fonts and Glyphs – Legacy character sets– The Unicode Character set – Procedure to Write XML Unicode.

**Text Book(s)**

1. Charles Arehart and Others. "Professional WAP with WML, WML script, ASP, JSP, XML, XSLT, WTA Push and Voice XML" , Shroff Publishers and Distributors Pvt. Ltd, 2014.

**Books for Reference**

1. Heather Williamson, "XML: The Complete Reference" , Tata McGraw-Hill Education India, 2001

**Core Elective III**  
**INTERNET OF THINGS**

**Course Outcomes**

On successful completion of the course, the students will be able to

1. Recall the basics of Internet of things and protocols
2. Define building blocks of Internet of Things and its characteristics
3. Summarize some of the application areas where Internet of Things can be applied
4. Explain the gist of the middleware for Internet of Things
5. Identify the application areas of IOT
6. Distinguish between Web of things and Internet of things

**Unit - I** **11 hrs**  
Definitions and Functional Requirements - Motivation - Architecture - Web 3.0 View of IoT - Ubiquitous IoT Applications - Four Pillars of IoT - DNA of IoT -The Toolkit Approach for End-user Participation in the Internet of Things. Middleware for IoT: Overview - Communication middleware for IoT - IoT Information Security

**Unit - II** **11 hrs**  
Protocol Standardization for IoT - Efforts - M2M and WSN Protocols - SCADA and RFID Protocols- Issues with IoT Standardization - Unified Data Standards -Protocols -IEEE 802.15.4 - BACNet Protocol Modbus - KNX - Zigbee-Network layer - APS layer -Security

**Unit - III** **11 hrs**  
Web of Things versus Internet of Things - Two Pillars of the Web - Architecture standardization for WoT Platform Middleware for WoT - Unified Multitier WoT Architecture - WoT Portals and Business Intelligence. Cloud of Things: Grid/SOA and Cloud Computing - Cloud Middleware - Cloud Standards - Cloud Providers and Systems - Mobile cloud Computing - The Cloud of Things Architecture.

**Unit - IV** **11 hrs**  
Integrated Billing Solutions in the Internet of Things Business Models for the Internet of things - Network Dynamics: Population Models - Information Cascades - Network Effects - Network Dynamics: Structural Models - Cascading Behavior in Networks - The Small-World phenomenon

**Unit – V** **11 hrs**  
The Role of the Internet of Things for Increased Autonomy and Agility in Collaborative Production Environments -Resource Management in the Internet of Things: Clustering, Synchronization and Software Agents. Applications -Smart Grid -Electrical Vehicle charging

**Text Book(s)**

1. Arshdeep Bahga, Vijay Madiseti, “Internet of Things – A hands- on approach”, Universities Press, 2015.

## **Books for Reference**

1. Carolina Fortuna, John Davies, "The Internet of Things: From Data to Insight", Wiley Publication, 2020.
2. Al-Sakib Khan Pathan, Danda B Rawat, Raja Datta, Uttam Ghosh, "Internet of Things and Secure Smart Environments Successes and Pitfalls", CRC Press, 2020



**Core Elective III**  
**OBJECT ORIENTED SYSTEM DEVELOPMENT**

**Course Outcomes**

On successful completion of the course the students will be able to

1. Recall the concept of Object-oriented design and understand the fundamentals of OOSD life cycle and familiarize with evolution of object-oriented model, classes and its notations
2. Define UML in order to express the design of software projects
3. Demonstrate how the object-oriented approach differs from the traditional approach in system analysis and design
4. Explain the importance of modeling and how the Unified Modeling Language (UML) represents an object-oriented system using a number of modeling views
5. Apply OOSD concepts to real world problems
6. Analyze, design, document the requirements through use case driven approach

**Unit - I**

**11 hrs**

Fundamentals of OOSD - Overview of Object-Oriented Systems Development: Two orthogonal view of the software - OOSD methodology – Need for an object orientation. Object basics: Object Oriented Philosophy- Objects – Attributes – Object respond to messages – Encapsulation and information hiding – class hierarchy Polymorphism – Object relationship and associations. OOSD life cycle: Software development process – OOSD Use case Driven Approach – Reusability.

**Unit - II**

**11 hrs**

Methodology, Modeling and UML - Object Oriented Methodologies: Rumbaugh et al.'s object modeling technique – The Booch methodology – The Jacobson et al. methodology – Patterns – Frameworks - The Unified approach. Unified Modeling Language : Static and dynamic models – Need for modeling - UML diagrams – UML class diagram – Use case diagram - UML dynamic modeling – packages and model organization.

**Unit - III**

**11 hrs**

Object Oriented Analysis - Object Oriented Analysis process: Business Object Analysis - Use case driven object oriented analysis – Business process modeling – Use-Case model – Developing effective documentation. Classification: Classifications theory – Approaches for identifying classes – Noun phrase approach – Common class patterns approach – Use-Case Driven approach – Classes, Responsibilities, and Collaborators - Naming classes - Identifying object relationships, attributes, and methods: Association – Super- Sub class relationship – Aggregation – Class responsibility – Object responsibility.

**Unit - IV**

**11 hrs**

Object Oriented Design - Object Oriented Design Process and Design Axioms - OOD process- OOD axioms – Corollaries – Design patterns. Designing classes: Designing classes – Class visibility – Refining attributes – Designing methods and protocols – Packages and managing classes. Access layer: Object Store and persistence – DBMS Logical and physical Database Organization and access control – Distributed Databases and Client Server Computing — Multidatabase Systems – Designing Access layer classes. View Layer: Designing view layer classes – Macro level process – Micro level process – The purpose of view layer interface – Prototyping the user interface.

**Unit – V**

**11 hrs**

Software Quality - Software Quality Assurance: Quality assurance tests – Testing strategies – Impact of Object Orientation on Testing - Test Cases- Test Plan – Continuous testing.

System Usability and Measuring User satisfaction: Usability Testing – User satisfaction test  
- A tool for analyzing user satisfaction. System Usability and Measuring User satisfaction:  
Introduction – Usability Testing.

