DEPARTMENT OF COMPUTER SCIENCE

RATHINAM COLLEGE OF ARTS AND SCIENCE (AUTONOMOUS)

Rathinam Techzone, Pollachi Road, Eachanari, Coimbatore – 641021

Celebrate life

RATHINAM

Syllabus for

M.Sc. Data Science & Business Analysis
(M.Sc. DS & BA)

(I & II Semester)

2018 – 2019 Batch onwards

RATHINAM COLLEGE OF ARTS AND SCIENCE (AUTONOMOUS)

Scheme of curriculum for M.Sc. Data Science & Business Analysis

for the Batch admitted during 2018 - 2019

Board of Studies – Computer Science (PG)

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

Vision and Mission of the Department:

Vision

To groom students to be a Centre of Excellence in the Discipline of Computing with various backgrounds into competitive software professionals and pioneering leaders in offering innovative solutions to lively worldwide challenges in tune with the needs of the society through effective Teaching and Learning.

Mission

- Provide an infrastructure for students to gain quality learning process in a research oriented expertise in theoretical and practical foundations of computer applications to develop real world applications catering to the global needs to make the students ready to meet the industry requirements.
- To develop human resource with sound knowledge-theory and practical in the discipline of computing and the ability to apply the knowledge to the benefit of the society at large through a continuous process of improvement.

Program Educational Objectives (PEO)

PEO1	:	Graduates of this programme will establish as effective professionals by learning technical skills for the computing field and can pursue higher education by accruing knowledge and research.
PEO2	:	Graduates of the programme will continue to learn and adopt latest technologies to solve real life problems.
PEO3	:	Graduates of the programme will continue to develop and update their knowledge and skills throughout their career.
PEO4	:	Graduates of the Data Science programme will have a sufficient understanding of the field of computer science including scientific principles, analysis techniques, and design methodologies.
PEO5	:	Students will get a solid basis in the informatics and statistical methodology necessary for working within Data Science.

Mapping of Institute Mission to PEO

Institute Mission	PEO's
To provide quality education at affordable cost, and to maintain academic and research excellence with a keen focus on INDUSTRY-INTEGRATED RESEARCH AND EDUCATION.	PEO1, PEO3

Mapping of Department Mission to PEO

Department Mission	PEO's
To provide quality education at affordable cost, and to maintain academic and research excellence with a keen focus on INDUSTRY-INTEGRATED RESEARCH AND EDUCATION.	PEO1, PEO3

Program Outcomes (PO):

PO1	**:**	To formulate the students and to attain ability to apply knowledge of theoretical foundations and formal methods in Data Science and Business Analytics.
PO2	**:**	An ability to analyze a problem, to design, implement, and evaluate a computer-based System, process, component, or program to meet desired needs.
PO3	**:**	An understanding of professionalism, ethics, security and social issues with responsibilities and how these are related to the local and the global impact of computing on Individuals, Organizations, and the Society.
PO4	**:**	Ability to design and conduct experiments, as well as to analyze and interpret data to develop computer programs and to solve them in an efficient manner.
PO5	**:**	To involve in project management techniques and teamwork, this is necessary for successful system designs and Implementations with the effective use of Data Analysis skills.
PO6	**:**	Ability to use and apply current technical concepts and practices in the core Data Analysis techniques of information related management, programming, business, web systems, and technologies.
PO7	**:**	Students will be able to demonstrate the use of modern tools, software, and to analyze.

