DEPARTMENT OF COMPUTER SCIENCE

RATHINAM COLLEGE OF ARTS AND SCIENCE (AUTONOMOUS)

RathinamTechzone, Pollachi Road, Eachanari, Coimbatore – 641021

Program for

M.Sc. Data Science and Business Analysis

(M.Sc. DSBA)

(I, II, III and IV Semester)

2021 – 2023 Batch onwards

RATHINAM COLLEGE OF ARTS AND SCIENCE (AUTONOMOUS)
Scheme of curriculum for M.Sc. Data Science and Business Analysis
Batch admitted during 2021 - 2023

Board of Studies – Computer Science (PG)

Rathinam College of Arts & Science (Autonomous), Coimbatore-21.
Admitted in M.SC DSBA from the academic year 2021 Onwards

Page 2 of 90
Regulations 2021

1. Vision and Mission of the Institution:

VISION
A world renowned INDUSTRY-INTEGRATED INSTITUTION that imparts knowledge, skill, and research culture in young men and women to suit emerging young India.

MISSION
To provide quality education at affordable cost, and to maintain academic and research excellence with a keen focus on INDUSTRY-INTEGRATED RESEARCH AND EDUCATION.

MOTTO
Meaningful INDUSTRY-READY education and research by all means

2. Vision and Mission of the Department:

VISION
To inculcate candidates with technical competency and skill professionalism to make them competent to undertake the current challenges in industrial sector with a focus on societal transformation.

MISSION
To import quality based education by enhancing the talent, innovative idea, and problem solving skill and to promote the researchproject by establishing industrial linkage and entrepreneurial setup.

3. Introduction

Battle for high-quality jobs is getting tougher in this competitive world, so it's more important than ever to have the best qualifications possible for better career opportunities. Studying for an MSc for sake, won't be enough by itself to get into the commendable level, but it could be extremely rewarding in other ways if one plans better. Masters in cross level programme from under graduation would widen the scope of career opportunity in this competitive world with higher quotient in selection for both the fields.

Upgrading one's qualifications and getting an MSc is a brilliant way to prove the potential and sustainable employers that one have what it takes to work in a high-profile position. Not only does it demonstrate the potential candidate can handle additional responsibility, it also suggests that added value and asset to the organization. That is the

reason, why many of the senior managers at leading companies tend to have a master's degree. Master in Business Analytics will prove the quality of a candidate in this "Data Era".

Career? Or Job? Both the terms are actionably seeming to be related but not for sustainable development and goal achievement. Career meant for long term, futuristic and sustainable. Job doesn't. Choosing career after university education is a deciding factor for one's living standards. At this current situation, heavily populated country like India, getting employment is a bigger challenge. Making better decision about career plan is the primary achievement, as career plan enhance the one's future value. After globalization, career evangelist come up with new era called "Data Era", as from last two years of data generation is approximately equal to the data generated in past all years. Hence, every industry dumped with 4V's of data – Volume, Velocity, Variety and Veracity, and they are need to be analysed for the better nurturement of business with meaning insights.

Irrespective of any business domain, industries are in dire need of analysing their huge volume of data in a proper way to come up with new insights which will help them in revenue generation and business development. There is a huge dearth of resource person on analysing the data and bringing with new ideas in parallel and vertical development of their business.

Business Analytics refers to the skills, technologies, practices that are applied on past data and/or processes to derive insights that can be used for future business planning. It is a field that is now applied across all domains and industries.

The comprehensive Business Analytics curriculum provides a framework through which participants learn to enhance their management skills, expand their knowledge of Business Analytics, and gain a strategic perspective of the industry. Business analytics refers to the ways in which enterprises such as businesses, non-profits, and governments can use data to gain insights and make better decisions. Business analytics is applied in operations, marketing, finance, and strategic planning among other functions. The ability to use data effectively to drive rapid, precise and profitable decisions has been a critical

strategic advantage for companies as diverse as Walmart, Google, Capital One, and Disney. For example, Capital One uses sophisticated analytic capabilities to match credit card offerings to customers more accurately than their competition. Walmart uses analytics to monitor and update its inventory in a way that allows it to serve its customers at an exceptionally low cost.The programme's courses and the final project are designed around the real-world integration of business disciplines. Apart from these courses, there are preparatory courses which will have to be completed before the programme begins.

Some of the important positions of Business Analyst Professionals in an organization are as follows:

Business Analyst

Associate Business Analyst

Statistician

Data Architect

Junior Statistician

Market Research Analyst

Data Analytics Manager

Health Care Analyst

4. How Two Year M.Sc (DSBA) Beneficial?

➢ 2-year Master of Science in Data Science and Business Analyst helps a student to gain employment in the following job areas:

Industry	Entry level (0-2 years exp.)	Mid-Level (3-5 years exp.)	Advanced level (5 years plus exp.)
Average Salary	Rs. 5,00,000 – 7,00,000	Rs. 7,00,000 – 10,00,000	Rs. 10,00,000 +

➢ 2-year Master of Science in Data Science and Business Analyst is also highly beneficial to start one's own business as an analytics outsourcing or an independent Business Analyst consultant.

5. Salient features:

➢ This program offers Business Analytics specialization focused on pre-processing, Storing and analytics of data for business environment.

➢ The Program is designed to impart strong knowledge of R programming and Python fundamentals, Business Management Concepts followed by conceptual and practical knowledge of Data Science Techniques and Big Data Analytics.

➢ The Program offers a unique value proposition by combining the important subject areas in each of these new age fields of study for Analyst industry.

➢ The program offers a wide range of technical and programming skill sets that complement the specialization subjects on Business Analytics.

➢ This program is primarily aimed at offering student's flexibility in making their career choices in Business Management, Project Management, Data Management, Data Analytics and Data Visualisation.

➢ The program ignite the spark of interest in widening their knowledge on business analytics techniques in different business domains.

6. Program Educational Objectives (PEO)

PEO1	:	Graduates of this programme will establish as effective professionals by learning technical skills in Business Analytics field and can pursue higher education by accruing knowledge and research.
PEO2	:	To impart sound theoretical foundation and In-depth practical knowledge to analyse the key business processes that drive the value chain of an organization throughout the entire product life cycle.
PEO3	:	Implement a classroom + practical oriented curriculum that helps students understand the Business Analytics Techniques and associated advanced techniques. To understand and analyse models, tools and techniques for enforcement of business analyst to different business industries.
PEO4	:	Provide solutions, assessments and validation to a broad range of situations by eliciting, planning, monitoring and analysing enterprise requirements.
PEO5	:	Provide a platform for students to understand various Business Analytics techniques of data preprocessing, storing, descriptive and predictive analytics.
PEO6	:	Prepare data for statistical analysis, perform basic exploratory and descriptive analysis, and apply statistical techniques to analyze data
PEO7	:	To learn and explore how visualization makes decision makers to understand the business in quick and taking rightful decisions.

7. Mapping of Institute Mission to PEO

Institute Mission	PEO's
To provide quality education at affordable cost, and to maintain academic standards and research excellence with a keen focus on INDUSTRY-INTEGRATED RESEARCH AND EDUCATION.	PEO1, PEO5, PEO7

8. Mapping of Department Mission to PEO

Department Mission	PEO's
To provide better understanding of Data Science and Business Analytics at top level standards and provide hands on experience in major industry demanded business analytics techniques with a keen focus on INDUSTRY-INTEGRATED RESEARCH AND EDUCATION.	PEO2, PEO3,PEO4, PEO6

9. Program Outcomes (PO):

PO1	:	To a given scenario, students will be able to analyze the problem, design strategies and technical requirement to solve them with the meaningful insights for business development.
PO2	:	Student will be able to understand the suitable statistical technique for algorithmic design of the given problem statement
PO3	:	Students will be able to apply clean the data and pre-process them to get ready for the model building and implement the model in the system for required decision making process.
PO4	:	Students will be able to apply their knowledge of machine learning for the better built model with bringing up of meaningful insights to the decision makers.
PO5	:	Students will be able to develop new or improved innovative business processes from gap analysis through process design in support of a company's strategic objectives in a socially responsible manner.
PO6	:	Students will be able to Learn and identify business opportunities and designsolutions and they will be able to discover how to optimize project investments
PO7	:	Students will be able to apply descriptive, predictive and prescriptive analytics to business modelling and decision-making

10. Correlation between the POs and the PEOs

Program Outcomes		PEO1	PEO2	PEO3	PEO4	PEO5	PEO6	PEO7
PO1	:	√	√		√			√
PO2	:	√		√		√	√	
PO3	:			√	√	√	√	√
PO4	:	√		√	√	√	√	
PO5	:	√	√	√	√			√
PO6	:	√	√		√		√	√
PO7	:	√	√			√	√	√

Components considered for Course Delivery is listed below:

1. Class room Lecture

2. Laboratory class and demo

3. Assignments

4. Mini Project

5. Project

6. Online Course

7. External Participation

8. Seminar

9. Internship

11. Mapping of POs with Course Delivery:

Program Outcomes	Course Delivery								
	1	2	3	4	5	6	7	8	9
PO1	√		√	√	√		√	√	√
PO2	√	√	√	√	√		√	√	
PO3	√	√	√				√	√	
PO4	√	√	√	√				√	
PO5	√	√		√	√				√
PO6	√	√	√	√	√		√	√	√
PO7	√	√	√	√	√	√	√	√	√

12.　Course Matrix

S.No	Sem	Part	Type	Subject	Credit	Hour	Int	Ext	Total
1	1	III	Theory	Business Fundamentals – I	3	3	40	60	100
2	1	III	Theory	Business Statistics and Probability	3	3	40	60	100
3	1	III	Theory	Analytics using Excel	3	3	40	60	100
4	1	III	Theory	R Programming Language	3	4	40	60	100
5	1	III	Theory	Business Economics	3	3	40	60	100
6	1	III	Theory	Database Management System	3	4	40	60	100
7	1	III	Practical	R Programming Language Lab	3	4	40	60	100
8	1	III	Practical	Database Management System Lab	3	4	40	60	100
1	2	III	Theory	Big Data Analytics	3	3	40	60	100
2	2	III	Theory	Machine Learning	3	3	40	60	100
3	2	III	Theory	Exploratory Data Analysis	3	3	40	60	100
4	2	III	Theory	Python Programming	3	3	40	60	100
5	2	III	Theory	Market Research and Analytics	3	3	40	60	100
6	2	III	Theory	Business Fundamentals – II	3	3	40	60	100
7	2	III	Practical	Big Data Analytics Lab	3	4	40	60	100
8	2	III	Practical	Python Programming Lab	3	4	40	60	100
9	2	III	Practical	Machine Learning Lab	3	4	40	60	100
1	3	III	Theory	Financial Econometrics	3	2	40	60	100
2	3	III	Theory	Data Visualization	2	2	40	60	100
3	3	III	Theory	Elective I	3	3	40	60	100
4	3	III	Theory	Elective II	3	3	40	60	100
5	3	III	Theory	Advanced Machine Learning	2	3	40	60	100
6	3	III	Theory	Data Analytics using SQL	2	4	40	60	100
7	3	III	Practical	Financial Econometrics Lab	2	3	40	60	100
8	3	III	Practical	Mini Project	3	4	40	60	100
9	3	III	ITR	Industrial Training Report-I	2	-	50	-	50
1	4	III	Theory	Artificial Neural Networks and Deep Learning	2	3	40	60	100
2	4	III	Theory	Elective III	3	4	40	60	100
3	4	III	Theory	Advanced Big Data Analytics	2	3	40	60	100
4	4	III	Theory	Major Project / Internship	10	15	40	60	500
					90	120	1250	1800	3450

List of Electives:

Elective	Subject Name
Elective – I	Operations Research
	Social Media Analytics
Elective – II	Cloud Computing
	Cloud Infrastructure Services
Elective – III	Natural Language Processing
	Cloud Web Services

13. Distribution of Credits Semester-wise

Semester	Credits
1	24
2	27
3	22
4	17
Total	**90**

14.   Syllabus:

SEMESTER -1
Syllabus

Title	Business Fundamental - I
Total Hours	**45**
Credits	**03**

Course Objectives:
- To understand the basic managerial functions of planning, organizing, staffing, directing, and controlling resources to accomplish organizational goals.
- To understand the concept of management and the role of the manager at each level of the organization
- To understand the skills that are necessary for a manager to be effective.

Course Focus on: Skill Development / Entrepreneurship / Employability / Research

Course Outcomes:

On successful completion of the module students will be able to:

- Developed working knowledge of fundamental terminology and frameworks in the four functions of management: Planning, Organizing, Leading and Controlling;
- Analyze organizational case situations in each of the four functions of management
- Identify and apply appropriate management techniques for managing contemporary organizations; and
- Understand the skills, abilities, and tools needed to obtain a job on a management track in an organization of their choice.

Module -1   **9 Hrs.**

Introduction to Management: Defining Management, Concept of Management, Nature, Importance, Management Skills, Levels of Management, Role of managers, Characteristics and Quality Managers, Evolution of Management thought, Organization and the environmental factors.

Business ethics and Social Responsibility: Concept, Shift to Ethics, Tools for Ethics.

Module -2   **9 Hrs.**

Planning : Nature and purpose of planning, Planning process, Types of plans, Process of planning, Barriers to Effective Planning, Objectives, Managing by objective (MBO) Strategies, Types of strategies, Policies, Decision Making, Types of decision, Decision Making Process, Rational Decision Making

Module - 3   **9 Hrs.**

Organizing: Nature and purpose of organizing, Organization structure, Formal and informal groups / organization, Line and Staff authority, Departmentation, Span of control, Centralization and Decentralization, Delegation of authority, Staffing, Selection and Recruitment, Orientation, Career Development, Career stages, Training, Performance Appraisal.

Module-4   **9 Hrs.**

Directing: Creativity and Innovation, Motivation and Satisfaction, Motivation Theories, Leadership Styles, Leadership theories, Communication, Barriers to effective communication, Organization Culture, Elements and types of culture, Managing cultural diversity.

Module-5   **9 Hrs.**

Controlling: Process of controlling, Types of control, Methods: Pre-control, Concurrent Control, Post-control, Budgetary and non-budgetary control Q techniques, Managing Productivity, Cost Control, Purchase Control, Maintenance Control, Quality Control, Planning operations.