**Text Book(s)**

1. Ali Bahrami, “Object Oriented Systems Development using UML”, McGraw-Hill, 2008.

**Books for Reference**

1. Alan Dennis, Barbara Haley Wixom, David Paul Tegarden,"System Analysis & Design, an Object-oriented Approach with UML, Wiley Publications, 2020.
2. Zeynep Altan,"Applications and Approaches to Object-Oriented Software Design: Emerging Research and Opportunities",IGI Global, 2019.

**Self- Paced Learning:**  
**ADVANCED PYTHON PROGRAMMING**

**Course Outcomes**

On successful completion of the course, the students will be able to

1. Recall the advanced features of Python.
2. Define the fundamentals of the most widely used Python packages.
3. Make use of the NumPy, Pandas and Matplotlib to perform data analysis and data Visualization.
4. Analyse data and perform data aggregation with Python.
5. Examine various Customization techniques in Matplotlib.

**Unit – 1**

**Introduction to NumPy:** Understanding Data Types in Python - The Basics of NumPy Arrays - Computation of NumPy Arrays - Aggregations - Comparisons - Masks - Boolean Logic - Fancy Indexing - Sorting Arrays - Structured Data.

**Unit – 2**

**Data Manipulation with Pandas:** Installing and Using Pandas - Introducing Pandas Objects - Data Indexing and Selection - operating on Data in Pandas - Handling Missing Data - Hierarchical Indexing - Combining Datasets.

**Unit – 3**

**Data Aggregation with Pandas:** Aggregate and Grouping - Pivot Tables - Vectorized String Operations - Working with Time Series - High-Performance Pandas.

**Unit – 4**

**Visualization with Matplotlib:** Simple Line Plots - Simple Scatter Plot - Visualizing Errors - Density and Contour Plots - Histograms - Binnings and Density.

**Unit – 5**

**Customization with Matplotlib:** Customizing Plot Legends - Customizing Colorbars - Multiple Subplots - Text and Annotations - Customizing Ticks - Configuration and Stylesheets - Three Dimensional Plotting - Geographic Data with Basemap and Visualisation with Seaborn.

**Text Book**

1. Jake Vander Plas, “Python Data Science Handbook Essential Tools for Working with Data”, O'Reilly Media, 1st edition, 2016.

**Book for Reference**

1. Karamagi, Robert Method, " Advanced Python Programming", Independently Published, 2020.
2. Wilkes, Matthew, "Advanced Python Development: Using Powerful Language Features in Real-World Applications", Apress Publication, 2020.

**IDC: MULTIVARIATE TECHNIQUES FOR DATA ANALYTICS**

**Course Outcomes**

On successful completion of the course, the students will be able to

1. Recall the meaning, objectives and applications of various types of analysis
2. Demonstrate factoring and clustering techniques
3. Explain discriminant and principal component analysis on data
4. Identify the suitability of analysis techniques and models
5. Apply appropriate multivariate techniques on data
6. Analyse the behaviour of data and interpret the results of simulation studies

**Unit - I**

**11 hrs**

**Introduction to Multivariate Techniques:** Measurement Scales (Metric and Non-metric Measurement Scales) - Classification of Multivariate Techniques (Dependence and Inter-dependence Techniques) - Applications of Multivariate Techniques in different disciplines.

**Unit - II**

**11 hrs**

**Factor Analysis:** Introduction to Factor Analysis - Meaning, Objectives and Assumptions - Designing a Factor Analysis Study - Deriving Factors - Assessing Overall Factors - Validation of Factor Analysis.

**Unit - III**

**11 hrs**

**Cluster Analysis:** Introduction to Cluster Analysis - Objectives and Assumptions - Research Design in Cluster Analysis - Hierarchical and Non-hierarchical Methods - Interpretation of Clusters - Validation of Profiling of Clusters.

**Unit - IV**

**11 hrs**

**Discriminant Analysis:** Introduction to Discriminant Analysis - Concepts, Objectives and Applications - Procedure for conducting Discriminant Analysis - Stepwise Discriminant Analysis - Mahalanobis Procedure - Logit Model.

**Unit – V**

**11 hrs**

**Principal Component Analysis:** Dimensionality Reduction - Deriving Orthogonal Projections - Lower Dimensional Subspaces - Characterization through Singular Value Decomposition and Eigenvalue Analysis - Rayleigh Quotient - Kernel PCA - Functional PCA.

**Text Book(s)**

1. Joseph F Hair, William C Black, “Multivariate Data Analysis”, Pearson Education, 7th edition, 2013.

**Books for Reference**

1. Warner, Rebecca M, "Applied Statistics II: Multivariable and Multivariate Techniques, SAGE Publications, 2020.
2. Hanif, Muhammad, Shahbaz, Muhammad Qaiser, Shahbaz, Saman Hanif, "Multivariate Techniques: An Example Based Approach”, Cambridge Scholars Publishing, 2019.

## DIGITAL IMAGE PROCESSING

### Course Outcomes

On successful completion of this course, the students will be able to

1. Recall the concepts, methods and algorithms of digital image processing
2. Define the ways of image transformation, image enhancement, image restoration, image compression techniques
3. Summarize the image reconstruction from projections
4. Interpret Image compression standards, image segmentation and representation techniques.
5. Examine the techniques for image enhancement and image restoration
6. Apply the fundamental concepts of a digital image processing system and analyze images in the frequency domain using various transforms.

### Unit - I

11 hrs

**Continuous and Discrete Images and Systems:** Light - Luminance - Brightness and Contrast - Eye - The Monochrome Vision Model - Image Processing Problems and Applications - Vision Camera - Digital Processing System - 2-D Sampling Theory - Aliasing - Image Quantization - Lloyd Max Quantizer - Dither - Color Images - Linear Systems And Shift Invariance - Fourier Transform - Z Transform - Matrix Theory Results - Block Matrices and Kronecker Products.