Correlation between the POs and the PEOs

Program Outcomes		**PEO1**	**PEO2**	**PEO3**	**PEO4**	**PEO5**
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

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		√	√	√		√	√		√

RATHINAM COLLEGE OF ARTS AND SCIENCE (AUTONOMOUS)

Scheme of curriculum for M.Sc. Data Science & Business Analytics

for the Batch admitted during 2018 - 2019

Board of Studies – Computer Science (PG)

S.No	Sem	Part	Type	Subject	Credit	Hour	Int	Ext	Total
1	I	III	Theory	Principles of Data Science	4	6	40	60	100
2	I	III	Theory	Operating System	4	6	40	60	100
3	I	III	Theory	Python Programming	4	6	40	60	100
4	I	III	Theory	Data Mining	4	6	40	60	100
5	I	III	Practical	Python Programming Lab	4	6	40	60	100
1	II	III	Theory	Advanced Database Management Systems	4	5	40	60	100
2	II	III	Theory	R Programming for Data Analytics	4	5	40	60	100
3	II	III	Theory	Machine Learning	4	5	40	60	100
4	II	III	Theory	Information Storage and Management	4	5	40	60	100
5	II	III	Theory	TCP / IP Protocol Suite	4	5	40	60	100
6	II	III	Practical	R Programming for Data Analytics Lab	4	5	40	60	100
1	III	III	Theory	Evolutionary Computing	4	5	40	60	100
2	III	III	Theory	Elective-I	4	5	40	60	100
3	III	III	Theory	Elective II	4	5	40	60	100
4	III	III	Practical	MapReduce Programming	4	5	40	60	100
5	III	III	Theory	IoT and Wireless Sensor Networks	4	5	40	60	100
6	III	III	Theory	Mini Project and Viva Voce	2	5	20	30	50
1	III	III	Theory	Data Storage Management	4	5	40	60	100
2	III	III	Theory	Elective III	4	5	40	60	100
3	III	III	Theory	Elective-IV	4	5	40	60	100
4	III	III	Theory	Cloud Infrastructure and Services	4	5	40	60	100
5	IV	III	Project	Project and Viva Voce	8	10	100	100	200
				Total	90	120			2250

List of Electives:

Semester	Elective	Paper title
Semester - III	**Elective - I**	Sentiment Analysis
		Social Media Mining
		Text Analytics
Semester - III	**Elective - II**	Time Series and Forecasting
		Graph Analytics for Big Data
		Semantic Data Models /Web
Semester - IV	**Elective - III**	Hadoop Programming
		Cloud Computing
		Agile Testing
Semester - IV	**Elective - IV**	Deep Learning
		Data Visualization
		Building Internet of Things

Mapping of Courses and POs:

S- Strong Correlations M – Medium Correlation B – Blank

Course Code	Course Name	Program Outcomes						
		P01	PO2	P03	P04	P05	P06	P07
	Principles of Data Science	S	M	S	S		S	
	Operating System	M	S		S	M	S	
	Python Programming	S	M	S		S	M	S
	Data Mining	M	S	M	S		M	S
	Python Programming Lab	S		M	S	M		S
	Advanced Database Management Systems	S	M		S		S	
	R Programming for Data Analytics		S		M	S		S
	Machine Learning	M		S	S	M		M
	Information Storage and Management		S	M	S		M	
	TCP / IP Protocol Suite	S	M	M		S	M	
	R Programming for Data Analytics Lab	M	S		M	S		S
	Evolutionary Computing		S	M	S		M	S
	Elective-I	M	S		M		S	M
	Elective II	S		M		M	S	
	MapReduce Programming	M		S	M		S	M
	IoT and Wireless Sensor Networks		S	M	S		M	
	Mini Project and Viva Voce	S	M		M	S		M
	Data Storage Management	M	S		M	S	M	
	Elective III	S		M	S		M	S
	Elective-IV		M		S	M	S	
	Cloud Infrastructure and Services	M	S			M		M
	Project and Viva Voce	S	M	S	M		S	S

Semester : I

Subject Code	Subject Title	Credit	Lecture	Tutorial	Practical	Type
	Core - I - Principles of Data Science	4	5	1	0	Theory

Introduction :

Data Science is the study of the generalizable extraction of knowledge from data. Being a data scientist requires an integrated skill set spanning mathematics, statistics, machine learning, databases and other branches of computer science along with a good understanding of the craft of problem formulation to engineer effective solutions

Course Outcome:

CO1	:	Describe the Data Science Process and how its components interact.
CO2	:	To formulate about big data history and its innovation.
CO3	:	To understand the hadoop framework and its architecture.
CO4	:	To have the ability to understand how big data analysed into data science and explained about the way of its implementation.
CO5	:	To particularly know about R Programming tool and its working principle.