Text Books:

1. Stephen P. Robbins and Mary Coulter, 'Management', Prentice Hall of India, 8th edition.

2. Charles W L Hill, Steven L McShane, 'Principles of Management', Mcgraw Hill Education, Special Indian Edition, 2007.

Reference Books:

1. Hellriegel, Slocum & Jackson, ' Management ,A Competency Based Approach', Thomson South Western, 10th edition, 2007.

2. Harold Koontz, Heinz Weihrich and Mark V Cannice, 'Management ,A global

Title	Business Statistics and Probability
Total Hours	45
Credits	03

Course Objectives:

- To understand the basic concepts of statistics and probability.
- To understand the description of data using statistical techniques
- To understand the summary of data using statistical measures
- To understand the statistical methods involved in hypothesis testing
- To understand the ANOVA and its importance in business performance.

Course Outcomes:

On successful completion of the module students will be able to:

- Able to know the importance of statistics in different research areas.
- Able to know the basic concepts of Statistics and its evolution.
- Able to apply suitable statistical measures to describe and summarize the data
- Able to apply t and f test for testing the statistical measures to know the significance.
- Able to apply ANOVA for testing significance of arithmetic mean and regression coefficients.

Module -1   **9 Hrs.**

Descriptive Statistics: Data and Data Sources, Types of Data, Measures of Central Tendency – Mean, median mode for raw and grouped data, measures of dispersion – Range, standard deviation, variance, coefficient of variation, mean deviation, mean absolute deviation, measures of symmetry: Skewness and Kurtosis.

Module -2   **9 Hrs.**

Elements of Probability and Sampling Distributions: Experiments and events, Basic Relations of Probability, Conditional Probability, Joint Probability, conditional probability on discrete case and continuous case, computing expectations by conditioning, introduction to Bayes theorem, problems related to Bayes Theorem, Discrete Probability Distribution (Binomial and Poisson), Continuous Probability Distribution (Normal). Various types of Probability and Non-probability Sampling, Sampling distribution of important statistic.

Module – 3   **9 Hrs.**

Hypothesis Testing : Introduction to testing of hypothesis, Statistical assumptions for parametric test, Level of significance, confidence level, Type I Error, Type II error, Critical value, power of the test, sampling distribution, small sample test – t test for one sample and two sample mean, F test to test the equality of two sample variance, Large Sample test – Z test for equality of single mean with population mean, equality of two sample mean, equality of single proportion with population proportion and equality of two sample proportions.

Module-4   **9 Hrs.**

Correlation and Regression Analysis: Correlation analysis, properties of correlation coefficients, significance of single correlation coefficient, significance of multiple correlation coefficients, concepts of multiple correlation and partial correlation, Introduction to linear model, concepts of factor, effect, residuals, dependency, independency, assumptions of linear model, difference between linear and nonlinear model, estimation of parameters of regression coefficients for simple and multiple linear regression model, properties of regression coefficients, significance of regression coefficients, diagnostic testing: auto correlation, multi collinearity, heteroscedasticity, normality, significance of estimated parameters in multiple linear regression,.

Module-5   **9 Hrs.**

Linear Model: Introduction to general linear model, assumptions of ANOVA, factors and levels in ANOVA, layout of one way ANOVA, skeleton of one way ANOVA, multiple comparison of sample means, one way analysis of variance with unequal sample sizes, two factor analysis of variance – introduction and parameter estimation, two way analysis of variance with interaction, Post ANOVA: testing of hypothesis for significance of mean using Fishers Least Significance Difference test (lsd), Tukeys test, Dunnet test, Duncan Multiple Range test.

Text Books:

mentals of mathematical statistics – SC Gupta and VK Kapoor, Sultan Chand & Sons
ation, New Delhi

Reference Books:

1.   Introduction to probability Models, Ninth Edition – Sheldon M. Ross, Elsevier Publication, Academic Press, UK

2.   Introduction to Probability and Statistics for Engineers and Scientists, Third Edition - Sheldon M. Ross, Elsevier Publication, Academic Press, UK

Title	**Analytics using Excel**
Total Hours	**45**
Credits	**03**

Course Objectives:

- Students will be able to work with Data Entry and various Functions and Formulae of Excel Workbook.
- This Module enables students to do Filtering and Conditional Formatting of data, work on various analysis techniques.
- Students will also be able to do Statistical Analysis techniques on data using Excel.
- Various Simulation techniques, Analysis and Forecasting methods will be taught.

Course Outcomes:

On successful completion of the module students will be able to:

- On successful completion of all modules students get the knowledge of create flexible data aggregations *using* pivot tables.
- Students get the knowledge how to represent the data visually *using* pivot charts.

Module -1   **9 Hrs.**

Functions and Formulas: Understanding Screen Layout - Creating Auto List & Custom List - Entering, Selecting and Editing Data - Understanding References (Relative, Absolute & Mixed) - Working on Various Functions & Formulas - Common Basic Functions - Logical Functions - Text Functions - Date & Time Functions - Lookup & Reference Functions - Mathematical Functions - Conditional Functions - Referring Data from Different Worksheet & Workbook Formula–Auditing -Various Calculation Techniques - Working on Ranges.

Module -2   **9 Hrs.**

Presentation of Data: Sorting Techniques - Various Data Filtering Techniques - Formatting Techniques - Conditional Formatting - Number Formatting - Table Formatting - Protecting Sheets & Files - Understanding Various Excel Window Techniques - Viewing Excel Spreadsheet in various Layouts - Advanced Printing Techniques - Templates - Themes.

Module – 3   **9 Hrs.**

Data Analysis Tools: Data Consolidation - Text to Columns - Flash Fill - Remove Duplicates - Advanced Data Validation Techniques - What-if Analysis - Goal Seek - Data Table - Solver – Scenarios; Working with Tables - Creating Charts - Understanding Sparklines (Line, Column, Win/Loss) - Pivot Tables & Pivot Charts.

Module-4   **9 Hrs.**

Data Analysis: Data Analysis ToolPak – Loading and Activating, ANOVA, correlation, covariance, Descriptive Statistics, Exponential Smoothing, F-Test 2-sample for variances, Fourier Analysis, Histogram, Moving Average, Random Number Generation, Rank and Percentile, Regression, Sampling, t-test, z-test.

Module-5   **9 Hrs.**

Simulations : Simulations, Decision Trees and Forecasting, when should we use simulation, simulation modeling cycle, Introduction to Monte Carlo Simulation, generating random values, discrete and continuous functions, Excel for simple simulation, Managerial applications of risk analysis, performing a simulation using @Risk, analyzing the simulation output, generating various plots. Simulation in forecasting, Advanced simulation techniques.

Text Books:

1. Excel 2016 Bible, John Walkenbach, Wiley, 1st Edition, 2015.
2. Excel Data Analysis - Modeling and Simulation, Hector Guerrero, Springer, 2010 Edition, 2014.
3. Excel Functions and Formulas, Bernd Held, Theodor Richardson, BPB Publications, 3rd Edition, 2017.

Reference Books:

1. Microsoft Excel 2013, Data Analysis and Business Modeling: Winston, PHI, 2014 Edition, 2014.
2. Excel Data Analysis for Dummies, Stephen L Nelson, E C Nelson, Wiley, 2nd Edition, 2014.

Title	R Programming Language
Total Hours	60
Credits	03

Course Objectives:

- To understand the basic concepts of R programming language.
- To understand the data structures in R programming language.
- To understand the important packages and functions in R programming language.
- To understand the procedure for summary statistics and parametric testing of hypothesis using R programming Language.
- To understand the functions for graphs and non-parametric testing of hypothesis in R programming Language.

Course Outcomes:

On successful completion of the module students will be able to:

- Know the procedure to read and write different format of data set into R environment.
- Understand the uniqueness in R programming with the help of apply function in R programming language.
- Apply different options in I/O operations in R programming Language.
- Know the interpretation of summary statistics and testing of hypothesis.
- Know the built-in functions for graphs and non-parametric testing of hypothesis in R.

Module -1   **9 Hrs.**

Introduction to R Environment: History and development of R Statistical computing programming language, installing R and R studio, getting started with R, creating new working directory, changing existing working directory, understanding the different data types, installing the available packages, calling the installed packages, arithmetic operations, variable definition in R, simple functions, vector definition and logical expressions, matrix calculation and manipulation using matrix data types, workspace management, help function in R environment.

Rathinam College of Arts & Science (Autonomous), Coimbatore-21.
Admitted in M.SC DSBA from the academic year 2021 Onwards

Page 20 of 90
Regulations 2021

Module -2   **12 Hrs.**

Data Structures and Control Statements

Introduction to different data types, vectors, atomic vectors, types and tests, coercion, lists, list indexing, function applying on the lists, adding and deleting the elements of lists, attributes, name and factors, matrices and arrays, matrix indexing, filtering on matrix, generating a covariance matrix, applying function to row and column of the matrix, data frame – creating, coercion, combining data frames, special types in data frames, operations in data frame, applying functions: lapply( ) and sapply( ) on data frames, control statements, loops, looping over non vector sets, arithmetic and Boolean operators and values, branching with **if**, looping with **for**, **if-else** control structure, looping with **while**, vector based programming.

Module – 3   **12 Hrs.**

I/O operations and String Manipulations

Introduction to I/O functions in R, accessing I/O devices, using of scan( ), readline ( ) function, comparison and usage of scan and readline function, reading different format files into R: text file, CSV file, Statistical package files, xls and xlsx files, reading data frame files, converting from one format to another using in built function, writing different file format in to the local machine directory, getting file directory information, accessing the internet : overview of TCP/IP, sockets in R, basics of string manipulations – **grep ( ), nchar ( ), paste( ), sprintf( ), substr( ), regexpr( ), strsplit( ),** testing of file name with given suffix.

Module-4   **12Hrs.**

R for Summary Statistics and Parametric Tests

Descriptive statistics – summary statistics for vectors, making contingency tables, creating contingency tables from vectors, converting objects in to tables, complex flat tables, making 'Flat' contingency tables, testing tables and flat table objects, cross tables, testing cross tabulation, recreating original data from contingency tables, switching class, mean (arithmetic, geometric and harmonic), median, mode for raw and grouped data, measure of dispersion – range, standard deviation, variance, coefficient of variation, testing

of hypothesis – small sample test, large sample test – for comparing mean, proportion, variance (dependent and independent samples), correlation and regression – significance of correlation and regression coefficients

Module-5   **15 Hrs.**

R for Graphs, Nonparametric Tests and ANOVA

Introduction to graphs, Box-Whisker Plot, Scatter plots, pairs plots, line chart, Pie Chart, Cleveland Dot Charts, Bar Charts, Customization of charts, non-parametric test: The Wilcoxon U-Test (Mann-Whitney): One and Two-Sample U-Test, Tests for association: Chi Square Tests, Monte Carlo simulation, Yates Correction for 2X2 Tables, single category goodness of fit tests, Analysis of Variance for one-way variation and two variation – with and without interaction,

Text Books:

1. The art of R programming – Norman Matloff, no starch Press, San Francisco.
2. R in Action – Robert I. Kabacoff, Second Edition, Dreamtech Press.

Reference Books:

1. Introduction to Scientific Programming and Simulation using R – Owen Jones, Robert Maillardet and Andrew Robinson, CRC Press
2. Advanced R – Hadley Wickham, CRC Press.

Title	Business Economics
Total Hours	45
Credits	03

Course Objectives:

- To equip the students of management with time tested tools and techniques of business economics to enable them to appreciate its relevance in decision-making.
- To explore the economics of information and network industries and to equip students with an understanding of how economics affect the business strategy of companies in these industries.

Course Outcomes:

On successful completion of the module students will be able to:

- It equip the students to Design competition strategies, including costing, pricing, product differentiation, and market environment according to the natures of products and the structures of the markets.
- It develop economic way of thinking in dealing with practical business problems and challenges.

Module -1   **9 Hrs.**

Basic Concepts of Economics: Introduction to Economics , Basic Economic Problem, Circular Flow of Economic Activity , Nature of the firm - rationale, objective of maximizing firm value as present value of all future profits, maximizing, satisficing, optimizing, principal agent problem, Accounting Profit and Economic Profit , Role of profit in Market System , Adam Smith and Invisible Hand.

Module -2   **9 Hrs.**

Demand Analysis and Forecasting: Determinants of Market Demand at Firm and Industry level – Elasticity of Demand - Market Demand Equation – Use of Multiple Regression for estimating demand – Case study on estimating industry demand (formulating equation and solving with the aid of software expected)

Demand and Supply: Market Equilibrium – Pricing under perfect competition, monopolistic competition, Case study on pricing under monopolistic competition, Oligopoly - product differentiation and price discrimination; price- output decision in multi-plant and multi-product firms.

Module – 3   **9 Hrs.**

Cost Concepts: Cost Concept, Opportunity Cost, Marginal, Incremental and Sunk Costs, Cost Volume Profit Analysis, Breakeven Point, Case Study on marginal costs.

Module-4   **9 Hrs.**

Risk Analysis and Decision Making: Concept of risk, Expected value computation, Risk management through Insurance, diversification, Hedging, Decision Tree Analysis, CaseStudy on Decision tree technique.

Module-5   **9 Hrs.**

Monetary and Fiscal Policy: Monetary and fiscal policy, Role of Fiscal and Monetary Policy, Money Markets, Concept of savings and investment, Business cycles , National income accounting concepts, Commercial banks and the central bank money and credit, Financial markets and asset prices

Text Books:

1. Managerial Economics,   by Peterson, Lewis, Sudhir Jain, Pearson, Prentice Hall Indian Economy by Datt&Sundaram 61st Edition, S Chand

2. Managerial Economics by D. Salvatore, McGraw Hill, New Delhi.

3. Thomas Sowell, "Economics – A Common Sense Guide to the Economy", Basic Books Publishers, ISBN 978-0-465-05684-2.

Reference Books:

1. Managerial Economics by Varshney and Maheshwari, Sultan Chand and Sons, New Delhi.

2. Managerial Economics by Dr. D. M. Mithani, Himalaya Publishing House

3. Managerial Economics by Joel Dean Prentice Hall, USA.

4. Managerial Economics by H L Ahuja S Chand & Co. New Delhi

Title	Database Management System
Total Hours	60
Credits	03

Course Objectives:

In this course students will be able to understand:

- Database design
- File systems
- Various database management systems
- Database design models and relational model
- ER Diagram
- SQL and SQL statements
- Database physical and logical design
- Transactions management

Course Outcomes:

On successful completion of the module students will be able to:

- Explain database and database management system
- Explain physical and logical view of a database management system
- List various types of database management systems
- Design database using different database design models
- Create ER Diagram for a database
- Write SQL queries to manipulate database
- Describe ACID properties of transactions
- Implement transactions in a database.