### Unit - II

11 hrs

**Image Transforms:** 2-D orthogonal and Unitary transforms - 1-D and 2-D DFT - Cosine - Sine - Walsh - Hadamard - Haar - Slant - Karhunen-loeve - Singular value Decomposition transforms.

### Unit - III

11 hrs

**Image Enhancement:** Point operations - contrast stretching - clipping and thresholding density slicing - Histogram equalization - modification and specification - partial operations - spatial averaging - low pass - high pass - band pass filtering - direction smoothing - medium filtering - generalized spectrum and homomorphic filtering - edge enhancement using 2-D IIR and FIR filters - color image enhancement.

### Unit - IV

11 hrs

**Image Restoration:** Image observation models - sources of degradation - inverse and Wiener filtering - geometric mean filter - nonlinear filters - smoothing splines and interpolation - constrained least squares restoration.

### Unit - V

11 hrs

**Image Data Compression and Image Reconstruction From Projections:** Image data rates - pixel coding - predictive techniques transform coding and vector DPCM - Block truncation coding - wavelet transform coding of images - color image coding. Random transform - back projection operator - inverse random transform - back projection algorithm - fan beam and algebraic restoration techniques.

**Text Book(s)**

1. Rafael Gonzalez, Richard E. Woods, "Digital Image Processing", Fourth Edition, PHI/Pearson Education, 2018

**Books for Reference**

1. Ashok, Alaknanda, Dumka, Ankur, Verma, Parag, "Advanced Digital Image Processing and Its Application in Data Science", Taylor & Francis Limited, 2020.
2. Dougherty, Edward R, "Digital Image Processing Methods, CRC Press, 2020.

## MACHINE LEARNING

### Course Outcomes

On successful completion of the course, the students will be able to

1. List out the fundamental issues and challenges of machine learning: data, model selection, model complexity, etc.
2. Define the underlying mathematical relationships within and across Machine Learning algorithms and the paradigms of supervised and un-supervised learning
3. Identify the strengths and weaknesses of renowned machine learning approaches.
4. Explain machine learning concepts and algorithms
5. Develop and implement various machine learning algorithms in a wide range of real-world applications
6. Examine the ways of Association Rule Learning and Reinforcement Learning

### Unit – I: Introduction

11 hrs

Learning –Types of Machine Learning –Supervised Learning –The Brain and the Neuron – Design a Learning System –Perspectives and Issues in Machine Learning –Concept Learning Task –Concept Learning as Search –Finding a Maximally Specific Hypothesis –Version Spaces and the Candidate Elimination Algorithm –Linear Discriminants –Perceptron – Linear Separability –Linear Regression.

### Unit – II : Linear Model

11 hrs

Multi-layer Perceptron –Going Forwards –Going Backwards: Back Propagation Error – Multi-layer Perceptron in Practice –Examples of using the MLP –Overview –Deriving Back-Propagation –Radial Basis Functions and Splines –Concepts –RBF Network –Curse of Dimensionality –Interpolations and Basis Functions –Support Vector Machines

### Unit – III : Dimensionality Reduction And Evolutionary Models

11 hrs

Dimensionality Reduction –Linear Discriminant Analysis –Principal Component Analysis – Factor Analysis –Independent Component Analysis –Locally Linear Embedding –Isomap – Least Squares Optimization –Evolutionary Learning –Genetic algorithms –Genetic Offspring: -Genetic Operators –Using Genetic Algorithms –Reinforcement Learning – Overview –Getting Lost Example –Markov Decision Process

### Unit – IV: Graphical Models

11 hrs

Markov Chain Monte Carlo Methods –Sampling –Proposal Distribution –Markov Chain Monte Carlo –Graphical Models –Bayesian Networks –Markov Random Fields –Hidden Markov Models –Tracking Methods

### Unit – V : Machine Learning Applications across Industries

11 hrs

Healthcare – Retail - Financial Services – Manufacturing – Hospitality - Cloud Based ML Offerings

### Text Book(s)

1. Stephen Marsland, “Machine Learning –An Algorithmic Perspective”, Second Edition, Chapman and Hall/CRC Machine Learning and Pattern Recognition Series, 2015.

### **Books for Reference**

1. Theodoridis, Sergios, "Machine Learning: A Bayesian and Optimization Perspective", Elsevier Science Publications, 2020.
2. Aboul Ella Hassanien, Ashraf Darwish, Roheet Bhatnagar, "Advanced Machine Learning Technologies and Applications", Springer Publication, 2020.



## TIME SERIES AND SAMPLING THEORY

### Course Outcomes

On successful completion of this course, the students will be able to

1. Recall the basic concepts of time series sampling theories.
2. Find seasonal variations in time series techniques
3. Demonstrate various sampling techniques
4. Illustrate the trends in seasonal variations
5. Apply time series and sampling theory techniques to select samples and analyze them wisely.
6. Analyze mean and variance of diverse samples

### Unit - I

11 hrs

**Introduction to Time Series:** Concepts of time series - Components of time series - Additive and multiplicative models for the analysis of time series - **Measurement of trend:** Graphic method - Semi Average method - **Method of Curve Fitting:** principle of least squares - Method of Moving Averages.

### Unit - II

11 hrs

**Time Series Measurements:** Measurement of Seasonal Variations - Method of simple average - Ratio-to-trend method - Ratio- to-Moving Average Method - Link Relatives method. **Measurement of Cyclic variations:** Residual approach - Random Component of a time series – Variate difference method

### Unit - III

11 hrs

**Simple random sampling:** WR and WOR - Use of Random number Table-Unbiased estimates of Mean and Variance - Sampling for attributes.

### Unit – IV

11 hrs

**Stratified Random Sampling:** Properties - Unbiased Estimate of the Mean and Variance of the Estimated Mean

### Unit – V

11 hrs

**Systematic Sampling:** Estimation of the Mean and Variance - Comparison of Simple, Stratified and Systematic Sampling - Population with Linear Trend - Circular Systematic Sampling.