Unit I: [12 Periods]

Data Evolution: Data Development Time Line – ICT Advancement- A Perspective – Data Growth-A Perspective – IT Components-Business Process – Landscape-Data to Data Science – Data Science : Data Science-A Discipline- Data Analytics - Relation: Data Science, Analytics, Big Data Analytics - Data Science Components Data Engineering, Data Analytics-Methods and Algorithm, Data Visualization – Data Science Big technology – Data Science ontology – Data science user – Data science use cases.

Unit II: [12 Periods]

Road map to Big Data- What Is Big Data - Characteristics of Big Data - Data in the Warehouse and Data in Hadoop- Wrapping It Up - Why Is Big Data Important - Big Data Use Cases: Patterns for Big Data Deployment: IT for IT Log Analytics - The Fraud Detection Pattern - They Said What? The Social Media Pattern - The Call Center Mantra: “This Call May Be Recorded for Quality Assurance Purposes” - Risk: Patterns for Modeling and Management - Big Data and the Energy Sector - Why IBM for Big Data? - A History of Big Data Innovation.

Unit III: [12 Periods]

Hadoop: Basic Concepts-An Overview of Hadoop- Hadoop Architecture - The Hadoop Distributed File System – HDFS Architecture - Map Reduce Application – Hadoop Ecosystem – Limitations of Hadoop – Hadoop Yarn – Yarn infrastructure – Application startup in Yarn – Hadoop Ecosystem – Components of Hadoop Eco System – Hadoop Installation – PIG Installation – HIVE Installation.

Unit IV: [12 Periods]

What are Big Data Analytics & Data Science –Big Data Analytics & Data Science – Are they the same? - Introduction to Machine Learning - Big Data Technology Potentials – Limitations of Big Data and Challenges- Big Data Roles Data Scientist , Data Architect, Data Analyst – Skills – Case Study : Big Data – Customer Insights – Behavioral Analysis – Big Data Applications - Marketing – Retails – Insurance – Risk and Security – Health care.

Unit V: [12 Periods]

R Programming : History and Overview of R - Getting Started with R - R Nuts and Bolts - Getting Data In and Out of R - Using the readr Package - Interfaces to the Outside World . Pig : Why Pig ? – Pig user interactive modes – Pig Latin – Dataset – Pig Commands and functions – Relational Operators – Evaluation function – Batch Mode – Embedded Mode – PIG vs SQL .

Reference :

V. Bhuvaneswari, T. Devi, "Big Data Analytics: A Practitioner's Approach" 2016. Paul C. Zikopoulos ,Chris Eaton, Dirk deRoos, Thomas Deutsch ,George

2. Understanding Big Data Analytics for Enterprise Class Hadoop and Streaming Data 2012
3. Roger D. Peng, R Programming for Data Science 2013.

Mapping of Course Outcomes with Program Outcomes:

Course Outcomes	Program Outcomes						
	PO1	**PO2**	**PO3**	**PO4**	**PO5**	**PO6**	**PO7**
CO1		H		M			
CO2	H				M		
CO3	M					L	
CO4		H			M		
CO5				L			H

H-High; M-Medium; L-Low

Semester : I

Subject Code	Subject Title	Credit	Lecture	Tutorial	Practical	Type
	Core - I – Operating System	4	5	1	0	Theory

Introduction :

A successful student will be able to understand the basic components of a computer operating system, and the interactions among the various components. The course will cover an introduction on the policies for scheduling, deadlocks, memory management, synchronization, system calls, and file systems.

Course Outcome:

CO1	:	To know about operating system structure, management operations and its services.
CO2	:	To have detailed study about process management and multi-threading models.
CO3	:	To understand about CPU scheduling algorithms and the methods of handling deadlocks.
CO4	:	To have a detailed description about memory management, paging concepts.
CO5	:	The ability to understand file concepts, implementation, recovery and its allocation methods.