Module -1   **9 Hrs.**

Overview of Database Systems: Introduction - Overview of Database Management - What is Database System - History of DBMS - Managing Structured Data - File Systems vs. DBMS - Basics of DBMS – DBMS Architecture -Overview of Relational Model - Database languages – Queries - Transaction Management - Structure & Design of a DBMS - Object Relational and semi-structured DB - Users & Administrators- Client/Server Architecture - Case Study.

Module -2   **12 Hrs.**

Database Design Models: The Relational Model - Relational Calculus - Introduction to Database Design - ER Diagrams – Entities, Attributes and Relationships. Design with ER Model - Conceptual Design for Large Enterprises - UML - Case Study**.**

Relational Model: The Relational Model Integrity Constraints - Key Constraints – Primary Key Constraints - Foreign Key Constraints - General Constraints - Relational Algebra- Selection and Projection- Set Operation - Relational Calculus - Tuple Relational Calculus- Domain Relational Calculus - Case Study.

Module – 3   **12 Hrs.**

Schema Refinement and Normal Forms : DB Design - Normal forms and Atomic Domain- Functional Dependencies and Decomposition - Database Design Process

SQL: SQL queries – Union – Intersect - and Except - NestedQueries – AggregateQueries- Null values- Joins – Views - Stored Procedures - User defined Functions – Triggers – Transactions - Case Study

Module-4   **15 Hrs.**

DB Application Development: DB Access from applications – embedded SQL, Cursors, and Dynamic SQL. Introduction to JDBC & SQL/J - Stored Procedures.

Overview of Storage and Indexing: Data on external storage - File Organizations and Indexing - Index Data Structures - Comparison of File Organizations - Indexes and

Performance Tuning.

Overview of Query Evaluation: System Catalog - Operator Evaluation - Algorithms for relational operations. Introduction to Query Optimization – Alternative Plans - Case Study.

Module-5   **12 Hrs.**

Transaction Management: Introduction to Transaction - ACID Properties Serializability-Transactions and Schedules - Concurrent Execution of Transactions - Lock-based concurrency control - Transaction support in SQL commit - rollback – save point - Introduction to Crash Recovery.

Physical Database Design and Tuning: Introduction to Physical Database design - Index Selection - Clustering. Overview of Database Tuning - Choices in tuning queries and Views - Case Study

Text Books:

1. Database Management Systems, Raghu Ramakrishnan and Johannes Gehrke 3rd Edition, McGraw Hill 2003.
2. **Database System Concepts, AbrahamSilberschatz, Henry F.Korth and S.Sudarshan, 5th Edition, McGraw Hill 2006.**

Reference Books:

1. Fundamentals of Database Systems, Elmasri and Navathe, 5thEdition, Addison-Wesley, 2007.

2. An Introduction to Database Systems, C.J. Date, A. Kannan, S. Swamynatham, 8th Edition, Pearson education, 2006.

Title	R Programming Language Lab
Credits	03

Programs:

Exercise -1

1. Install and configure R, set working directory.
2. Install Packages and calling installed packages
3. R studio environment and functionalities of R studio
4. Implement basic R operations (data input, missing values, importing data into R using different formats : xlsx, CSV, Text files)
5. Use R as a calculator
6. Explore various functionalities of dataframes.
7. Create data set using data frames, list and tables.
8. Calculate the remainder after dividing 31079 into 170166719
9. Calculate the interest earned after 5 years on an investment of $2000,
10. Assuming an interest rate of 3% compounded annually.
11. Use R to calculate the area of a circle with radius 7 cm.
12. Do you think there is a difference between 48:14^2and 48:(14^2)?
13. Usingrep()and seq()as needed, create the vectors?
    000001111122222333334444 and 1234512345123451234512345
14. Create the vector
    ## [1]00011110001111000111100011110000111100011
    ## [34] 1 1
    and convert it to a factor. Identify the levels of the result, and then
    change the level labels to obtain the factor:
    ## [1] Male MaleMale Female FemaleFemaleFemale Male Male
    ## [10] Male Female FemaleFemaleFemale Male MaleMale Female
    ## [19] Female FemaleFemale Male MaleMale Female FemaleFemale
    ## [28] Female Male MaleMale Female FemaleFemaleFemale
    ## Levels: Male FemaleExplore various functionalities of plots

Exercise -2

1. Create the contingency table for the given raw data.

2. Create the interactive user input code line in r using readline ( ) function.

3. Create the contingency table for the given vector format data.

4. Convert the contingency table to original format of the given data.

5. Analyse and give interpretation of summary statistics for the given data.

6. Calculate mean, median and mode for the grouped data and compare the results for the given data.

7. Analyse the given data for non-parametric tests and give the interpretations.

8. Use R for test the given data

   In order to compare the effectiveness of two sources of nitrogen, namely ammonium chloride (NH4Cl) and urea, on grain yield of Coarse cereal, an experiment was conducted. The results on the grain yield of Coarse Cereal (kg/plot) under the two treatments are given below.

NH4Cl :   13.4, 10.9, 11.2, 11.8, 14.0, 15.3, 14.2, 12.6, 17.0, 16.2, 16.5, 15.7.

Urea :   12.0, 11.7, 10.7, 11.2, 14.8, 14.4, 13.9, 13.7, 16.9, 16.0, 15.6, 16.0.

   Assess which source of nitrogen is better for Coarse Cereal.

9. Use R to test the given data and interpret the results.

   In a health survey of school children, it is found that the mean hemoglobin level of 55 boys is10.2 per 100ml with a SD of 2.1. Can we consider this group as taken from a population with a mean of 11.0 g/100ml?

10. In a hearing survey among 246 town school children, 36 were found with conductive hearing loss and among 349 village school children 61 were found with conductive hearing loss. Does this present any evidence that conductive hearing loss is as common among town children as among village children?

11. In an experiment to compare two types of Goat foods A and B, the following results of increase in weight were observed in Goats.

Pig No.	1	2	3	4	5	6	7	8
Increase in weight due to A	49	53	51	52	47	50	52	53

Increase in weight due to B	52	55	52	53	50	54	54	53

Assuming the two samples are independent can we conclude food B is better than food A?

12. Before an increasing in exercise duty on tea, 800 persons out of a sample of 1000 persons were found to be tea drinkers. After an increasing in duty, 800 people were tea drinkers in a sample of 1200 people. Using SE of a proportion, state whether there is a significant decrease in consumption of tea after the increase in the exercise duty.

13. Use R for test the given data

A health status survey in a few villages revealed that the normal serum protein value of children in that locality is 7.0 g/100ml. A group of 16 children who received high protein food for a period of six months had serum protein values shown below. Can we consider that the mean serum protein level of those who were fed on high protein diet is different from that of the general population?

S.No. (Child No.)	1	2	3	4	5	6	7	8
Protein level (g%)	7.10	7.70	8.20	7.56	7.05	7.08	7.21	7.25

S.No. (Child No.)	9	10	11	12	13	14	15	16
Protein level (g%)	7.36	6.59	6.85	7.90	7.27	6.56	7.93	8.56

14. Students were selected to training. Their performance was noted by giving a test and the marks recorded out of 50. They were given effective 6 months training and again they were given a test and marks were recorded out of 50.

Farmers	1	2	3	4	5	6	7	8	9	10
Before training	25	20	35	15	42	28	26	44	35	48
After training	26	20	34	13	43	40	29	41	36	46

By applying the t-test can it be concluded that the students have benefited by the training?

15. 100 individuals of a particular race were tested with an intelligence test and classified into two classes. Another group of 120 individuals belong to another race were

administered the same intelligence test and classified into the same two classes. The following are the observed frequencies of the two races:

Race	Intelligence		
	Intelligent	Non-intelligent	Total
Race I	42	58	100
Race II	55	65	120
Total	97	123	220

Test whether the intelligence is anything to do with the race.

16. Obtain the correlation coefficient between the heights of father(X) and of the son (Y) from the following data

X	65	66	67	68	69	70	71	72
Y	67	68	65	68	72	72	69	71

And also test its significance. Using R functions.

17. Analyse the given data for analysis of variance and interpret the same for all the possible values.

18. Consider the inbuilt data set cars.

   Find Correlation between possible variables and pairwise correlation

   Find regression line between appropriate variables

   Display the summary statistics and comment on the results

Title	Database Management System Lab
Credits	03

List of Experiments:

Tables to be used in the exercise:

Student (Student_id, Sname, DepNo, email)

CollDept (DeptNo, Dname,HOD)

Faculty(Faculty_id,fname,dept,designation,salary)

Employee (EmpID, name, job, hiredate, sal, deptno,MgrID,age)

Department (deptno, dname, loc)

a. Create :

   a) Above given tables with appropriate data types and constrains
   b) A query to display the name and age of students, whose age is more than 30 years.

   c) Display the employee number, name, salary, and salary increased by15%.
   d) A query that displays the names and indicates the amounts of their annual salaries with asterisks. Sort the data in descending order of salary.

b. Retrieve data from multiple tables by using all types of below specified joins:

   Inner Join:

   a) Retrieve only the information about departments to which at least one employee is assigned.

   b) Retrieve only the information about those employees who are assigned to a department.

   c) Retrieve the information of all the employees along with their Department Name if they are assigned to any department.

   d) Retrieve the information of all the departments along with the detail of EmployeeName belonging to each Department, if any is available.

   e) Retrieve the ID of the employees and the IDs of their respective managers from the employee table.

c. Using group functions, retrieve suitable results by using HAVING, GROUP BY and ORDER BY clauses

   a) Find the age of highest paid employee who is at least 30 years old for each department with at least two such employees.

   b) Find those departments for which the average age of employee is the minimum over all departments.

    c) Find the sum of salary of all the employees in a each department having department number greater than 10.

d. Write a query that displays the names with the first letter capitalized and all other letters lowercase, and the length of the names, for all names whose name starts with J, A, or M. Give each column an appropriate label. Sort the results by the names.

e. Perform below specified DDL operations:

    a) Create another table by name NEWDEPT from DEPT table's deptno, dname

        a. Columns and another column by name dept_head.

    b) Apply all the given constraints properly.

        Perform below specified operations with these tables.
          1. Rename the table.

          2. Add one column (sex) to that table which contains either M or F.

          3. Drop column from the table.

          4. Drop the table.

f. Create one simple view on students which contains all students records belong to department MCA.

g. Try to insert one row data through the view and verify it in the base table.

h. Create a sequence and insert data to students using that sequence for providing new students number.

i. Create a PL/SQL block that selects the maximum department number in the DEPARTMENTS table and stores it in a variable. Print the results to the screen.

j. Create a PL/SQL block to insert a new department number into the Departments table. Use maximum dept number fetched from above and adds 10 to it.

k. Create a PL/SQL block to delete the department created in exercise 16.   Print to the screen the number of rows affected.

l. Write a PL/SQL block which accepts employee name, basic and should displayEmployee name, PF and net salary.

    HRA=30% of basic salary

    DA=75% of basic salary

    Net salary=basic+HRA+DA-PF

    If the basic is less than 8000 PF is 5% of basic salary.

If the basic is between 8000 and 15000 PF is 7% of basic salary.

If the basic is between 15000 and above PF is 8% of basic salary.

m. Write a PL/SQL block to award an employee with the bonus. Bonus is 15% of commission drawn by the employee. If the employee does not earn any commission then display a message that 'employee does not earn any commission'. Otherwise add bonus to the salary of the employee. The block should accept an input for the employee number.

n. Write a PL/SQL block which accepts employee number and finds the average salary of the employees working in the department where that employee works. If his salary is more than the average salary of his department, then display messaging that 'employee's salary is more than average salary' else display 'employee's salary is less than average salary'.

o. Using Cursors

a) Write a program that gives all employees in department MCA and 15% pay increase. Display a message displaying how many Employees were awarded the increase.

b) Display the names of employees who are working for Department MCA.

p. Create a procedure that deletes rows from the employee table. It should accept 1 parameter, job; only delete the employee's with that job. Display how many employees were deleted.

q. Write a PL/SQL block to invoke the procedure. Write a simple before statement-level trigger that displays a message prior to an insert operation on the EMP table.

r. Write an after statement-level trigger. Whenever an insert, update, or delete operation occurs on the EMP table, a row is added to the empauditlog table recording the date, user, and action.

s. Write a row-level trigger that calculates the commission of every new employee belonging to department 30 that is inserted into the EMP table.

t. Design mini database project using procedures and triggers. Front-end can be done using Java or ASP.Net or any other similar languages.

Rathinam College of Arts & Science (Autonomous), Coimbatore-21.
Admitted in M.SC DSBA from the academic year 2021 Onwards

Page 34 of 90
Regulations 2021

SEMESTER -2
Syllabus

Rathinam College of Arts & Science (Autonomous), Coimbatore-21.
Admitted in M.SC DSBA from the academic year 2021 Onwards

Page 35 of 90
Regulations 2021

Title	Big Data Analytics
Total Hours	45
Credits	03

Course Objectives:

- Introduce the concepts of big data analytics to the students.
- Introduces the big data framework - its characteristics and use cases associated with it.
- A concise introduction to Hadoop framework will prepare students to handle industry scenarios of big data analytics.

Course Outcomes:

On successful completion of the module students will be able to:

- Understand the characteristics of big data.
- Explore Hadoop framework and its components.
- Use HDFS and Map Reduce to analyze various industry use cases of big data analytics.
- Understand the YARN Infrastructure.
- Learning different Sorting - Shuffling.