### Text Book(s)

1. Gupta, S.C. and Kapoor, V.K, “Fundamentals of Applied Statistics”, Sultan Chand & Co., 2012 (Units 1,2,3).
2. William G. Cochran, “Sampling Techniques”, John Wiley Sons,2007.(Unit 4,5)

### Books for Reference

1. Bob Mather, "Time Series with Python: How to Implement Time Series Analysis and Forecasting Using Python", Abipro Pty Limited, 2020.
2. Nielsen, Aileen, "Practical Time Series Analysis: Prediction with Statistics and Machine Learning", O'Reilly Media, 2019.

**Software Lab-IV**  
**IMAGE PROCESSING**

**Course Outcomes**

On successful completion of the course, the students will be able to

1. Recall the basic image operations and transform images
2. Illustrate how to employ various detection techniques for images classification
3. Demonstrate the methods to create better image solutions.
4. Show how to process video using various motion detection techniques
5. Develop efficient programs to disentangle problems that occur while processing the images
6. Analyze image processing techniques to detect accurate results

**Lab Exercises:**

1. Installing and configuring Open CV
2. Basic Image Operations – Colors, Pixel manipulation, filtering, blur, dilation and erosion
3. Scale, rotate and transform images
4. Skew Detection and Correction
5. Object Detection and Thresholding
6. Contour Object Detection and Edge Detection
7. Digit Detection
8. Handling Video Input
9. Motion Detection using Open CV
10. Build a CAR Number Plate Detection System.

**Software Lab-V**  
**MACHINE LEARNING**

**Course Outcomes**

On successful completion of the course, the students will be able to

1. Recall dataset characteristics
2. Define various extraction techniques
3. Explain a wide range of Machine Learning techniques
4. Apply classification and regression in machine learning on various systems
5. Develop efficient programs to resolve problems using algorithms and machine learning techniques
6. Discover machine learning techniques to solve real world problems

**Lab Exercises:**

1. Cleaning and Pre-processing Data using Pandas and NumPy
2. Data Imputation, Normalization and EDA using Pandas and NumPy
3. Feature Extraction and Feature Engineering
4. Simple Linear Regression Models
5. Complex Regression Models for Non-Linear Problems
6. Classification using Tree based Models
7. Ensemble Models for Regression and Classification
8. Building Recommender Systems
9. Artificial Neural Networks for Classification and Regression
10. Deep Learning models for Classification and Regression

**INTERNSHIP****Course Outcomes**

On successful completion of the course, the students will be able to

1. Explore career alternatives prior to graduation.
2. Assess interests and abilities in their field of study.
3. Learn to appreciate work and its function in the economy.
4. Develop work habits and attitudes necessary for job success.
5. Develop communication, interpersonal and other critical skills in the job interview process.
6. Acquire employment contacts leading directly to a full-time job following graduation from college.

**Internship:**

The Student has to undergo 40 Days of Internship Programme in the Industry

Sl. No	Area of Work	Maximum Marks
1.	(i) Work Related performance – Work Attitude / Academic Preparation / Problem Solving Ability / Adaptability / Overall Attendance / Progress towards learning goals.	20
	(ii) Organizational skills – Time Management Skills / Planning Skills	20
	(iii) Relationships with others - Willingness to cooperate with co-workers / Ability to work with supervisor / Acceptance of constructive comments / Ability to take direction	20
	(iv) Communication skills – Oral Communication / Written Communication / Listening skills	20
2.	VIVA VOCE EXAMINATION	20
	<b>TOTAL</b>	<b>100</b>

Core Elective IV  
**SOFTWARE PROJECT MANAGEMENT**

**Course Outcomes**

On successful completion of the course the students will be able to

1. Find out how to align projects with organizational strategy to improve project success rates
2. Define the requirements of clients at each stage of the software development life cycle
3. Explain the framework of project management and Learn to monitor and control the project
4. Demonstrate sound knowledge of agile methods
5. Identify the team, cost, quality and resource management and control risks in project management
6. Analyze the scope, cost, timing, and quality of the project, at all times

**Unit - I**

**11 hrs**

**Project Management Framework:** Project - Project management - Relationship among Project, Program and Portfolio management - Project and operations management- Role of project manager Project management body of knowledge - Enterprise Environmental factors.  
**Project life cycle and Organization:** Overview of project life cycle - Projects vs Operational Work - Stakeholders - Organizational influences on project management. **The Standard for Project Management of a Project:** Project management processes for a project: Common project management process interactions - Projects management process groups - Initiating process group - planning process group - Executing process group Monitoring and controlling process group - Closing process group.

**Unit - II**

**11 hrs**

Choosing Methodologies and Technologies – Software Processes and Process Models – Choice of Process Models – The Waterfall Model– Prototyping – other ways of categorizing prototype - **Agile Methods** – Extreme Programming–Selecting the Most Appropriate Process Model- Need of Agile - Iterative vs Incremental-Agile Manifesto and Mindset – Lean, Scrum and Kanban methods-uncertainty, Risk, and lifecycle selection-Scrum Elements overview-5 levels of planning-Scrum Process overview-Agile Team- roles and responsibilities- Epic-feature- User Stories-PBI-The Sprint.

**Unit - III**

**11 hrs**

**The Project Management Knowledge Areas:** Project integration management: Develop project charter - Develop project management plan - Direct and manage project execution - Monitor and control project work Perform integrated change control - Close project or phase. Project scope management: Collect requirements - Define Scope - Create WBS - Verify Scope - Control Scope. Project team management: Define activities - Sequence activities - Estimate activity resources - Estimate Activity Durations - Develop Schedule - Control Schedule

**Unit - IV**

**11 hrs**

**Project cost management:** Estimate costs - Determine budget - Control costs. Project Quality Management: Plan quality - perform quality assurance - Perform quality control.

Project Human Resource Management: Develop human resource plan - Acquire project team - Develop project team - Manage project team. **Project Communications Management:** Identify stakeholders - Plan communications - Distribute information - Manage stakeholder expectations - report performance.

**Unit – V**

**11 hrs**

**Project Risk Management:** Plan risk management - Identify risks - Perform qualitative risk analysis - Perform quantitative risk analysis - plan risk responses - Monitor and control risks.