Unit I: [12 Periods]

Introduction - What operating system do?- Computer System Organization - Computer-System Architecture - Operating-System Structure - Operating-System Operations - Process Management - Memory Management - Storage Management - Protection and Security - Operating System Services - User and Operating System Interface - System Calls - Types - System Programs

Unit II: [12 Periods]

Process Management - Process Concept - Process Scheduling - Operations on Process - Interprocess Communications - Threads - Multithreading models - Thread Libraries - Threading Issues - Process Synchronization - Critical-section Problem - Semaphores - Monitors

Unit III: [12 Periods]

CPU Scheduling - Scheduling Criteria - Scheduling Algorithms - Dead lock - System Model - Deadlock Characterization - Methods for Handling Deadlocks - Deadlock Prevention - Deadlock Avoidance - Deadlock Detection - Recovery from Deadlock.

Unit IV: [12 Periods]

Memory Management - Main Memory - Swapping - Contiguous Memory Allocation - Paging - Structure of the Page Table - Virtual Memory - Demand Paging - Copy-on Write - Page Replacement - Allocation of Frames

Unit V: [12 Periods]

File-System Interface - File Concept - Access Methods - Directory and Structure - File System mounting - File Sharing - File System Implementation - Directory implementation - Allocation methods - Fee-space management - Recovery

Reference :

1. Abraham Silberschatz, Peter Baer Galvin and Greg Gagne, "Operating System Concepts", 9th Edition, John Wiley and Sons Inc., 2012.
2. William Stallings, "Operating Systems – Internals and Design Principles", 7th Edition, Prentice Hall, 2011.
3. D M Dhamdhere, "Operating Systems: A Concept-Based Approach", Second Edition, Tata McGraw-Hill Education, 2007.
4. http://nptel.ac.in/

Mapping of Course Outcomes with Program Outcomes:

Course Outcomes	Program Outcomes						
	PO1	PO2	PO3	PO4	PO5	PO6	PO7
CO1	L	H					
CO2		M	H				
CO3					H		H
CO4			M				M
CO5		L			H		

H-High; M-Medium; L-Low

Semester : I

Subject Code	Subject Title	Credit	Lecture	Tutorial	Practical	Type
	Core - I – Python programming	4	5	1	0	Theory

Introduction :

On successful completion of the course the student should have understood the concepts in Python and its application. Python programming is intended for software engineers, system analysts, program managers and user support personnel.

Course Outcome:

CO1	:	To gain knowledge in core python, syntax, basic programs and its objects.
CO2	:	To understand about numbers, strings and built-in functions.
CO3	:	To know about and work on python lists, operators, dictionaries and set types.
CO4	:	To understand about file concepts in python, exceptions and an introduction to object oriented programming.
CO5	:	To have a detailed study about internet and network programming and regular expressions.

Unit I: [12 Periods]

Core Python: What is Python? - Origins - Features – Comparing Python- Getting Started: Comments – Operators -Variables and Assignment – Numbers – Strings - Lists and Tuples - Dictionaries - Functions – Python Basics: Statements and Syntax - Variable Assignment – Identifiers – Basic Style Guidelines - Memory Management-First Python Programs - Python Objects: Python Objects - Standard Types – Other Built-in-types - Internal types - Standard Type Operators - Standard Type Built-in functions – Categorizing the Standard types - Unsupported types.

Unit II: [12 Periods]

Numbers: Introduction to Numbers – Integers – Double Precision Floating Point Numbers - Complex numbers – Operators - Built-in and Factory Functions - Other Numeric Types – Sequences: Strings, Lists and Tuples: Sequences – Strings - Strings and Operators – String-Only Operators - Built-in Functions – String Built-in-Methods – Special Features of Strings - Unicode - Related Modules.