Module -1   **9 Hrs.**

Introduction to Big Data: Defining Data, Types of Data, Structured Data, Semi Structured Data, Unstructured Data, How data being Generated, Sources of Data Generation, Rate at which Data is being generated, Different V's, Volume, Variety, Velocity, Veracity, Value, How single person is contributing towards Big Data, Significance for Big Data, Reason for Big Data, Understanding RDBMS and why it is failing to store Big Data. Future of Big Data, Big Data use cases for E-Commerce Industry, Banking Sector, Medical Sector, Entertainment Sector, Fraud Detection, Security Intelligence, 360 degree View of Customer, Internet of Things

Module -2   **9 Hrs.**

Introduction to Hadoop: Introduction – Why Hadoop – What is Hadoop – History and Milestone of Hadoop – Core Components of Hadoop – Difference between Regular File System and HDFS – Common Hadoop Shell Commands – Hadoop Configuration.

Module -3   **9 Hrs.**

Hadoop Distributed File System (HDFS): Hadoop Distributed File System, Design of HDFS, HDFS Concept, How files are stored in HDFS, Hadoop File system, Replication factor, Name Node, Secondary Name Node, Job Tracker, Task tracker, Data Node, FS Image, Edit-logs, Check-pointing Concept, HDFS federation, HDFS High availability, Architectural description for Hadoop Cluster, When to use or not to use HDFS, Block Allocation in Hadoop Cluster, Read operation in HDFS, Write operation in HDFS, Hadoop Archives, Data Integrity in HDFS, Compression & Input Splits, Advantages of Hadoop, Unix Shell commands and HDFS commands

Module – 4   **9 Hrs.**

Processing Unit: What is MapReduce - History of MapReduce - How does MapReduce works - Input files - Input Format types Output Format Types - Text Input Format - Key Value Input Format - Sequence File Input Format - Input split - Record Reader - MapReduce overview - Mapper Phase - Reducer Phase - Sort and Shuffle Phase - Importance of MapReduce

Data Flow - Counters - Combiner Function - Partition Function - Joins - Map Side Join - Reduce Side Join - MapReduce Web UI - Job Scheduling - Task Scheduling - Fault Tolerance - Writing MapReduce Application - Driver Class - Mapper Class - Reducer Class - Serialization - File Based Data Structure - Writing a simple MapReduce program to Count Number of words - MapReduce Work Flows

Module-5   **9 Hrs.**

Introduction to YARN and Apache Hive:YARN: YARN Architecture, YARN Components, Resource Manager, Node Manager, Application Master, Difference between Hadoop 1.x and 2.x Architecture

Apache Hive: Apache Hive, Features of Apache Hive, Command Line Interface, History of Apache Hive, Hive Data Types & Files Formats, Creating Managed Table, External Table, Partitioned Table, Dropping Tables, Alter Table, Loading data into Managed Table, Inserting Data into Tables from Queries, Dynamic Partitions inserts, Exporting data, SELECT from clauses, WHERE Clauses, GROUP BY Clauses, JOIN Statements, ORDER BY, SORT BY, DISTRIBUTE BY, CLUSTER BY, bucketing, UNION ALL, View, Hive Metastore

Text Books:

1. Seema Acharya (Author) - SubhashiniChellappan - Big Data and Analytics (2015).

Wiley Publication.
2. Data Science and Big Data Analytics: Discovering - Analyzing - Visualizing and Presenting Data (2015) - EMC Education Services

Website:

1. http://hadoop.apache.org/docs/current/
2. https://hadoop.apache.org/docs/r1.2.1/mapred_tutorial.html
3. https://cwiki.apache.org/confluence/display/Hive/LanguageManual

Title	Machine Learning
Total Hours	45
Credits	03

Course Objectives:

- Understand the underlying concepts about Machine Learning.
- To understand the importance of Linear Algorithms
- Understand the concepts of Non-Linear Algorithms.
- To understand the Data Preprocessing and Dimensionality Reduction, Clustering Analysis
- Understand the concept of Time Series Analysis and Forecasting.

Course Outcomes:

On successful completion of the module students will be able to:

- Examine with many of the ML introduction Concepts.
- Demonstrate about importance of Linear Algorithms
- Demonstrate about importance of Non-Linear Algorithms
- Examine with Data Preprocessing, Clustering Analysis
- Experiment with Time Series Analysis and Forecasting

Module -1   **9 Hrs.**

Introduction to Machine Learning: Supervised Machine Learning, Unsupervised Machine Learning, Semi-Supervised Machine Learning, Bias Error, Variance Error, Bias-Variance Trade-Off , Statistical Fit, Overfitting in Machine Learning, Underfitting in Machine Learning, A Good Fit in Machine Learning, How to Limit Overfitting.

Module -2   **9 Hrs.**

Linear Algorithms: Linear Regression, Logistic Regression, Gradient Descent for Machine Learning, Batch Gradient Descent, Stochastic Gradient Descent, Linear Regression Using Gradient Descent, Logistic Regression by Stochastic Gradient Descent, Linear Discriminant Analysis, Representation of LDA Models, Making Predictions with LDA, Preparing Data For LDA

Module – 3   **9 Hrs.**

Non-Linear Algorithms: Classification and Regression Trees, Decision Trees, CART Model Representation, Naive Bayes, K-Nearest Neighbors, Learning Vector Quantization, Support Vector Machines, Maximal-Margin Classifier, Soft Margin Classifier, Support Vector Machines (Kernels), Training SVM With Gradient Descent, Learn an SVM Model from Training Data,Make Predictions with SVM Model

Module-4   **9 Hrs.**

Data Preprocessing and Dimensionality Reduction, Clustering Analysis: Dealing with missing data ,Handling categorical data, Partitioning a dataset in training and test sets, Bringing features onto the same scale , Selecting meaningful features , Assessing feature importance with random forests, Unsupervised dimensionality reduction via principal component analysis , Supervised data compression via linear discriminant analysis , Using kernel principal component analysis for nonlinear mappings, Grouping objects by similarity using k-means ,Organizing clusters as a hierarchical tree ,Locating regions of high density via DBSCAN.

Module-5   **9 Hrs.**

Time Series Analysis and Forecasting: Time Series Patterns, Horizontal Pattern, Trend Pattern, Seasonal Pattern, Forecast Accuracy,

Moving Averages and Exponential Smoothing, Moving Averages, Weighted Moving Averages, Exponential Smoothing, Trend Projection, Linear Trend Regression, Holt's Linear Exponential Smoothing, Nonlinear Trend Regression, Seasonality and Trend, Seasonality Without Trend, Seasonality and Trend, Models Based on Monthly Data, Time Series Decomposition, Calculating the Seasonal Indexes, Deseasonalizing the Time Series, Using

the Deseasonalized Time Series to Identify Trend.

Text Books:

1. Master Machine Learning Algorithms by Jason Brownlee
2. Statistics for Business and Economics by Anderson, Sweeney and Williams
3. Deeper Insights into Machine Learning by BIRMINGHAM(Packt)

Reference Books:

1. Using Multivariate Statistics - Barbara G. Tabachnick, Linda S. Fidell, Pearson
2. Introduction to Machine Learning - EthemAlpaydm, The MIT Press

Title	**Exploratory Data Analysis**
Total Hours	**45**
Credits	**03**

Course Objectives:

- To understand importance of data and its types in Exploratory Data Analysis.

- To understand difference between EDA and summary statistics in context of interpretation.

- To understand the importance of data pre-processing for Exploratory Data Analysis.

- To understand the significance of missing value imputations in better EDA interpretations.

- To understand the importance measure of central tendency in describing the quick view of data set.

- To understand the importance of measure of dispersion and its interpretation in spread ness of data

Course Outcomes:

On successful completion of the module students will be able to:

- Understand the data and its types for the appropriate exploratory data analysis.
- Understand the importance of Exploratory Data Analysis over summary statistics.
- Understand the importance Univariate statistics in EDA
- Plot Univariate statistical graphs for the better representation and interpretation.
- Plot bivariate statistical graphs for the better representation and interpretation.

Module -1   **12 Hrs.**

Introduction to Data and its types: Definition and importance of data, classification of data : based on observation – Cross Sectional, times series and panel data, based on measurement – ratio, interval, ordinal and nominal, based on availability – primary, secondary, tertiary, based on structural form – structured, semi structured and unstructured, based on inherent nature – quantitative and qualitative, concepts on sample data and population, small sample and large sample, statistic and parameter, types of statistics and its application in different business scenarios, frequency distribution of data.

Module -2   **12 Hrs.**

Introduction to Exploratory Data Analysis (EDA): Definition of EDA, difference between EDA with classical and Bayesian Analysis, comparison of EDA with Classical data summary measures, goals of EDA, Underlying assumptions in EDA, importance of EDA in data exploration techniques, introduction to different techniques to test the assumptions involved in EDA, role of graphics in data exploration, introduction to unidimensional, bidimensional and multidimensional graphical representation of data.

Module – 3	12 Hrs.

Data Preparation: Introduction to data exploration process for data preparation, data discovery, issues related with data access, characterization of data, consistency and pollution of data, duplicate or redundant variables, outliers and leverage data, noisy data, missing values, imputation of missing and empty places, with different techniques, missing pattern and its importance, handling non numerical data in missing places.

Module-4	12 Hrs.

Univariate Data Analysis: Description and summary of data set, measure of central tendency – mean: Arithmetic, geometric and harmonic mean – Raw and grouped data, confidence limit of mean, median, mode, quartile and percentile, interpretation of quartile and percentile values, measure of dispersion, concepts on error, range, variance, standard deviation, confidence limit of variance and standard deviation, coefficient of variation, mean absolute deviation, mean deviation, quartile deviation, interquartile range, concepts on symmetry of data, skewness and kurtosis, robustness of parameters, measures of concentration.

Module-5	12 Hrs.

Bivariate Data Analysis: Introduction to bivariate distributions, association between two nominal variables, contingency tables, Chi-Square calculations, Phi Coefficient, scatter plot and its causal interpretations, correlation coefficient, regression coefficient, relationship between two ordinal variables – Spearman Rank correlation, Kendall's Tau Coefficients, measuring association between mixed combination of numerical, ordinal and nominal variables.

Text Books:

1. Exploratory Data Analysis – John W Tukey, Addison Wesley Publishing Company
2. Exploratory Data Analysis in Business and Economics - An Introduction Using SPSS, Stata and Excel – Thomas Cleff, Springer Publication

Reference Books:

1. Graphical Exploratory Data Analysis - S.H.C. du Toit A.G.W. Steyn R.H. Stumpf, Springer Publication
2. Hand book of Data Visualization – Chun-houh Chen, Wolfgang Härdle, Antony Unwin, Springer Publication.

Title	Python Programming
Total Hours	45
Credits	03

Course Objectives:

- To understand the history and development of Python Programming Language.
- To understand the data structures and looping concepts in Python Programming Language.
- To understand the important packages and functions in Python Programming Language.
- To understand the importance of Python Programming Language in data wrangling or munging.
- To understand the impact of Python Programming Language in statistical analysis.

Course Outcomes:

On successful completion of the module students will be able to:

- Understand the core programming concepts of Python Programming Language.
- Know the Looping and condition statements in Python Programming Language
- Understand the different options in Data Management in Python Programming Language.
- Understand the importance of data transformation and its need in Python Programming Language
- Know elementary to advanced statistical methods in Python Programming environment.

Module -1   9 Hrs.

Introduction to Python Environment : History and development of Python, Why Python? Grasping Python's core philosophy, Discovering present and future development goals, Working with Python : Getting a taste of the language, Understanding the need for indentation, Working at the command line or in the IDE, Visualizing Power, Using the Python Ecosystem for Data Science, Accessing scientific tools using SciPy, Performing fundamental scientific computing using NumPy, Performing data analysis using pandas, Implementing machine learning using Scikit- learn, Plotting the data using matplotlib, Parsing HTML documents using Beautiful Soup, Setting Up Python for Data Science, Getting Continuum Analytics Anaconda, Getting Enthought Canopy Express, Getting pythonxy, Getting WinPython, Installing Anaconda on Windows, Linux and MAC

Module -2   9 Hrs.

Data Structures, Looping and Branching: Working with Numbers and Logic, Performing variable assignments, Doing arithmetic, Comparing data using Boolean expressions, Creating and Using Strings, Interacting with Dates, Creating and Using Functions, Calling functions in a variety of ways, Using Conditional and Loop Statements, Making decisions using the if statement, Choosing between multiple options using nested decisions, Performing repetitive tasks using for, Using the while statement, Storing Data Using Sets, Lists, and Tuples : Performing operations on sets, Working with lists, Creating and using Tuples, Defining Useful Iterators, Indexing Data Using Dictionaries.

Module – 3   9 Hrs.

Data Management : Working with Real Data, Uploading small amounts of data into memory, Streaming large amounts of data into memory, Sampling data, Accessing Data in Structured Flat- File Form, Sending Data in Unstructured File Form, Managing Data from Relational Databases, Interacting with Data from NoSQL Databases, Accessing Data from the Web, Juggling between NumPy and pandas, Validating Your Data, Removing duplicates, Manipulating Categorical Variables, Dealing with Dates in Your Data, Dealing with Missing Data, Slicing and Dicing: Filtering and Selecting Data, Concatenating and Transforming Working with HTML Pages, Working with Raw Text, Working with Graph Data.

Module-4   **9 Hrs.**

Data Transformation: Understanding classes in Scikit- learn, Playing with Scikit- learn, Defining applications for data science, Performing the Hashing Trick, Using hash functions, Demonstrating the hashing trick, Working with deterministic selection, Considering Timing and Performance, Benchmarking with timeit, Working with the memory profiler, Performing multicore parallelism, Demonstrating multiprocessing.

Module-5   **11 Hrs.**

Unit V: Python for Statistics: Exploring Data Analysis, The EDA Approach, Defining Descriptive Statistics for Numeric Data, Measuring central tendency, Measuring variance and range, Working with percentiles, Defining measures of normality, Counting for Categorical Data, Understanding frequencies, Creating contingency tables, Creating Applied Visualization for EDA, Inspecting boxplots, Performing t- tests after boxplots, Observing parallel coordinates, Graphing distributions, Plotting scatterplots, Using covariance and correlation, Using nonparametric correlation, Considering chi- square for tables, Using the normal distribution, Creating a Z- score standardization, Transforming other notable distributions, Detecting Outliers in Data, Clustering, Reducing dimensionality.