**Project Procurement Management:** Plan - Conduct - Administer - Close procurements.

**Text Book(s)**

1. BOB Huges, Mike Cotterell, Rajib Mall, “Software Project Management”, McGraw Hill, Fifth Edition,2011.

**Books for Reference**

1. Adolfo Villafiorita,"Introduction to Software Project Management"CRC Press, 2016
2. C.Ravindranath Pandian, “Applied Software Risk Management-A Guide for Software Project Managers”, Auerbach Publications, 2015.

Core Elective IV  
**NATURAL LANGUAGE PROCESSING**

**Course Outcomes**

On successful completion of the course, the students will be able to

1. Recall the basic concepts of NLP
2. Define the syntax, semantics and pragmatics of NLP
3. Demonstrate the probabilistic models employed in NLP
4. Explain the core ideologies on the concepts of NLP
5. Apply Morphology and other techniques to recognize parts of speech
6. Build models by deploying the suitable syntax and semantics

**Unit - I** **11 hrs**

Overview: Origins and challenges of NLP- Theory of Language -Features of Indian Languages - Issues in Font -Models and Algorithms- NLP Applications.

**Unit - II** **11 hrs**

Phonology - Computational Phonology - Words and Morphemes - Segmentation - Categorization and Lemmatization - Word Form Recognition - Valency - Agreement - Regular Expressions - Finite State Automata - Morphology- Morphological issues of Indian Languages - Transliteration.

**Unit - III** **11 hrs**

Probabilistic Models of Pronunciation and Spelling - Weighted Automata - N- Grams - Corpus Analysis - Smoothing - Entropy - Parts-of-Speech - Taggers - Rule based - Hidden Markov Models - Speech Recognition.

**Unit - IV** **11 hrs**

Basic Concepts of Syntax - Parsing Techniques - General Grammar rules for Indian Languages - Context Free Grammar - Parsing with Context Free Grammars - Top Down Parser - Earley Algorithm - Features and Unification - Lexicalised and Probabilistic Parsing.

**Unit – V** **11 hrs**

Representing Meaning -Computational Representation - Meaning Structure of Language - Semantic Analysis -Lexical Semantics - WordNet - Pragmatics - Discourse - Reference Resolution - Text Coherence - Dialogue Conversational Agents.

**Text Book(s)**

1. Daniel Jurafsky and James H. Martin “Speech and Language Processing”, Prentice Hall, 2019.

**Books for Reference**

1. Brojo Kishore Mishra, Raghvendra Kumar, "Natural Language Processing in Artificial Intelligence" Apple Academic Press, 2020.
2. Yuli Vasiliev, "Natural Language Processing with Python and SpaCy: A Practical Introduction", No Starch Press, 2020.

**IDC:SENTIMENT ANALYSIS**

**Course Outcomes**

On successful completion of the course, the students will be able to

1. Recall various tools used for sentiment analysis
2. Find solutions and observations from classified data
3. Classify data based on the sentiment analysis performed
4. Identify the results and summarize the opinions in an effective way
5. Apply the aspects of sentiment classification in social media platforms
6. Analyze numerous methodologies to perform sentiment classification

**Unit - I** **11 hrs**

Need for Sentiment Analysis - Problem of Sentiment Analysis - Subjectivity - Stance - Words to Discourse Pragmatics - Natural Language Processing issues - Opinion Definition - Sentiment analysis Tasks - Opinion Summarization - Types of opinion - Subjectivity and emotion - Author and Reader Standpoint

**Unit - II** **11 hrs**

Sentiment classification using Supervised learning - unsupervised learning - rating prediction - cross- domain sentiment classification - cross-language sentiment classification - Sentence subjectivity and classification - subjectivity classification - sentence sentiment classification - conditional sentences - sarcastic sentences - cross-language subjectivity and sentiment classification - Discourse information for sentiment classification

**Unit - III** **11 hrs**

Aspect sentiment classification - rules of opinions and compositional semantics - aspect extraction - identifying resource usage aspect - simultaneous opinion lexicon expansion and aspect extraction - Grouping aspects into categories - entity, opinion hold and timing extraction - coreference resolution and word sense disambiguation - aspect and entity extraction - sentiment lexicon generation - corpus based approach - dictionary based approach - desirable and undesirable facts

**Unit - IV** **11 hrs**

Aspect based opinion summarization - improvements to aspect-based opinion summarization - contrastive view summarization - traditional summarization - Analysis of comparative opinions - identifying comparative sentences - identifying preferred entities - opinion search and retrieval - opinion spam detection - types of spam detection - supervised and un-supervised approach - group spam detection

**Unit – V** **11 hrs**

Detecting fake or deceptive opinions - Quality of Review - Quality as regression model - other methods - Case study - sentiment analysis applications - tools for sentiment analysis - Semantria - Meltwater - Google Analytics - Face book Insights - Tweetstats.

**Text Book(s)**

1. Bing Liu, “Sentiment Analysis - Mining opinion, Sentiments and Emotions”, Cambridge University Press,2015.



## **Books for Reference**

1. Antonio Moreno, Carlos A. Iglesias, "Sentiment Analysis for Social Media", MDPI AG, 2020.
2. Bing Liu, "Sentiment Analysis: Mining Opinions, Sentiments, and Emotions", Cambridge University Press, 2020.