Unit III: [12 Periods]

Lists - Operators - Built-in Functions – List Type Built-in Methods – Special Features of Lists – Tuples – Tuple Operators and Built-in Functions – Special Features of Tuples - Related Modules – Mapping and Set Types: Mapping Type: Dictionaries - Mapping Type Operators - Mapping Type Built-in and Factory Functions - Mapping Type Built-in-Methods - Set Types - Set Type Operators - Built-in Functions - Set Type Built-in Methods - Related Modules - Conditionals and loops.

Unit IV: [12 Periods]

Files and Input/output: File Objects – File Built-in Functions – File Built-in Methods - File Built-in Attributes - Standard Files – Command-Line Arguments - File System - File Execution - Persistent Storage Modules - Related Modules - Errors and Exceptions: What Are Exceptions? – Exceptions in Python – Detecting and Handling Exceptions – Object-Oriented Programming: Classes – Class Attributes – Instances.

Unit V: [12 Periods]

Regular Expression: Introduction / Motivation - Special Symbols and Characters - REs and Python – Regular Expressions Examples - Network Programming: What Is Client/Server Architecture – Sockets: Communication Endpoints - Internet Client Programming- Transferring files – Electronic Mail - GUI Programming: Introduction - Tkinter and Python Programming - Database Programming: Introduction - Python (DB – API)

Reference :

1. Chun, J Wesley, Core Python Programming, Prentice Hall, 2nd Edition, Pearson, 2007 Reprint 2010.
2. Wesley J Chun Core python Application Programming,3rd Edition
3. Lutz, Mark, Learning Python, 5th Edition, O Rielly.

Mapping of Course Outcomes with Program Outcomes:

Course Outcomes	Program Outcomes						
	PO1	**PO2**	**PO3**	**PO4**	**PO5**	**PO6**	**PO7**
CO1	M	H					
CO2			H	M			
CO3	H		M				
CO4			H			L	
CO5					M		H

H-High; M-Medium; L-Low

Semester : I

Subject Code	Subject Title	Credit	Lecture	Tutorial	Practical	Type
	Core - I – Data Mining	4	5	1	0	Theory

Introduction :

To introduce the basic concepts of Data Warehouse and Data Mining techniques. Examine the types of the data to be mined and apply pre processing methods on raw data. Discover interesting patterns, analyze supervised and unsupervised models and estimate the accuracy of the algorithms.

Course Outcome:

CO1	:	To learn about basic concepts of data mining, current trends and algorithm.
CO2	:	To have a detailed study about clustering techniques and its algorithm implementation using MAT lab tool.
CO3	:	To formulate about decision trees and its algorithm.
CO4	:	To learn about web mining, text mining and its processing features.
CO5	:	To have a detailed study about data warehousing, OLAP and its current technology along with cloud usage.

Unit I: [12 Periods]

Introduction: What Is Data Mining – KDD versus Data Mining – Data Mining Techniques – ues and challenges in Data Mining – Data Mining applications - Current Trends in Data ning. Association Rules – Concept- Methods to discover Association rules- A priori algorithm Partition algorithm- Discussion on Different Algorithm

Unit II: [12 Periods]

Clustering techniques: Introduction – Clustering Paradigms – Partitioning Algorithm – K-Means, ARA-CLARANS- Hierarchical Algorithm- DBSCAN-BIRCH-CURE-Categorical Clustering gorithm- STIRR-ROCK-CACTUS. Implementation of clustering techniques using Mat lab Tool.

Unit III: [12 Periods]

Classification Technique: Introduction - Decision Trees: Construction Principle – Attribute lection measure – Tree Pruning - Decision Tree construction Algorithm – CART – ID3 - inforest - CLOUDS - BOAT, Pruning Technique- Cross Validation – Bootstrap-– Classifier rformance.

Unit IV: 12 Periods]

Introduction: Web Mining – Content Mining – Structure Mining – Web Usage Mining – Text Mining – Unstructured Test- Text Pre-processing - Text clustering – Spatial mining – Spatial mining tasks – Spatial clustering – Spatial trends – Case Studies : Big Data , IoT.