Text Books:

1. Python for Data Science for Dummies - Luca Massaron and John Paul Mueller, John Wiley & Sons, Inc.

Reference Books:

1. Python for Data Analysis - Wes McKinney, O'Reilly Media, Inc.
2. Data Science from Scratch - Joel Grus, O'Reilly Media, Inc.
3. Python Scripting for Computational Science - Hans PetterLangtangen

Title	Market Research and Analytics Seminar
Total Hours	45

Credits	**03**

Course Objectives:

- To familiarize students with a variety of popular techniques used in the collection and analysis of marketing research information and, within the constraints of the course, to develop their proficiency in their use and interpretation.
- Provide students with an understanding of what marketing research can and cannot realistically achieve for management decisions.

Course Outcomes:

On successful completion of the module students will be able to:

- Gain perspective and practice in applying techniques and interpreting findings.
- Develop, design and execute marketing research projects.
- Study emerging trends in marketing research
- Implement the process of research design through collection of data.

This class will consist of a lecture, discussion and hands on work. Students will also be responsible for a major research project. As such, a high degree of commitment, involvement and energy is critical to the successful completion of this course. Due to the nature of the marketing research process, the course may consist of a high degree of outside field work, computer analysis and group work. Students are required to read all assigned materials before coming to class. The instructor will typically give an overview of the assigned topic in the first portion of the class session. This part of the class seeks to stimulate discussion of key concepts as well as to provide a forum for the exchange of ideas among class members.

Any topic can be selected for Case Study not restricted to:

- Trade Analytics

- FMCG Marketing

- Industrial Products – B2B Marketing

Suggested Readings:

1. An Introduction to Marketing Research, Smith and Albaum (2010), Qualtrics Survey University

Title	Business Fundamental – II
Total Hours	45
Credits	03

Course Objectives:

- To provide students with an understanding of the principles of human behavior in organizations with relevance to the Indian business context.
- To provide students with an understanding of different organizational behavior concepts like motivation, communication, culture, human resource and conflict

Course Outcomes:

On successful completion of the module students will be able to:

- Equip the students to understand and implement different cultures, ethics and motivation required in organization management.
- Develop leadership, group behavior, reaction and handling situations during conflict in a professional way

Module -1 **9 Hrs.**

Introduction to OB: Introduction Concept of OB; Management roles, skills and activities; Disciplines that contribute to OB; Opportunities for OB – Globalization - Indian workforce diversity - customer service - innovation and change - networked organizations - work-life balance - people skill - positive work environment – ethics, Challenges for OB Manager,

Learning: Individual Behavior Learning, attitude and job satisfaction: Concept of learning, conditioning, shaping and reinforcement Concept of attitude, components, behavior and attitude Job satisfaction: causation; impact of satisfied employees on workplace Comparison of job satisfaction amongst Indian employees with other cultures

Module -2   **9 Hrs.**

Motivation: Concept, Theories (Hierarchy of needs, X and Y, Two factor, McClelland, Goal setting, Self-efficacy, Equity theory), Job characteristics model, Redesigning job and work arrangements, Employee involvement, Flexible benefits, Intrinsic rewards

Personality and Values: Concept of personality, MBTI, Big Five model. Relevance of values, Indian values, Linking personality and values to the workplace (person-job fit, person-organization fit)

Perception, Decision Making and Emotions: Perception and judgments, Factors, Linking perception to individual decision making, Decision making in organizations, Ethics in decision making, Emotional labour, Emotional Intelligence

Module – 3   **9 Hrs.**

Communication: Importance, Types, Barriers to communication, Communication as a tool for improving Interpersonal Effectiveness

Group Behavior : Groups and Work Teams, Concept, Five stage model of group development, Group think and shift, Indian perspective on group norms Groups and teams, Types of teams, Creating team players from individuals, Team building and team based work (TBW).

Leadership: Concept, Trait theories, Behavioral theories (Ohio and Michigan studies), Contingency theories (Fiedler, Hersey and Blanchard, Path-Goal), Authentic leadership, Mentoring, self-leadership, online leadership, Inspirational Approaches (transformational, charismatic), Comparison of Indian leadership styles with other countries. Exercises, games and role plays may be conducted to develop team and leadership skills.

Module-4   **9 Hrs.**

Organizational Culture: Organizational Culture and Structure Concept of culture, Impact (functions and liability), Creating and sustaining culture, Employees and culture, Creating positive and ethical cultures.

Concept of structure: Prevalent organizational designs, New design options.

Human Resource Management: Introduction to HRM, Selection, Orientation, Training & Development, Performance Appraisal, Incentives

Module-5 **9 Hrs.**

Decision Making: Decision Making, Process of Decision Making, Using data to make better decisions, data driven decision making, Shut down Decision, Make or Buy Decision, Joint Product Decisions, Product Mix Decisions, Replacement Decisions.

Text Books:

1. Organisational Behaviour by Stephen P. Robbins, Timothy A. Judge and Seema Sanghi, 13th Ed, Pearson Education ltd.

Reference Books:

2. Luthans Fred., "Organizational Behaviour", McGraw Hill.
3. Hellriegel, Slocum and Woodman, Organisational Behavior, South-Western, Thomson Learning, 9th edition, 2001.
4. BehaviorIn Organizations, Jerald Greenberg, 8th ed, Pearson Education.
5. Arnold, John, Robertson, Ivan t. and Cooper, Cary, l., "Work psychology: understanding human behavior in the workplace", Macmillan India Ltd., Delhi.
6. Dwivedi, R. S., "Human relations and organizational behaviour: a global perspective", Macmillan India Ltd., Delhi.
7. Jan Williams, "Financial and Managerial Accounting" – The basis for business decisions, Tata McGraw Hill Publishers
8. http://www.vssut.ac.in/lecture_notes/lecture1421946477.pdf

Rathinam College of Arts & Science (Autonomous), Coimbatore-21.
Admitted in M.SC DSBA from the academic year 2021 Onwards

Page 49 of 90
Regulations 2021

Title	Big Data Analytics Lab
Credits	03

Experiment 1:

Prepare list of software, infrastructure for setting up single node Hadoop cluster.

Experiment 2:

You need to perform 20 basics Linux commands on single node Hadoop cluster.

Experiment 3:

You need to perform 20 basics Hadoop commands on single node Hadoop cluster.

Experiment 4:

You need to program Mapper Class, Reducer Class and Driver Class for map reduce word count Job.

Experiment 5:

You need execute word count job based on 0 reducer, 2 reducer, Default reducer & 4 reducer and observe different outputs for word count job.

Experiment 6: (Below experiment should be implemented on Apache Pig Environment)

In this task you have 2 files named as Student and Results. You need to use PIG commands for this task.

Step1: Upload this file to Lab through winSCP.

Student: Contains names and roll number of students.

Results: Contains roll number and results of students whether they passed or failed.

Problem Statement: You need to print the name of all the students who failed or passed in the exam based on the given data.

(Faculty will share data with students)

Experiment 7:

Description: Georgia Salary/Travel data provided as CSV file with this assignment for the Fiscal Year 2010 and Organization Type of Local Boards of Education, produce a distinct list of all Job Titles along with the total number of employees aligned with each Job Title & the minimum/maximum/average salaries for each of the identified Job Titles

Expected Steps:

-Store the given input file salaryTravelReport.csv into the HDFS Location

- Load the salary file and declare its structure

- Loop through the input data to clean up the number fields. Take out the commas from the salary and travel fields and cast to a float

- Trim down to just Local Boards of Education

- Further trim it down to just be for the year in question

- Bucket them up by the job title

- Loop through the titles and check how many are there under each title

- Determine the minimum, maximum and average salaries for every title

- Guarantee the order on the way out

- Dump the results on the console

- Save results back to HDFS.

Experiment 8: (Below experiment should be implemented on Apache Hive Environment)

The dataset provided - MovieLens data sets are collected by the GroupLens Research Project at the University of Minnesota. It represents users' reviews of movies.

This data set consists of:

* 100,000 ratings (1-5) from 943 users on 1682 movies.

* Each user has rated at least 20 movies.

* Simple demographic info for the users (age, gender, occupation, zip)

u.data

-- The full u data set, 100000 ratings by 943 users on 1682 items.

Each user has rated at least 20 movies.

Users and items are numbered consecutively from 1.

The data is randomly ordered.

This is a tab separated list:

user id | item id | rating | timestamp

The time stamps are Unix seconds since 1/1/1970 UTC

u.user

-- Demographic information about the users;

This is a tab separated list:

user id | age | gender | occupation | zip code

The user ids are the ones used in the u.data data set.

Find the below problemstatement:

1. Create au_data table.

2. See the field descriptions of u_data table.

3. Load data into u_data table from a local text file.

4. Show all the data in the newly created u_data table.

5. Show the numbers of item reviewed by each user in the newly created u_data table.

6. Show the numbers of users reviewed each item in the newly created u_data table.

Experiment 9

You need to follow Experiment Number 09 and find out solution for below sets of problem:

1. Create au_user table.

2. See the field descriptions of u_user table.

3. Load data into u_user table from a local text file.

4. Show all the data in the newly created user table.

5. Count the number of data in the u_user table.

6. Count the number of user in the u_user table genderwise.

7. joinu_data table and u_user tables based on userid and show the top 10 results.

Experiment 10: File Management tasks in Hadoop

- Create a directory in HDFS at given path(s).
- List the contents of a directory.
- Upload and download a file in HDFS.
- See contents of a file.
- Copy a file from source to destination.
- Copy a file from/ To local file system to HDFS
- Move file from source to destination.
- Remove a file or directory in HDFS.
- Display last few lines of a file
- Display the aggregate length of a file.

Title	Python Programming Lab
Credits	03

List of Experiments:

1. Write a python program to find biggest number among four numbers using if-else
2. Write a python program to find given number is prime or not
3. Write a python program to find given number is palindrome or not
4. Write a python program to print multiplication table of given number
5. Write a python program to find mean of a n numbers using list
6. Write a python program to find given number is exist or not in the list, if exists print all its places
7. Write a python program to return sum of n numbers from a function using list
8. Write a python program to manipulate student details using dictionary and lists
9. Write a python program to return student details from a function using list and sno as parameter

10. Write a python program to manipulate employee details using classes and objects
11. Write a python program to read and write student details from and to a file using IO
12. Write a python program to read content from student.csv file and find total number of students, maximum and minimum marks

Title	Machine Learning Lab
Credits	03

List of Experiments:

1. Product review prediction analysis for genuine ratings using ML.
2. Stock Market Analysis and Prediction using Time Series Analysis
3. Artificial Intelligence Dietician using ML
4. House Price prediction
5. Online AI Shopping with E-wallet system
6. Personality prediction using resume analysis
7. Lung Disease Predictions Using ML
8. Bitcoin Price Prediction
9. To Find Air Quality Index
10. Customer Segmentation for Banking Domain
11. Credit Card Fraud Detection

Rathinam College of Arts & Science (Autonomous), Coimbatore-21.
Admitted in M.SC DSBA from the academic year 2021 Onwards

Page 54 of 90
Regulations 2021

12. Air Passengers Classification

SEMESTER -3
Syllabus

Title	Financial Econometrics
Total Hours	45
Credits	03

Course Objectives:

- To understand the basic concepts of time series analysis
- To understand the elementary time series models and model evaluation techniques
- To understand the integration process of non-stationary data set
- To understand the importance of ARMA and ARIMA models for forecasting
- To understand the basic concepts and estimation procedure for VAR models
- To understand the method to select the appropriate number of order of variables
- To understand the VECM model for Cointegrated series of variables
- To understand the ARCH and GARCH models
- To understand the basics of multivariate time series analysis techniques

Course Outcomes:

On successful completion of the module students will be able to:

- Understand the different elementary models related to time series analysis.
- Apply different model evaluation technique to identify better model to forecast.
- Understand the importance of stationarity in building time series models.
- Understand the use of Granger Causality and JohensenCointegration method.
- Apply VAR model to the dynamic behaviour of financial time series conditions.
- Select the order of Vector Auto Regression model for better forecast of time series data.
- Apply VECM in the appropriate place to overcome the Cointegration problem.
- Build the model using ARCH and GARCH technique for non-constant variance data.

Module -1   **9 Hrs.**

Time Series in Financial Econometrics: Introduction to time series plot in history, time series data and cross sectional data, difference between time series and cross sectional data, time series and stochastic process, means, variances, covariance, stationarity, importance of stationarity in time series analysis, components of time series analysis: trend, seasonal, cyclical and irregular, white noise process, random walk, elementary time series models with zero mean, model evaluation techniques: Bias, MAD, MSE, MAPE.

Module -2   **9 Hrs.**

Univariate time series analysis – I: Models related to stationary data, Auto Regressive model, Moving Average model, Stationarity of data, concepts on unit root, impacts of unit root in estimating the model parameters, tests related to unit root: Dickey Fuller test, Augmented Dickey Fuller test, KPSS Test, The Phillips Peron Test, seasonal unit roots, periodic integration and unit root testing.

Module – 3   **9 Hrs.**

Univariate time series analysis – II: ARMA (p,q) process, ACF (Auto Correlation Function) and PACF (Partial Auto Correlation Function) of an ARMA (*p,q*) process, forecasting ARMA process, integration of non-stationary data, first order integration and second order integration, ARIMA (*p,i,q*), estimation of parameters of ARIMA model, Wald Test Statistic for significance of coefficients.

Module-4   **9 Hrs.**

Spectral Analysis: Spectral densities, periodogram, he Spectral Representation and Spectral Distribution, Sampling Properties of the Sample Spectral Density, time invariant linear filters, the spectral density of ARMA (Auto Regressive Moving Average), smoothing the Spectral Density, Bias and variance, bandwidth, Confidence Intervals for the Spectrum, Leakage and Tapering, auto regressive spectrum estimation.