## PROJECT WORK

### Course Outcomes

On successful completion of the course, the students will be able to

1. Show leadership skills and learn time management
2. Explain various tools to be applied to a specific problem
3. Demonstrate individual skills to present and organize projects
4. Identify the art of the documentation
5. Make use of the concepts of data science
6. Take part in a team as well as manage it to deliver stunning outcomes

### Project Work

SL	Area of Work	Maximum Marks
1.	<b>PROJECT WORK:</b> (i) Plan of the Project	40
	(ii) Execution of the plan / Collection of data / Organization of materials / Fabrication Experimental study / Hypothesis, Testing etc., and Presentation of the report.	90
	(iii) Individual Initiative	20
2.	VIVA VOCE EXAMINATION	50
	<b>TOTAL</b>	<b>200</b>

Note: PASSING MINIMUM - 50% MARKS

## COMPREHENSIVE EXAMINATION

### Course Outcomes

On successful completion of the course, the students will be able to

7. Understand the basic and advanced methods of big data technology
8. Define Database Models and Database Management Skills
9. Show a comprehensive understanding of cloud technologies and architectures
10. Develop the Web Application
11. Explain machine learning concepts
12. Analyse the scope, cost, timing, and quality of the project, at all times

### Unit - I

**Introduction to Big Data Analytics:** Big Data Overview – Big Data Ecosystem - Data Analytics Lifecycle – Data Discovery – Data Preparation – Model Planning – Model Building – Communicate Results – Operationalize. **Relational and parallel Database Design:** Basics, Entity Types, Relationship Types, ER Model, ER-to- Relational Mapping algorithm. Normalization: Functional Dependency, 1NF, 2NF, 3NF, BCNF, 4NF and 5NF. **Distributed and Object based Database:** Architecture, distributed data storage, Distributed transactions, Commit protocols, Concurrency control, Query Processing.

### Unit - II

**Cloud Computing:** Evolution of Cloud Computing -Essential Characteristics of cloud computing - Operational models - Service models. Google cloud platform - Amazon AWS - Microsoft Azure - Pivotal cloud foundry and Open Stack. **Introduction to Django Framework** - Django Framework components - Installing Python and Django - Setting Up Django Projects - Working with Django Models - Models and fields - Import CSV Data - Django Admin - Query with Django ORM.

### Unit – III

**Introduction to Machine Learning:** Learning –Types of Machine Learning –Supervised Learning –The Brain and the Neuron –Design a Learning System –Perspectives and Issues in Machine Learning –Concept Learning Task –Concept Learning as Search –Finding a Maximally Specific Hypothesis –Version Spaces and the Candidate Elimination Algorithm –Linear Discriminants –Perceptron –Linear Separability. **Introduction to Time Series:** Concepts of time series - Components of time series - Additive and multiplicative models for the analysis of time series - Measurement of trend by (i) Graphic method, (ii) Semi Average method, (iii) Method of Curve Fitting by principle of least squares, (iv)Method of Moving Averages.

### Unit - IV

**Introduction:** Soft Computing Constituents – Soft Computing Vs Hard Computing – Characteristics - Applications - **Artificial Neural Network (ANN):** Fundamental Concept – Application Scope – Basic Terminologies – Neural Network Architecture – Learning Process – **Basic Models of ANN:** McCulloch- Pitts Model – Hebb Network. **Introduction:** Data-Analytic Thinking - The Ubiquity of Data Opportunities -Data Science, Engineering, and Data-Driven Decision Making - Data Processing and “Big Data” - Data and Data Science Capability as a Strategic Asset. Business Problems and Data Science Solutions - From Business Problems to Data Mining Tasks - Supervised Versus Unsupervised Methods - Data

Mining and Its Results - The Data Mining Process - Other Analytics Techniques and Technologies.

## Unit – V

**Introduction to Algorithm:** Algorithm Definition – Algorithm Specification –Asymptotic Notations. Elementary Data Structures: Stacks and Queues – Trees – Graphs. **Project Management Framework:** Project - Project management - Relationship among Project, Program and Portfolio management - Project and operations management- Role of project manager Project management body of knowledge - Enterprise Environmental factors. **Project life cycle and Organization:** Overview of project life cycle - Projects vs Operational Work - Stakeholders - Organizational influences on project management.

### Books for Reference

1. EMC Education Services “Data Science & Big Data Analytics: Discovering, Analysing, Visualizing and Presenting Data”, Wiley & Sons, Inc Publications, 2016
2. Abraham Silberschatz, Henry F Korth , S Sudarshan, “Database System Concepts”, 6th edition , McGraw-Hill International Edition , 2015
3. Buyya, Vecciola and Selvi, “Mastering Cloud Computing: Foundations and Applications Programming”, Tata McGraw Hill, 2013
4. Nigel George , “Build Your First Website with Django 2.1: Master the Basics of Django While Building a Fully-Functioning Website”, Paperback, 2018.
5. Stephen Marsland, Machine Learning –An Algorithmic Perspective, Second Edition, Chapman and Hall/CRC Machine Learning and Pattern Recognition Series, 2014.
6. Gupta, S.C. and Kapoor, V.K, “Fundamentals of Applied Statistics”, Sultan Chand & Co., 2012
7. S.N. Sivanandam, S.N. Deepa, “Principles of Soft Computing”, 2<sup>nd</sup> Edition, Wiley India Pvt. Limited, 2011
8. Provost, Foster, and Tom Fawcett, “Data Science for Business: What you need to know about data mining and data-analytic thinking”, O’Reilly Media, Inc., 2013
9. Ellis Horowitz, Satraj Sahni and Sanguthevar Rajasekaran, “Fundamentals of Computer Algorithms”, Universities Press, Second Edition, Reprint 2009.
10. BOB Huges, Mike Cotterell, Rajib Mall, “Software Project Management”, McGraw Hill, Fifth Edition,2011.

Core Elective V  
SOFT COMPUTING

Course Outcomes

On successful completion of the course, the students will be able to

1. Recall the fuzzy logic and the concept of fuzziness involved in various systems and fuzzy set theory
2. Find and get to the bottom of the concepts of fuzzy sets, knowledge representation using fuzzy rules, approximate reasoning, fuzzy inference systems, and fuzzy logic
3. Explain neural network theory and fuzzy logic theory
4. Interpret artificial neural networks and fuzzy theory from an engineering perspective
5. Apply appropriate learning rules for each of the architectures and learn several neural network paradigms and its applications.
6. Examine different applications of these models to solve engineering and other problems

**Unit - I** **11 hrs**

**Introduction:** Soft Computing Constituents – Soft Computing Vs Hard Computing – Characteristics - Applications - Artificial Neural Network (ANN): Fundamental Concept – Application Scope – Basic Terminologies – Neural Network Architecture – Learning Process – Basic Models of ANN: McCulloch- Pitts Model – Hebb Network – Linear Separability.