Unit V: [12 Periods]

Data warehousing: Introduction - Definition - Multidimensional data model - OLAP operations - Warehouse schema - Data warehousing architecture- Meta data - OLAP Engine - Data warehouse backend process - Data Warehouse Technology - Warehousing Software - Cloud data warehousing - Other features. Data Warehousing Case Study: Government, Tourism and Industry

Reference :

1. Arun K Pujari, Data Mining Techniques, Universities Press. 2012
2. Jiawei Han, Micheline Kamber, "Data Mining Concepts and Techniques", Morgan Kaufmann Publishers, 2012
3. Data Mining, The textbook, Charu C.Agarwal, Springer.

Mapping of Course Outcomes with Program Outcomes:

Course Outcomes	Program Outcomes						
	PO1	**PO2**	**PO3**	**PO4**	**PO5**	**PO6**	**PO7**
CO1	M	H					
CO2					L		H
CO3				H		M	
CO4		M				L	
CO5				L	M		H

H-High; M-Medium; L-Low

Semester : I

Subject Code	Subject Title	Credit	Lecture	Tutorial	Practical	Type
	Core - I – Python Programming Lab	**4**	**5**	**1**	**0**	**Practical**

Course Outcome:

To apply the principles of python programming, to write clear and effective python code. To create applications using python programming and to implement them in database using SQLite. To develop web applications and web services using python programming.

List of Programs:

1. Write python program to print Hello World.

2. Write python program to Hello World using string variable

3. Write python program to store data in list and then try to print them.

4. Write python program to do basic trim and slice on string.

5. Write python program to print list of numbers using range and for loop

6. Write python program to store strings in list and then print them.

7. Write python program to let user enter some data in string and then verify data and print welcome to user.

8. Write python program in which an function is defined and calling that function prints Hello World

9. Write python program in which an function(with single string parameter) is defined and calling that function prints the string parameters given to function.

10. Write python program in which an class is define, then create object of that class and call simple print function define in class.

Semester : II

Subject Code	Subject Title	Credit	Lecture	Tutorial	Practical	Type
	Core - I – Advanced Database Management Systems	**4**	**5**	**1**	**0**	**Theory**

Introduction :

Introduce basic concepts and major techniques in DBMS implementations. These include concepts and techniques for data storage, query processing, concurrency control and transaction management. Introduce research development ability in databases through technical survey and presentation.

Course Outcome:

CO1	:	Analyze an information storage problem and derive an information model expressed in the form of an entity relation diagram and other optional analysis forms, such as a data dictionary.
CO2	:	Use an SQL interface of a multi-user relational DBMS package to create, secure, populate, maintain, and query a database. Use a desktop database package to create, populate, maintain, and query a database.
CO3	:	Relational model concepts. Referential integrity, entity integrity, and other constraints. Defining a relational schema from an ER diagram.
CO4	:	Describe data models and schemas in DBMS, To understand the features of database management systems and Relational database. To use SQL-the standard language of relational databases. To understand the functional dependencies and design of the database. To understand the concept of Transaction and Query processing
CO5	:	Master the basic Concepts and appreciate the applications of database systems.

Unit I: [12 Periods]

Introduction to Database Systems:- Overview – Data Abstraction – Data Models – Database System Architecture – Instances and Schemes – Data independence – DDL – DML – Database Users – Entity-Relationship Model:- Entity Sets – Keys – ER diagram – Structure – Extended E-R features – Design of an ER Database Schema – Reduction of E-R Schema to Tables.

Unit II: [12 Periods]

Relational Model:- Structure of Relational Databases – Relational Algebra – Extended Relational Algebra Operations – Modification of Database – Views – Tuple Relational Calculus – Domain Relational Calculus. **SQL :-** Background – Basic Structure – Set Operations - Aggregate Functions – Null Values – Nested Sub queries – Views – Complex Queries – Modification of the database –

Joined Relations – Data-definition language.