Module-5   **9 Hrs.**

Multivariate Time Series Analysis - VAR Estimation : Introduction to multivariate time series analysis, Concepts of Vector Auto Regression, multivariate least square estimation, asymptotic properties of Lease square estimation, Introduction to Vector Error Correction Models, Cointegrated Processes (Johensen Co-integration technique), Common Stochastic Trends, Deterministic Terms in Cointegrated Processes, Forecasting Integrated and Cointegrated Variables, Introduction to Univariate GARCH models, multivariate GARCH, estimation of GARCH models.

Text Books:

1. Introductory Econometrics A modern Approach - Jeffrey M. Wooldridge, South-Western Cengage Learning.
2. Basic Econometrics, Fifth Edition - Damodar N. Gujarati, Dawn C. Porter, McGraw-Hill/Irwin Publication.

Reference Books:

1. Introduction to Time Series and Forecasting– Peter J. Brockwell Richard A. Davis, Springer
2. Time Series Analysis with applications in R - Jonathan D. Cryer • Kung-Sik Chan, Second Edition, Springer
3. New Introduction to Multiple Time Series Analysis, Helmut Lütkepohl, Springer

Title	Data Visualization
Total Hours	**45**
Credits	**02**

Course Objectives:

- To understand the importance of data visualization in the business and engineering, application and role of visualization tools in creating the advanced techniques., importance of vector visualization in multivariate data analysis and the impact of visualization techniques in ease the analytics decision-making process

Course Outcomes:

On successful completion of the module students will be able to:

- Explain the application of different visualization techniques for different business representation.
- Apply and use the static graphical technique for the better understanding about the business problem.
- Apply and use multivariate statistical graphical technique for testing the assumptions of multivariate data.
- Apply and use graphical validation techniques for better model fit.
- Use the customization in graphical representation for user friendly graphics.

Rathinam College of Arts & Science (Autonomous), Coimbatore-21.
Admitted in M.SC DSBA from the academic year 2021 Onwards

Page 59 of 90
Regulations 2021

Module -1   **9 Hrs.**

Introduction to Data Visualization: Brief history of data visualization, scientific design choices in data visualization- choice of graphical form, grammar of graphical techniques of large amount of data, crucial need of visualization techniques, challenges in visualization techniques, classification of visualization techniques for qualitative and quantitative data, power of visualization techniques, introduction to different visualization techniques.

Module -2   **9 Hrs.**

Static Graphical Techniques: Introduction to bar graph, basic understanding of making basic bar graph, grouping bars together, bar graphs on counts, customization of bar graphs by changing colour, size, title, axis units, changing width and spacing of the bar chart, adding labels to bar graph, application of bar graph in business.

Module – 3   **9 Hrs.**

Multivariate Graphical Techniques: Introduction to correlation matrix, application of correlation matrix in the multivariate analysis, network graph, basics of heat map, difference between heat map and tree map, introduction to higher dimensional scatter plot, axis adjustment in the higher dimensional scatter plot, addition of prediction surface of higher dimensional scatter plot.

Module-4   **9 Hrs.**

Graphical Validation: Basics of multivariate statistical visual representations and its results, dendrogram, importance of dendrogram in grouping (cluster analysis), Scree Plot, importance of Scree Plot, application of Scree Plot in determining number of clusters and factors, QQ plot, importance of QQ plot in distribution of data for the further quantitative analysis, PP plot, applications and usage of PP Plot for distribution detection.

Module-5	**9 Hrs.**
Customization: Introduction to annotations – adding : text, mathematical expression , lines, arrows, shaded shapes, highlighting the texts and items, adding error bars, swapping x and y axis, changing the scaling, positioning of axis and arranging tick marks and labels, appearance of axis labels, circular graphs, using themes, theme elements, creating the own themes, legends : removing the legends, position of legends, legend title, labels in legends.	

Text Books:

1. DATA VISUALIZATION PRINCIPLES AND PRACTICE, SECOND EDITION - AlexandruTelea, CRC Press.
2. Hand book of Data Visualization – Chun-houh Chen, Wolfgang Härdle, Antony Unwin, Springer Publication.

Reference Books:

1. R Graphics Cook Book, Winston Chang, First Edition, O'Reilly Publication.
2. ggplot2 Elegant Graphics for Data Analysis – Hadley Wickham, Springer Publication

Title	**Operations Research**
Total Hours	**60**
Credits	**03**

Course Objectives:

- Understand modelling techniques and linear equations
- Understand the need of using operations research
- Simplification of LPP by graphical method
- Identify situations in which linear programming techniques can be applied
- Understand fundamental concepts and general mathematical structure of a linear programming model
- To understand and solve various transportation problems.

Course Outcomes:

On successful completion of the module students will be able to:

- Solve Linear Equations
- Work with Linear Programming
- Work with Network Models and Solving Problems
- Work with gaming and queuing theory
- Solve and examine situations that generate queuing problems

Module -1   **9 Hrs.**

Introduction To Operations Research (OR): Operations Research Definition and Scope, History of Operation Research, Features of Operations Research, OR approach to problem solving, Modelling in Operation Research, Principles of Modelling, Methodology of Operations Research, Management applications of Operations Research, Characteristics of Operations Research, Role of Operations Research in decision making.

Module -2   **12 Hrs.**

Linear Programming Problem(LPP): Introduction, Structure of LPP, Advantages and Limitations of LPP, Applications of LPP, Mathematical Model of LPP, Guidelines of Model Formulation, Examples of LP Model Foundation – Production – Marketing - Finance- Agricultural - Transportation - Personnel, Graphical Solution Methods of LPP, Simplex Algorithm – Minimization and Maximization Case for Linear Programming

Module – 3   **12 Hrs.**

Transportation Problems: Introduction of Transportation Problems, Transportation Algorithm, Test for Optimality, Maximization Transportation Problem, Trans-shipment Problem
Assignment Problems: Introduction to Assignment problem, Solution Methods, Hungarian Method, Travelling Salesman Problem

Module-4   **12 Hrs.**

Game Theory: Introduction, Two-Person Zero-Sum Game, Maximin-Minimax Principle, Games without saddle points–Mixed Strategies, Graphical Solution of 2×n and m×2 games,

Rathinam College of Arts & Science (Autonomous), Coimbatore-21.
Admitted in M.SC DSBA from the academic year 2021 Onwards

Page 62 of 90
Regulations 2021

Dominance Property. Replacement Theory: Introduction, Replacement of items that deteriorate, Replacement of items that fails.

Module-5   **15 Hrs.**

Queuing Theory: Queuing problem, characteristics, general structure of queuing system, probabilistic queuing models (Poisson-exponential single server model with infinite population), applications of queuing theory.
Markov Chains: Introduction, characteristics, Applications of Markov Chain, State and Transition Probabilities, Multi-Period Transition Properties, Steady-State Conditions

Text Books:

1. N.D. Vohra, "Quantitative Techniques in Management",6 thEd.,2004 , BPB.

2. Operations Research Theory and Applications – J K Sharma , Macmillan Publication, ISBN: 978-9350-59336-3

Reference Books:

1. V.K.Kapoor, "Operations Research Techniques for Management",1st Ed., 2001. Sultan Chand

2. K. Swarup, P.K.Gupta and M. Mohan, "Operations Research", 12th Ed.,2006, Sultan Chand

3. Hamady A. Taha, "Operations Research", 7 th Ed., 2005, Wesley.

Title	Social Media Analytics
Total Hours	60
Credits	03

Course Objectives:

- To understand the basic concepts and importance social media analytics.
- To understand the procedure for analysing twitter data and access the same through R platform.
- To understand the procedure for analysing Facebook data and access the same through R platform.
- To understand the procedure for analysing Instagram and access the same through R platform.
- To understand the procedure for analysing GitHub data and access the same through R platform

Course Outcomes:

On successful completion of the module students will be able to:

- Understand the important terminologies and analytics techniques in social media analytics.
- Analyse the twitter data and conclude the important finding and insights of the society thought on particular issues.
- Analyse the facebook data and conclude the important finding and insights of the society thought on particular issues.
- Analyse the Instagram profile and find out the interesting insights.
- Analyse the GitHub profile and find out the latest trending article in GitHub

Module -1   **12 Hrs.**

Introduction to Social Media Analytics : History and Evolution of social media, impact of social media in growth of business, Social media and its importance, Various social media platforms, Social media mining, Challenges for social media mining, Social media mining techniques: Graph mining and text mining, The generic process of social media mining: Getting authentication from the social website, Data visualization R packages, The simple word cloud, Sentiment analysis Word cloud, Pre-processing and cleaning in R.

Module -2   **12 Hrs.**

Analytics on Twitter: Introduction, Twitter and its importance, Understanding Twitter's APIs: Twitter vocabulary, Creating a Twitter API connection: Creating a new app, Finding trending topics, Searching tweets, Twitter sentiment analysis: Collecting tweets as a corpus, Cleaning the corpus, Estimating sentiment.

Module – 3   **12 Hrs.**

Analytics on Facebook : Introduction, importance of Facebook, Creating an app on the Facebook platform, Rfacebook package installation and authentication, Installation, A closer look at how the package works, A basic analysis of your network, Network analysis and visualization: Social network analysis, Degree, Betweenness, Closeness, Cluster, Communities, Getting Facebook page data, Trending topics analysis, Influencers: based on single post and multiple post, Measuring CTR performance for a page, Spam detection, Recommendations to friends.

Module-4   **12 Hrs.**

Analytics on Instagram: Definition and overview Instagram and its role in social awareness, Creating an app on the Instagram platform, Installation and authentication of the instaR package, Accessing data from R: Searching public media for a specific hashtag, Searching public media from a specific location, Extracting public media of a user, Extracting user profile, Getting followers, Getting comments, Number of times hashtag is used, Building a dataset: User profile, User media, Travel-related media, Popular personalities: Who has the most followers? Who follows more people? Who shared most media? Overall top users, Most viral media, Finding the most popular destination, Locations with most likes, Locations most talked about, Clustering the pictures, Recommendations to the users.

Module-5	**12 Hrs.**
Analytics on GitHub : Introduction to GitHub, creating an app on GitHub, GitHub package installation and authentication, Accessing GitHub data from R, Building a heterogeneous dataset using the most active users, Building additional metrics, Exploratory data analysis, EDA – graphical analysis: Which language is most popular among the active GitHub users? What is the distribution of watchers, forks, and issues in GitHub? How many repositories had issues? What is the trend on updating repositories? Compare users through heat map, EDA – correlation analysis: How Watchers is related to Forks, Correlation with regression line, Correlation with local regression curve, Correlation on segmented data, Correlation between the languages that user's use to code, how to get the trend of correlation?	

Text Books:

1. Mastering Social Media Mining with R– Sharan Kumar Ravindran, Vikram Garg, PACKT Publishing.

Reference Books:

2. Social Media Mining with R - Nathan Danneman, Richard Heimann, PACKT Publishing.
3. SOCIAL MEDIA MINING An Introduction - REZA ZAFARANI, MOHAMMAD ALI ABBASI, HUAN LIU, CAMBRIDGE University Press.

Title	Cloud Computing
Total Hours	**45**
Credits	**03**

Course Objectives:

- To provide students with the fundamentals and essentials of Cloud Computing.

- To provide students a sound foundation of the Cloud computing so that they are able to start using and adopting Cloud Computing services and tools in their real life scenarios.

- To enable students exploring some important cloud computing driven commercial systems such as GoogleApps, Microsoft Azure and Amazon Web Services and other businesses cloud applications.

Course Outcomes:

On successful completion of the module students will be able to:

- Analyze the Cloud computing setup with its vulnerabilities and applications using different architectures.

- Design different workflows according to requirements and apply map reduce programming model.

- Apply and design suitable Virtualization concept, Cloud Resource Management and design scheduling algorithms

- Create combinatorial auctions for cloud resources and design scheduling algorithms for computing clouds

- Assess cloud Storage systems and Cloud security, the risks involved, its impact and develop cloud application

- Broadly educate to know the impact of engineering on legal and societal issues involved in addressing the security issues of cloud computing

Rathinam College of Arts & Science (Autonomous), Coimbatore-21.
Admitted in M.SC DSBA from the academic year 2021 Onwards

Page 67 of 90
Regulations 2021

Module -1   9 Hrs.

Fundamentals of Cloud Computing: Cloud Computing Basics – History of Cloud Computing, Characteristics of Cloud Computing, Need for Cloud computing, Advantages and Possible Disadvantages of cloud computing, Cloud Deployment Models – Public, Private, Hybrid, Community, Other deployment Models. Evolving Data Center into Private Cloud, Datacentre Components, Extracting Business value in Cloud Computing – Cloud Security, Cloud Scalability, Time to Market, Distribution over the Internet, Cloud Computing Case Studies.

Module -2   9 Hrs.

Cloud Delivery Models : Introduction to Cloud Services, Infrastructure as a Service (IaaS) – Overview, Virtualization, Container, Pricing Models, Service Level Agreements, Migrating to the Cloud, IaaS Networking options, Virtual Private Cloud(VPC), IaaS Storage – File and Object storage, Data Protection, IaaS security, Benefits, Risks and Examples of IaaS. Platform as a Service (PaaS) – Overview, IaaS vs PaaS, PaaS Examples, benefits and risks. Software as a Service (SaaS) – Introducing SaaS, SaaS Examples – Office 365, Google G Suite, Salesforce.com , Evaluating SaaS – user and vendor perspective, Impact of SaaS, Benefits and risks of SaaS. Other Services on Cloud, Cloud Delivery Models Considerations

Module – 3   9 Hrs.

Cloud Platforms:

Introducing Cloud Platforms, Evaluating cloud platforms, Cloud Platform technologies – Amazon Web Services, Microsoft Azure, Google Cloud Platform, Salesforce.com, Impact of Cloud platforms. Private Cloud Platforms – Introducing Private clouds – Microsoft Azure stack, Open stack, AWS Greengrass, Impact of Private clouds

Cloud Migration : Delivering Business Processes from the Cloud: Business process examples, Broad Approaches to Migrating into the Cloud, The Seven-Step Model of Migration into a Cloud, Efficient Steps for migrating to cloud., Risks: Measuring and assessment of risks, Company concerns Risk Mitigation methodology for Cloud computing, Case Studies

Module-4   9 Hrs.