**Unit - II** **11 hrs**

**Supervised Learning Networks:** Perceptron Networks – Adaline and Madaline Networks – Back Propagation Network – Radial Basis Function Network. Associative Memory Networks – BAM – Hopfield Network - Boltzmann Machine. Unsupervised Learning Networks: Kohonen Self Organizing Network – Counter Propagation Network – ART Network.

**Unit - III** **11 hrs**

**Fuzzy Sets:** Basic Concept – Crisp Set Vs Fuzzy Set - Operations on Fuzzy Set – Properties of Fuzzy Sets – **Fuzzy Relations:** Concept – Fuzzy Composition – Fuzzy Equivalence and Tolerance Relation - Membership Functions: Features – Fuzzification – Methods of Membership value assignments – Defuzzification – Methods.

**Unit - IV** **11 hrs**

Fuzzy Arithmetic – Extension Principle – Fuzzy Measures – Fuzzy Rules and Fuzzy Reasoning: Fuzzy Propositions – Formation of Rules – Decomposition of Rules – Aggregation of Rules – Approximate Reasoning – Fuzzy Inference and Expert Systems – Fuzzy Decision Making – Fuzzy Logic Control Systems.

**Unit – V** **11 hrs**

**Genetic Algorithm:** Fundamental Concept – Basic Terminologies – Traditional Vs Genetic Algorithm - Elements of GA - Encoding - Fitness Function – Genetic Operators: Selection – Cross Over - Inversion and Deletion - Mutation – Simple and General GA – The Schema Theorem - Classification of Genetic Algorithm – Genetic Programming – Applications of GA.

**Text Book(s)**

1. S.N. Sivanandam, S.N. Deepa, "Principles of Soft Computing", 2<sup>nd</sup> Edition, Wiley India Pvt. Limited, 2011

**Book for Reference**

1. Ashish Mishra, G. Suseendran, Trung-Nghia Phung, "Soft Computing Applications and Techniques in Healthcare", CRC Press, 2020.
2. Mangey Ram, Suraj B. Singh, "Soft Computing: Techniques in Engineering Sciences", De Gruyter Publication, 2020.

**Core Elective V  
DEEP LEARNING**

**Course Outcomes**

On successful completion of the course, the students will be able to

1. Recall Deep Learning techniques
2. Explain Boltzmann machines and computer vision of Deep learning
3. Identify the neural network technology and their functions
4. Examine deep learning concepts and acquaint themselves with the usage of ANN, CNN, RNN techniques
5. Apply the Boltzmann machine principle to appropriate environments
6. Analyze algorithms to spot problems in existing DL techniques

**Unit - I: Artificial Neural Networks** **11 hrs**

The Neuron - Activation Function - Gradient Descent - Stochastic Gradient Descent - Back Propagation - Business Problem.

**Unit - II: Convolutional Neural Networks** **11 hrs**

Convolution Operation - ReLU layer - Pooling - Flattening - Full Conversion Layer - Softmax and Cross- Entropy.

**Unit - III: Recurrent Neural Networks** **11 hrs**

RNN intuition - Tackling Vanishing Gradient Problem - Long Short-Term Memory - Building a RNN - Evaluating the RNN - Improving the RNN - Tuning the RNN.

**Unit - IV: Boltzmann Machines** **11 hrs**

Introduction to Boltzmann Machine - Energy-Based Models - Restricted Boltzmann Machine - Contrastive Divergence - Deep Belief Networks - Deep Boltzmann Machine.

**Unit - V: Computer Vision** **11 hrs**

Viola-Jones Algorithm - Haar-like Features - Integral Image - Training Classifiers - Adaptive Boosting - Cascading - Face Detection with Open CV.

**Text Book**

1. Josh Patterson and Adam Gibson, "Deep Learning A practitioners Approach",Shroff publishers & Distributors, First edition 2017

**Book for Reference**

1. Chittaranjan Pradhan, Himansu Das, Nilanjan Dey, "Deep Learning for Data Analytics: Foundations, Biomedical Applications, and Challenges",Elsevier Science publication, 2020.
2. Gaurav Meena, Kamal Kant Hiran, Mehul Mahrishi, Paawan Sharma, " Machine Learning and Deep Learning in Real-Time Applications",IGI Global Publication, 2020.

Core Elective VI  
**DATA SCIENCE FOR BUSINESS**

**Course Outcomes**

On successful completion of the course, the students will be able to

1. Recall data science techniques and methodologies to business environment
2. Demonstrate analytic thinking capacity to address business problems
3. Interpret the findings using visualization techniques
4. Make use of fundamental algorithmic ideas to process data
5. Apply modeling and predictive strategies and make estimation
6. Build models with efficient decision making abilities

**Unit - I** **11 hrs**

**Introduction:** Data-Analytic Thinking - The Ubiquity of Data Opportunities -Data Science, Engineering, and Data-Driven Decision Making - Data Processing and “Big Data” - Data and Data Science Capability as a Strategic Asset. Business Problems and Data Science Solutions - From Business Problems to Data Mining Tasks - Supervised Versus Unsupervised Methods - Data Mining and Its Results - The Data Mining Process - Other Analytics Techniques and Technologies.

**Unit - II** **11 hrs**

**Introduction to Predictive Modeling:** From Correlation to Supervised Segmentation - Models, Induction, and Prediction - Supervised Segmentation -Visualizing Segmentations - Trees as Sets of Rules - Probability Estimation.

**Unit - III** **11 hrs**

**Overfitting and Its Avoidance:** Generalization - Overfitting - Overfitting Examined -From Holdout Evaluation to Cross-Validation - Learning Curves - Overfitting Avoidance and Complexity Control.

**Unit - IV** **11 hrs**

**Distance Metrics in Business:** Similarity, Neighbors, and Clusters - Similarity and Distance - Nearest-Neighbor Reasoning – Some Important Technical Details Relating to Similarities and Neighbors - Clustering - Stepping Back: Solving a Business Problem Versus Data Exploration.