Unit III: [12 Periods]

Integrity & Security: - Domain Constraints – Referential Integrity – Assertions – Triggers – Security & Authorization – Authorization in SQL – Encryption and Authentication. **Relational Database Design: -** First Normal Form – Second Normal form – Boyce-Codd Normal Form – Third Normal Form – Fourth Normal Form.

Unit IV: [12 Periods]

Storage & File Structures :- Overview of Physical Storage Media – Magnetic Disks – RAID – Tertiary Storage – Storage Access – File Organization – Organization of Records in Files – Data Dictionary Storage. **Indexing and Hashing: -** Basic Concepts – Ordered Indices – B^+-Tree Index Files – B-Tree Index Files – Static Hashing – Dynamic Hashing Index – Definition in SQL – Multiple-Key Access.

Unit V: [12 Periods]

Transactions :- Transaction Concept – Transaction State – Implementation of Atomicity and Durability – Concurrent Executions – Serializability – Recoverability – Implementation of Isolation – **Testing for Serializability Concurrency Control :-** Lock – Based Protocols – Timestamp Based Protocols – Validation – Based Protocols – Multiple Granularity – Deadlock Handling.

Reference :

1. Silberschatz, Korth, Sudarshan, "Database System Concepts", 4th edition – McGraw Hill Higher Education, International Edition 2002.
2. Abraham Silberschatz, Henry F. Korth, S. Sudarshan "DATABASE SYSTEM CONCEPTS", 6th edition – McGraw Hill
3. Fred R McFadden, Jeffery A Hoffer, Mary B.Prescott, "Modern Database Management", Fifth edition, Addison Wesley, 2000
4. Silberschatz, Korth, Sudarshan Database System Concepts, Fourth Edition The McGraw−Hill Companies, 2001
5. Hans-Petter Halvorsen Structured Query Language 2016.01.08

Mapping of Course Outcomes with Program Outcomes:

Course Outcomes	Program Outcomes						
	PO1	**PO2**	**PO3**	**PO4**	**PO5**	**PO6**	**PO7**
CO1	M	H		L			
CO2				M	L		H
CO3		H		H	M		
CO4	L				M		M
CO5			M				H

H-High; M-Medium; L-Low

Semester: II

Subject Code	Subject Title	Credit	Lecture	Tutorial	Practical	Type
	Core - I - R PROGRAMMING FOR DATA ANALYTICS	**4**	**5**	**1**	**0**	**Theory / Practical**

Introduction :

R is a programming language and software environment used for statistical analysis, data modeling, and graphical representation and reporting. R is best tool for software programmers, statisticians and data miners who looking forward for to easily manipulate and present data in compelling ways

Course Outcome:

CO1	:	The data science course provides an in-depth understanding of the R language, R-studio, and R packages.
CO2	:	The data science training course also includes various statistical concepts such as linear and logistic regression, cluster analysis and forecasting and hypothesis testing.
CO3	:	Understand and use the various graphics in R for data visualization, Gain a basic understanding of various statistical concepts and strings and Lists
CO4	:	Define, understand and use the various apply functions and DPLYP functions ,Arrays and functions
CO5	:	Understand and use linear, non-linear regression models, and classification techniques for data analysis various types of file .

Unit I: [12 Periods]

R – OVERVIEW : Evolution of R- Features of R - R – ENVIRONMENT SETUP: Local Environment Setup- R – BASIC SYNTAX - R Command Prompt - R Script File – Comments - R – DATA TYPES : Vectors – Lists – Matrices – Arrays – Factors - Data Frames - R – VARIABLES : Variable Assignment - Data Type of a Variable - Finding Variables - Deleting Variables .

Unit II: [12 Periods]

R – OPERATORS: Types of Operators - Arithmetic Operators - Relational Operators - Logical Operators - Assignment Operators - Miscellaneous Operators - R – DECISION MAKING : R - If Statement - R – If...Else Statement- The if...else if...else Statement - R – Switch Statement - R – LOOPS : R - Repeat Loop - R - While Loop - R – For Loop - Loop Control Statements - R – Break Statement - R – Next Statement - R – FUNCTION : Function Definition - Function Components - Built-in Function - User-defined Function - Calling a Function - Lazy Evaluation of Function.