Cloud Computing - Challenges, Risk and Mitigation:
Cloud Storage, Application performance, Data Integration, Security. Ensuring Successful Cloud Adoption: Designing a Cloud Proof of Concept, Vendor roles and capabilities, moving to the Cloud. Impact of Cloud on IT Service Management.

Risks and Consequences of Cloud Computing: Legal Issues, Compliance Issues, Privacy

and Security.

Module-5   **9 Hrs.**

Managing the Cloud: Managing and Securing Cloud Services, Virtualization and the Cloud, Managing Desktops and devices on the cloud, SOA and Cloud computing, Managing the Cloud environment, Planning for the Cloud – Economic Cost Model and Leveraging the Cloud, Cloud computing resources, Cloud Dos and Don'ts.

Text Books:

1. Kirk Hausman, Susan L. Cook, TelmoSampaio, " CLOUD ESSENTIALS CompTIA® Authorized Courseware for Exam CLO-001", John Wiley & Sons Inc., 2013
2. Judith Hurwitz , Robin Bloor , Marcia Kaufman , Fern Halper, "Cloud Computing for Dummies", Wiley Publishing Inc., 2010

Reference Books:

1. Erl," Cloud Computing: Concepts, Technology & Architecture", Pearson Education, 2014
2. Srinivasan, "Cloud Computing: A Practical Approach for Learning and Implementation "Pearson Education, 2014

Title	Cloud Infrastructure Services
Total Hours	45
Credits	03

Course Objectives:

- To provide students with the fundamentals and essentials of Cloud Computing.

- To provide students a sound foundation of the Cloud computing so that they are able to start using and adopting Cloud Computing services and tools in their real life scenarios.

- To enable students exploring different services offered on Amazon Platform like Storage, Management tools, Analytical Services and Business Intelligence.

Course Outcomes:

On successful completion of the module students will be able to:

- Understand the Cloud Service Models and Cloud Deployment Models
- Create and configure the compute, storage and database services in the cloud which help them to work with analytic services.
- Monitor and get the logs related to various services.
- Deploy and configure the data analytics projects.
- Select Cloud services to analyse big data and create statistical models.

Module -1   **9 Hrs.**

Introduction to Cloud Computing : Introduction to Cloud Computing, Cloud Service Models, Cloud Deployment Models, Cloud Computing Security, Introduction to Amazon Web Services, AWS Compute Options, AWS Virtual Private Cloud, AWS Identity and Access Management, AWS Lambda, Pricing Concepts

Module -2   **9 Hrs.**

AWS Storage : Amazon Storage, S3 Storage Basics, Buckets and Objects, Creating A Web Server Using S3 Endpoints, Managing Voluminous Information with EBS, Glacier Storage Service , Describe Amazon DynamoDB, Understand key aspects of Amazon RDS,RDS Database engines, benefits of RDS, Launch an Amazon RDS instance, Amazon ElastiCache.

Module – 3   **9 Hrs.**

AWS Management Tools: Amazon CloudWatch, Accessing a CloudWatch using different method, CloudWatch Metrics, CloudWatch Alarms, Monitoring Amazon RDS, Collect Metrics and Logs with CloudWatch Agent, Benefits and features of AWS CloudTrail, CloudTrail Concepts, Creating and updating Trail, Working with log files. AWS Command Line Interface (CLI), AWS Personal Health Dashboard, Simple Notification Service(SNS).

Module-4   **9 Hrs.**

AWS Analytic Services: Introduction to AWS Analytic Services, Features and benefits of Amazon Athena, Creating Database, creating table and running a query in Athena, Querying AWS CloudTrail Logs, Benefits and features of Amazon EMR, EMR Architecture, Launch cluster and run a hive Script to process data, Plan and configure clusters, features of Amazon Elasticsearch(ES) Service, Creating and configuring Amazon ES Domains, Overview of

Amazon kinesis, Configuring input and output stream for Kinesis Data Analytics, features of AWS Glue

Module-5   **9 Hrs.**

Cloud Business Intelligence: Features of Amazon RedShift, Launch an Amazon Redshift Cluster, Connect to the Sample Cluster and Run Queries, Load sample data from S3, AWS Data Pipeline, Pipeline Components, Instances, and Attempts, DataNode, Database and Activities, Overview of Quicksight, Editions of Amazon Quicksight, Data Sources, Data Sets, Functions and Operators, Working with Analysis, Working with Visuals, Working with Stories, Working with Dashboards, Overview of Amazon SageMaker, Features of SageMaker, Create and configure notebook instances, create and configure training jobs, creating and configuring models and end points.

Text Books:

1. AWS Certified Solutions Architect Official Study Guide.by John Stamper, Sean Senior, Kevin E. Kelly, Biff Gaut, Tim Bixler, Hisham Baz, Joe Baron
2. Cloud Analytics with Google Cloud Platform: An end-to-end guide to processing and analyzing big data using Google Cloud Platform by SanketThodge
3. Rajaraman, Anand and Ullman, Jeff. (2008). Mining of Massive Datasets. New York: Cambridge Press.

Reference Books:

1. Isson, Jean-Paul and Harriott, Jesse. (2012). Win with Advanced Business Analytics: Creating Business Value from Your Data, 1st edition. New York: Wiley.

2. Shmueli, Galit, Patel, Nitin and Bruce, Peter. (2010). Data Mining for Business Intelligence. New York: Wiley.

3. Rajaraman, Anand. (2011). Mining of Massive Datasets. New York: Cambridge University Press.

Title	**Advanced Machine Learning**
Total Hours	**45**
Credits	**02**

Course Objectives:

- Implement Machine learning techniques using tensorflow
- Assess ensemble models involved in machine learning concepts
- Understand reinforcement learning concepts of machine learning
- Test the built models using validation techniques
- Deploy the machine learning models on cloud or local server

Course Outcomes:

On successful completion of the module students will be able to:

- Examine with advanced machine learning concepts
- Demonstrate different ensemble methods of machine learning
- Implement reinforcement learning in real world scenarios.
- Demonstrate and deploy machine learning concepts
- Deploy machine learning algorithms on cloud

Module -1   **9 Hrs.**

Advanced Machine learning with TensorFlow: Introduction, Tensorflow operations, declaring tensors, working with metrics, declaring operations, implementing activation functions, operations in computational graph, layering nested operations, working with multiple layers, implementing loss functions, implementing back propagation, working with batch and stochastic training, evaluating models, Implementing unit tests, multiple executors, productionalizingtensorflow

Module -2   **9 Hrs.**

Ensemble Methods: Bagging and Random forest, Bootstrap method, Bootstrap aggregation, Variable Importance, Boosting, AdaBoost, Boosting ensemble method, AdaBoost ensemble, CatBoost, Learning with ensembles, Implementing a simple majority vote classifier, Leveraging weak learners via adaptive boosting.

Module – 3   **9 Hrs.**

Reinforcement Learning: Introduction, formal framework, different components to learn a policy, value based methods for RL, Q-learning, fitted Q-learning, Deep Q-networks, double DQN, dueling network architecture, distributional DQN, Multi step learning, concepts of generalization, feature selection, modifying objective function, hierarchical learning, bias-over fitting tradeoff.

Module-4   **9 Hrs.**

Model Evaluation and Hyper-parameter Tuning: Streamlining workflows with pipelines, K-fold cross validation, Model performance measures, debugging algorithms with learning and validation curves, fine-tuning machine learning models via grid search, looking at different performance evaluation metrics, Ranking metrics, Classification metrics, regression metrics, Bootstrapping and Jackknife, Hold-out validation, difference between model validation and testing.

Module-5   **9 Hrs.**

Machine Learning Deployment: Serializing fitted scikit – learn estimators, setting up a SQLite database for data storage, developing web application with Flask, turning the classifier into a web application, turning a regression problem into a web application, pickle model, deploying web application to a public server, Cloud deployment using AWS and Google.

Text Books:

1. Master Machine Learning Algorithms, Jason Brownlee

2. Deeper Insights into Machine Learning, Birmingham, Packt

3. An Introduction to Deep Reinforcement Learning, Vincent François-Lavet, Peter Henderson, Riashat Islam, Marc   G. Bellemare and Joelle Pineau

4. Tensorflow machine learning cookbook, Nick McClure, Packt

Reference Books:

1. Advanced machine learning with python, John hearty, Packt

2. https://cloud.google.com/ml-engine/docs/deploying-models

3. https://towardsdatascience.com/simple-way-to-deploy-machine-learning-models-to-cloud-fd58b771fdcf

Rathinam College of Arts & Science (Autonomous), Coimbatore-21.
Admitted in M.SC DSBA from the academic year 2021 Onwards

Page 75 of 90
Regulations 2021

Title	Data Analytics using SQL
Total Hours	45
Credits	02

Course Objectives:

- Understand modelling techniques and linear equations
- Understand the need of using operations research
- Simplification of LPP by graphical method
- Identify situations in which linear programming techniques can be applied
- Understand fundamental concepts and general mathematical structure of a linear programming model
- To understand and solve various transportation problems.

Course Outcomes:

On successful completion of the module students will be able to:

- Explain SQL benefits and its syntax
- Write SQL statements to create, update, delete database objects
- Write SQL queries to retrieve data from tables
- Demonstrate built-in functions to retrieve data and format
- Write complex sub queries to retrieve data from multiple tables

Module -1   **9 Hrs.**

Introduction to SQL: Introduction to Structure Query Language (SQL), SQL History & Evolution, Features of SQL, Understanding of SQL process, Benefits and Role of SQL along with different market forces, Types of SQL, SQL Standards, SQL and Networking, Centralized architecture, File Server Architecture, Client Server Architecture, Multitier Architecture, Understanding concept for OLAP and OLTP Applications, Difference between OLAP and OLTP, SQL and Database Management, Data warehouse Concept

Module -2   **12 Hrs.**

SQL Statements & Executions: Types of SQL Statement, Data Definition language, Data Control language, Data Manipulation Language, Types of execution, Direct Invocation, Embedded SQL, Module Binding, Call-level interface, Data types, Constants, Numeric Constants, String Constants, Time & date Constants, Symbolic Constants, Expressions, Built in function, Null Values, Primary and Foreign Key Concept

Module – 3   **12 Hrs.**

Starting with basic SQL Syntax: Types of Tables, Create Database statement, Drop database Statement, Use statement, Create table Statement, Drop table Statement, Create index Statement, Drop index Statement, Describe Statement, Truncate Statement, Alter table Statement, Insert INTO Statement, Update table Statement, Delete table Statement, Commit Statement

Create SQL Tables, Specify Column data types, Create user Defined Types, Specify Column Default Values, Alter SQL Tables, Updating Data, Using WHERE Clause, Using Logical operations, AND operations, OR operations, Deleting SQL table

Module-4   **12 Hrs.**

Extracting Information & Manipulating Data: Select Statement, Returning only Distinct Rows, Using Aliases, Filtering Results using WHERE Clause, Logical Operations and Operator Precedence, NOT operator, BETWEEN Operator, LIKE Operator, IN Operator, Ordering Results with ORDER BY

Understanding SQL Arithmetic, basic Math operations, ABS() function, POWER() function, SQRT() function, RAND() function, CEILING() function, FLOOR() function, ROUND() function, SUBSTRING() function, Case Conversion Functions, REVERSE() function, TRIM() function, LENGTH() function, SOUNDEX() function, DIFFERENCE() function, DATE() function

Module-5   **15 Hrs.**

Grouping & Multi-table Queries: Grouping Results, Summarizing and Aggregating Data, Counting results, Adding Results, Averaging Results, MAX & MIN functions, using HAVING clause with GROUP BY Statements, Implicit Versus Explicit Groups, Counting DISTICT Values

Simple Joins/ Equi-Joins, Parent / child queries, Inner Joins, Multiple Joins, Cross Joins, Self Joins, Outer Joins, Right Joins, Left Joins, Full-outer Joins, Creating joins with more than two tables, Equi-Joins Versus Non-Equi Joins, Union operations

Text Books:

1. Beginning SQL, Paul Wilton and John W. Colby, Published by: Wiley Publishing, Inc

Reference Books:

1. SQL: The Complete Reference, James R. Groff and Paul N. Weinberg, McGraw-Hill/Osborne
2. Learning SQL, ALAN Beaulieu, O'REILLY

Title	Financial Econometrics Lab
Credits	04

List of Experiments:

Exercise -1

1. For the given data find out the time series component present in it.
2. Install Packages and calling installed packages related with time series in R
3. Understanding the function of ts packages in R

    4. Plotting of time series data and conclude the possible analysis for the same.

Exercise -2

    1. Create the moving average model for the given data – Simple Average.

    2. Create the moving average model for the given data – Moving Average.

    3. Create the moving average model for the given data – Weighted Moving Average.

    4. Fit naïves forecasting model for the given data.

    5. Fit Smoothing forecasting model for the given data – Exponential Smoothing (Holts Method).

Exercise -3

    1. Model Evaluation Techniques using – Error or Bias, MAD, MAPE, MSE.

Exercise -4

    1. Testing the stationarity of the given data

    2. Testing the autocorrelation

Exercise -5

    1. ACF and PACF

    2. Correlogram

Exercise -6

    1. Auto Regressive Model

    2. Moving Average Model

Exercise -7

    1. Fit ARMA for the given data and forecast the same for the next time period

    2. Fit ARIMA for the given data and forecast the same for the next time period

Exercise – 8

1. Testing of Spurious (Non Sense) Regression

2. Unit Root Test

3. Heteroscedasticity

4. Granger Causality

Exercise – 9

1. VAR

Exercise – 10

1. ARCH model fit

2. GARCH model fit

SEMESTER -4
Syllabus

Title	Artificial Neural Networks and Deep Learning
Total Hours	45
Credits	02

Course Objectives:

- To understand the importance of neural network system and its components and to understand the neural network learning and adaptation in data science.
- Know the importance of single layer perceptron in neural network model also to understand the advantage of multilayer perceptron over single layer perceptron.
- Able to apply the concepts of associative memory in retrieving information from data.
- To apply the concepts of self-organizing feature maps and its application and also to understand broad application of neural networks in different field of businesses.