**Unit – V** **11 hrs**

**Decision Analytic Thinking:** characteristics of a Good Model - Visualizing Model Performance - Representing and Mining Text - Other Data Science Tasks and Techniques.

**Text Book(s)**

1. Provost, Foster, and Tom Fawcett, “Data Science for Business: What you need to know about data mining and data-analytic thinking”, O’Reilly Media, Inc., 2013.(Unit 1,2)
2. Asllani, Arben, “Business Analytics with Management Science Models and Methods”, FT Press, 2014.(Unit 3,4,5)



**Books for Reference**

1. Probyto Data Science and Consulting Pvt. Ltd, "Data Science for Business Professionals: A Practical Guide for Beginners",BPB Publication, 2020.
2. Chirag Shah,"A Hands-On Introduction to Data Science", Cambridge University Press, 2020.

**Core Elective VII**  
**BUSINESS DATA ANALYTICS**

**Course Outcomes**

On successful completion of the course, the students will be able to

1. Recall core statistical techniques, data retrieval, analysis and mining
2. Find significant insights into data-driven decision making
3. Explain the ideologies of the data and the business worlds to acquire outstanding results
4. Examine the purpose of employing business analysis tools within an organization
5. Apply critical thinking skills involved in the critical thinking processes to draw inferences
6. Analyze data for business decision making

**Unit - I** **11 hrs**

**Introduction to Business Analytics:** Business Analytics - Process - Identifying Data -Types and Stages of Data Analytics - Business Intelligence and Data Engineering - Exploratory Data Analytics - Communicating Business Analytics Results

**Unit - II** **11 hrs**

**Critical Thinking in Data Science:** Asking Data Science Questions - Good Question-Critical Thinking, Reasoning - Weak Sense Critical Thinking and Strong Sense Critical Thinking - Question Meetings - Question Types - Key areas of Questioning - Challenging Evidence - Statistical Inference.

**Unit - III** **11 hrs**

**Prescriptive Analytics:** Introduction - Prescriptive Analytics Workflow - Case Studies.  
**Experimental Analytics:** Introduction - Workflow and Case Studies.

**Unit - IV** **11 hrs**

**Business Strategy Analyticsf:** Problems on Analysis - Goal Seek - Solvers for Optimization - Linear and non-linear mixture problems - Importance of Creating and Evaluating Business Scenarios.

**Unit – V** **11 hrs**

**Marketing Analytics:** Case Studies on Customer Analytics - Marketing Performance Measurement and Management - Employee Performance Analytics - Operation Analytics - Accounting Analytics.

**Text Book(s)**

1. David Stephenson , “Big Data Demystified: How to use big data, data science and AI to make better business decisions and gain competitive advantage”, Pearson, 2018

**References**

1. Jay Liebowitz, "Data Analytics and AI",CRC Press, 2020.
2. Matt Taddy, “Business Data Science: Combining Machine Learning and Economics to Optimize, Automate, and Accelerate Business Decisions”, McGraw-Hill Education, 2019.

## SOFT SKILLS

### Course Outcomes

On successful completion of the course, the students will be able to

1. Recall the nuances of grooming such as, good manners and etiquettes
2. Demonstrate the public speaking skills via extempore speeches and prepared speeches
3. Model their own resumes and present them before the interview panel for their mock interview
4. Explain and learn the ten parameters of group discussion, perform on the stage with their associates, which is videotaped, reviewed and evaluated.
5. Analyse and self-discover themselves to develop their own personalities facilitated with scientific psychological personality tests
6. Apply analytical skills to solve real world problems

#### Unit - I

11 hrs

**Basics of Communication:** Definition of communication, Process of Communication, Barriers of Communication, Non-verbal Communication, Effective Communication: The Art of Listening, Exercises in Kinesthetics, Production of Speech, Organization of Speech, Modes of delivery, Conversation Techniques, Dialogue, Good manners and Etiquettes, Politeness markers & Listening links

#### Unit - II

11 hrs

**Resume Writing:** Resume, Types of Resume, Chronological, Functional and Mixed Resume, Steps in preparation of Resume, structure and framework for writing resume, Intensive training / personalized training on resume writing. Interview Skills: Common interview questions, Attitude, Body Language, The mock interviews, Phone interviews, Behavioral interviews.

#### Unit - III

11 hrs

**Group Discussion:** Group Discussion Basics, GD Topics for Practice, Points for GD Topics, Case-Based and Article based Group Discussions, Points for Case Studies, and Notes on Current Issues for GDS & Practicum with video coverage. Team Building: Team Vs Group – Synergy, Stages of Team Formation, Broken Square-Exercise, Win as much as you win-Exercise, Leadership – Styles, Work ethics. **Personal Effectiveness:** Self Discovery, Self Esteem, Goal setting, Problem-solving, Conflict and Stress Management

#### Unit - IV

11 hrs

**Numerical Ability:** Average, Percentage, Profit and Loss, Problems on ages, Simple Interest, Compound Interest, Area, Volume and Surface Area, Time and Work, Pipes and Cisterns, Time and Distance, Problems on Trains, Boats and Streams, Calendar, Clocks, Permutations and Combinations, Probability.

#### Unit - V

11 hrs

**Test of Reasoning:** Series Completion, Analogy, Data Sufficiency, Blood Relations, Assertion and Reasoning, Logical Deduction, Direction. Non-Verbal Reasoning: Series, Classification

### Text Book

1. Melchias, G., Balaiah John., John Love Joy, “ Winners in the making”, St.Joseph’s College, Trichy-2, 2015
2. Aggarwal, R. S., “ Quantitative Aptitude”, S.Chand & Sons, 2017

**Books for Reference**

1. Disha Experts, "Quantitative Aptitude & Data Interpretation", 3rd Edition, Disha Publications, 2019.
2. Aggarwal, R.S., "A Modern Approach to Verbal and Non Verbal Reasoning", S. Chand & Co, 2018.