Course Outcomes:

On successful completion of the module students will be able to:

- Identify the basic concepts of neural networks and its components
- Explain the detailed concepts of single layer perceptron neural networks
- Analyse neural network learning and adaptation techniques
- Examine the detailed concepts of multilayer perceptron neural networks
- Explain the different associative memory concepts in retrieving information in data

Module -1   **9 Hrs.**

Neural Network System: Introduction to biological neurons and their artificial models, history of artificial neural systems development, Simple Memory and Restoration of Patterns, basic concepts related to neural networks, Models of neural networks: feedforward and feedback networks, neural processing. Neural Network Learning and Adaptation: Learning as approximation or Equilibria Encoding, Outstar learning rule, summary and comparison of artificial neural network learning rules.

Module -2   **9 Hrs.**

Single Layer and Multilayer Perceptron: Single Layer Perceptron: Introduction to single layer perceptron, classification model, features and decision tree, multi category single layer perceptron neural networks.

Multilayer Perceptron: Introduction to multilayer perceptron neural networks, linearly non separable pattern classification, delta learning rule for multilayer perceptron networks, generalized delta learning rule, expert systems applications, learning time sequences.

Module – 3   **9 Hrs.**

Associative Memories: Introduction to associative memories, basic concepts, linear associators, Basic Concepts of Recurrent Auto associative Memory: retrieval algorithm and storage algorithm, performance considerations, capacity of autoassociative recurrent memory, memory conversion vs. corruption, fixed point concepts, Bidirectional Associative Memory, advantages and limitations.

Module-4   **9 Hrs.**

Hopfield Networks: Introduction, Binary Hopfield Networks, Setting of Weights in Hopfield Nets | Bidirectional Associative Memory (BAM) Principle, Walsh Functions, Network Stability, Summary of the Procedure for Implementing the Hopfield Network, Continuous Hopfield Models, The Continuous Energy (Lyapunov) Function, Hopfield Network Case Study: Character Recognition, Hopfield Network Case Study: Traveling Salesman Problem.

Module-5   **9 Hrs.**

Applications of Neural Algorithms and Systems: Introduction to neural network algorithms and its appropriate uses, linear programming modelling network, Multilayer Feedforward Network for Printed Character Classification, Handwritten Digit Recognition, Neural Networks Control Applications, Networks for Robot Kinematics, Connectionist Expert Systems for Medical Diagnosis.

Text Books:

1. Introduction to Artificial Neural Systems – Jacek M Zurada, West Publishing Company.

Reference Books:

1. Artificial Intelligence 3e: A Modern Approach – Stuart J. Russell and Peter Norv

Title	Natural Language Processing
Total Hours	60
Credits	03

Course Objectives:

- To understand the pre-processing of text for text analytics.
- To understand the importance of considering syntactic parsing.
- To understand the mechanism of text analytics generation in processing of natural language.
- To understand the importance of corpus creation in text analytics
- To understand the different statistical techniques used in text analytics

Course Outcomes:

On successful completion of the module students will be able to:

- Know the basic concepts of text analytics and its important terminologies
- Know the key role of syntactic parsing and semantic analysis in text analytics
- Know the importance of corpus creation in text analytics
- Know the important statistical techniques used in text analytics

Module -1   **12 Hrs.**

Introduction to Text Analytics : Introduction to text pre-processing, terminologies related with text processing, challenges of text pre-processing, tokenization, sentence segmentation, introduction to lexical analysis, finite state morphonology, finite state morphology, morphology vs lexcal analysis, paradigm based lexical analysis.

Module -2   **12 Hrs.**

Syntactic Parsing and Semantic Analysis: Introduction to syntactic parsing, The Cocke–Kasami–Younger Algorithm, parsing as deduction, Implementing Deductive Parsing, LR Parsing, Constraint-based Grammars, Issues in Parsing, Basic Concepts and Issues in Natural Language Semantics, Theories and Approaches to Semantic Representation, Relational Issues in Lexical Semantics, Fine-Grained Lexical-Semantic Analysis.

Module – 3	**12 Hrs.**

Natural Language Generation   : Introduction to natural language generation, simple Examples of Generated Texts, The Components of a Generator: Components and level of representation, Approaches to Text Planning: The Function of the Speaker, Desiderata for Text Planning, Pushing vs. Pulling, Planning by Progressive Refinement of the Speaker's Message, Planning Using Rhetorical Operators, Text Schemas, The Linguistic Component: Surface Realization Components, Relationship to Linguistic Theory, Chunk Size, Assembling vs. Navigating, Systemic Grammars, Functional Unification Grammars

Module-4	**12 Hrs.**

Corpus Creation: Introduction and definition of corpus in natural language processing, corpus size, Balance, Representativeness, and Sampling, Data Capture and Copyright, Corpus Markup and Annotation, Multilingual Corpora, Multimodal Corpora, Corpus Annotation Types, Morphosyntactic Annotation, Treebanks: Syntactic, Semantic, and Discourse Annotation, The Process of Building Treebanks, application of Treebanks.

Module-5	**12 Hrs.**

Statistical Techniques in Text Analytics: Introduction to statistics and its importance in text analytics, general linear model, binary linear classification, one versus all method for multi-category classification, maximum likelihood estimation in parameter estimation in linear classification techniques, concepts of generative and discriminative models, introduction to sequence prediction model and its application in text analytics.

Text Books:
Hand Book of Natural Language Processing, Second Edition – NITIN INDURKHYA FRED J. DAMERAU, CRC Press.

Reference Books:
1. Mining Text Data - Charu C. Aggarwal, ChengXiangZhai, Springer
2. Text Mining Classification, Clustering, and Applications - Ashok N. Srivastava, Mehran Sahami, CRC Press

Title	Cloud Web Services
Total Hours	60
Credits	03

Course Objectives:

- To get introduced to Cloud computing and various cloud service providers.

- To provide students a sound foundation of Cloud Computing so that they are able to start using and adopting Cloud Computing services and tools in their real life scenarios.

- To know about AWS and how it drastically changed the cloud market.

- To explore IaaS, PaaS and SaaS offerings of AWS.

- To learn how important security principles are implemented in virtualization and Cloud platforms in managing issues and challenges in Cloud deployment using Amazon platform.

- Broadly educate on legal and societal issues involved in addressing the security issues of cloud computing.

- Analyse the trade-offs between deploying applications in the cloud and deploying applications over local/on-premise infrastructure.

Course Outcomes:

On successful completion of the module students will be able to:

- Understand the benefits of using AWS.

- Analyze IaaS, PaaS and SaaS offerings of AWS.

- Apply AWS Compute services.

- Design virtual private network as per requirements and different organizational scenarios.

- Assess cloud Storage systems and Cloud security, the risks involved, its impact and develop cloud application.

Module -1   12 Hrs.

Introduction to Cloud Computing and Amazon Web Services: Introduction to Cloud Computing, Cloud Service Delivery Models (IAAS, PAAS, SAAS), Cloud Deployment Models (Private, Public, Hybrid and Community), Cloud Threats and Importance of Cloud Security, Case Study

Introduction to Amazon Web Services, Other leading cloud service providers in the market, Why Amazon? Use Cases, AWS Storage Options, AWS Compute Options, AWS Database Options, AWS Workflow Automation and Orchestration Options, AWS Systems Management and Monitoring Options, AWS Virtual Private Cloud Introduction, Explore various Digital Transformational Services by AWS in Artificial Intelligence, Analytics and IoT with some related use cases, AWS Pricing Concept.

Module -2   12 Hrs.

AWS Compute: Introduction to Elastic Cloud Compute (EC2), Amazon Machine Images (AMIS), Instance Lifecycle, Instance Types And Uses, Modifying Existing Images, Creating New Images of Running Instances, Converting an Instance Store AMI to an EBS AMI, EBS Volume Types, Elastic IPs, Associating and releasing Elastic IP addresses to and from an EC2 Instance, Elastic Load Balancing, Auto scaling EC2 Instances, Introduction to AWS Lambda, Automating EC2 instances using code in AWS Lambda.

Module – 3   12 Hrs.

Web Applications and Security: Introduction to Elastic Beanstalk, Deploying Scalable Application On AWS, Selecting And Launching An Application Environment, Provisioning Application Resources with Cloud formation, Introduction to CloudWatch, Describe Amazon Cloud Watch metrics and alarms, AWS Messaging Services(SNS,SQS,SES).

Introduction to AWS Security, Describe Amazon Identity and Access Management (IAM), AWS Directory Service, AWS Key Management Service, Securing Data at Rest and In Motion.

Module-4   **12 Hrs.**

AWS Storage: Amazon Storage, S3 Storage Basics, Buckets and Objects, S3 Versioning, Hosting a static website on Amazon S3 using S3 website endpoints, Managing Voluminous Information with Elastic Block Storage (EBS) , EBS Features, Glacier Storage Service , Describe Amazon Dynamo DB and its benefits, Applications and use cases of Dynamo DB, Understand key aspects of Amazon RDS, Launch an Amazon RDS instance.

Module-5   **12 Hrs.**

AWS Networking: Introduction to AWS Networking , Access Control Lists (ACLs), Network ACL Basics and Rules, Default and Custom Network ACL, Managing Access with ACLs, Setting Up a Security Group, Setting up Virtual Private Cloud (VPC) and Internet Gateway, Setting up Virtual Private Network(VPN), Two services of VPN: AWS Site-to-Site VPN and AWS Client VPN, Understanding Customer Gateway,Virtual Private Gateway, Transit Gateway, Setting Up Dedicated Hardware For VPC, Scenario 1:VPC with a Single Public Subnet (Standalone Web), Scenario 2: VPC with Public and Private Subnets using NAT (multi-tier website), Scenario 3:VPC with Public and Private Subnets And AWS Site-to-Site VPN Access (Extension of your Corporate Network into the cloud and also directly access the Internet from your VPC) Scenario 4: VPC with a Private Subnet Only and AWS Site-to-Site VPN Access. (Extension of your Corporate Network into the cloud using Amazon's infrastructure without exposing your network to the Internet), Route53 for DNS System, Cloud Front.

Text Books:

1. Cloud Computing: Principles and Paradigms, Editors: RajkumarBuyya, James Broberg, Andrzej M. Goscinski, Wiley, 2011.
2. https://awsdocs.s3.amazonaws.com/gettingstarted/latest/awsgsg-intro.pdf

Reference Books:

1. Cloud Computing Explained: Implementation Handbook for Enterprises, John Roton, Recursive Press (November 2, 2009).

Title	Advanced Big Data Analytics
Total Hours	45
Credits	02

Course Objectives:

- The main goal of this course is to provide an overview of an exciting growing field to use core Spark APIs to operate on data. The course studies fundamentals of data stream processing applications.

- The emphasis of the course is on the theoretical foundations and research issues surrounding the algorithmic, analytical, software engineering, and distributed processing architectural foundations of stream processing.

- It also provides a solid programming foundation and an understanding of the practical aspects of building stream processing

Course Outcomes:

On successful completion of the module students will be able to:

- Design and build stream processing systems, services and applications.
- Produce data pipelines and query large data sets using Spark SQL and DataFrames
- Co-Ordinate with relational data using the GraphFrames APIs.
- Integrate current research trends in data-stream processing.
- Design and build stream processing systems, services and applications.

Module -1   **9 Hrs.**

Streaming Data & Architectures: Sources of Streaming Data, Operational Monitoring, Web Analytics, Online Advertising, Social Media, Mobile Data and the Internet of Things, Why Streaming Data is Different, Loosely Structured, Highly-Cardinality Storage

Real-Time Architecture Components, Collection, Data Flow, Processing, Storage, Delivery, Features of a Real-Time Architecture, High Availability, Low Latency, Horizontal Scalability, Languages for Real-Time Programming, Understanding MapReduce Failure for Streaming Data.

Module -2   **9 Hrs.**

Introducing Real-Time Processing Tool: Apache Spark, Why Apache Spark, Evolution of Apache Spark, Architecture Apache Spark, Features of Apache Spark, Spark Deployment, Standalone, Hadoop YARN, Spark MapReduce, Components of Apache Spark, Spark core, Spark SQL, Spark Streaming, Spark Machine Learning, Spark GraphX, Spark Shell, Resilient Distributed Dataset (RDD) Basic, Spark Context, RDD Transformations, Creating RDDs, RDD Operations, Programming with RDD, Transformations, Actions, Lazy Evaluation,

Converting between RDD Types.

Module – 3   9 Hrs.

Data Processing using R & MLL: Spark Data Frame, Starting Session, Creating Data Frames, Spark Data Frame Operations, Rows, Columns, Grouping and Aggregations, Operating on Columns, User defined Functions, Data type mapping between R and Spark, Structured Streaming

Architecture, Data types, Classification and Regression, Development Environment, Classification with Naïve Bayes, Theory, Clustering with K-Means, Theory, Artificial Neural Networks, Theory, Use Cases.

Module-4   9 Hrs.

Processing Live Data Streams & SQL: Streaming overview, Basic Concepts, Linking, Streaming Context, Discretized Streams, Input DStreams and Receivers, Streaming Sources, TCP streams, File streams, Transformations on DStreams, Operations, DataFrame and SQL Operations, Check pointing, Memory Tuning Fault Tolerance

SQL Context Datasets, DataFrame, Data Sources, , Importing, Processing Text Files, Processing JSON Files, Processing Parquet Files, Using SQL, User Defined Functions, Using Hive, Performance tuning, Distributed SQL Engine, Compatibility with Apache hive

Module-5   9 Hrs.

Data Processing using GraphX: Overview, Getting Started, Property Graph, Graph operations, Summary List, Property Operator, Structural Operator, Join, Neighbourhood Aggregations, Pregel API, Graph Builders, Vertex and Edge RDDs, Optimized Representation, Graph Algorithm.

Text Books:
1. Real-Time Analytics, By Byron Ellis, Wiley publication
2. Mastering Apache Spark, By Mike Frampton, Packt Publishing

Reference Books:
1. Learning Spark, By Holden Karau, Andy Kowinski&MateiZaharia, Published by

Rathinam College of Arts & Science (Autonomous), Coimbatore-21.
Admitted in M.SC DSBA from the academic year 2021 Onwards

Page 90 of 90
Regulations 2021

O'Reilly Media, Inc