Unit III: [12 Periods]

R – STRINGS: Rules Applied in String Construction - String Manipulation - R – VECTORS : Vector Creation Accessing Vector Elements - Vector Manipulation - R – LISTS : Creating a List - Naming List Elements - Accessing List Elements - Manipulating List Elements - Merging Lists - Converting List to Vector - R – MATRICES : Accessing Elements of a Matrix - Matrix Computations

Unit IV: [12 Periods]

R – ARRAYS - Naming Columns and Rows - Accessing Array Elements - Manipulating Array Elements - Calculations Across Array Elements - R – FACTORS : Factors in Data Frame - Changing the Order of Levels - Generating Factor Levels - R – DATA FRAMES ; Extract Data from Data Frame - Expand Data Frame - R – PACKAGES - R – DATA RESHAPING - Joining Columns and Rows in a Data Frame - Merging Data Frames - Melting and Casting - Melt the Data - Cast the Molten Data

Unit V: [12 Periods]

R – CSV FILES - R – EXCEL FILE - R – BINARY FILES - R – XML FILES - R – PIE CHARTS - R – HISTOGRAMS – R Regressions - R – DATABASES : RMySQL Package - Connecting R to MySql - Inserting Data into the Tables - Creating Tables in MySql - Dropping Tables in MySql- R – NORMAL DISTRIBUTION : dnorm() - pnorm() - qnorm() - rnorm()- R – ANALYSIS OF COVARIANCE

Reference :

1. R programming Simple Easy learning Copyright 2016 by Tutorials Point (I) Pvt. Ltd.
2. Roger D. Peng, "R Programming for Data Science" Lean Publishing, 2014
3. Alain F. Zuur, Elena N. Ieno, Erik H.W.G. Meesters, "A Beginner's Guide to R" Springer, 2009
4. Introductory R A beginner's guide to data visualization and statistical analysis and programming in R , Robert Knell 2014
5.Using R for Data Analysis and Graphics Introduction, Code and Commentary , J H Maindonald Centre for Mathematics and Its Applications, Australian National University ,2002

Mapping of Course Outcomes with Program Outcomes:

Course Outcomes	Program Outcomes						
	PO1	**PO2**	**PO3**	**PO4**	**PO5**	**PO6**	**PO7**
CO1	H	L					
CO2			M	L			
CO3	L		H				
CO4			L			M	
CO5					M		M

H-High; M-Medium; L-Low

Semester : II

Subject Code	Subject Title	Credit	Lecture	Tutorial	Practical	Type
	Core - I – Machine Learning	**4**	**5**	**1**	**0**	**Theory**

Introduction :

To have a good understanding of the fundamental issues and challenges of *machine learning*: data, model selection, model complexity, etc. Be able to design and implement various *machine learning* algorithms in a range of real-world applications.

Course Outcome:

CO1	:	To understand and learn about the underlying technologies in machines.
CO2	:	To understand and learn about Linear Machines
CO3	:	To familiar with decision trees and logic programming
CO4	:	To understand and learn about clustering methods and its implementation
CO5	:	To familiar with Q learning, reinforcement learning and deductive learning

Unit I: [12 Periods]

Introduction: What is Machine Learning? - Wellsprings of Machine Learning - Varieties of Machine Learning - Learning Input-Output Functions : Types of Learning – Input Vectors – Outputs - Training Regimes - Noise - Performance Evaluation - Learning Requires Bias - Boolean Functions : Representation - Classes of Boolean Functions - Using Version Spaces for Learning - Neural Networks

Unit II: [12 Periods]

Linear Machines - Networks of TLUs - Training Feedforward Networks by Backpropagation : Notation - The Backpropagation Method - Computing Weight Changes in the Final Layer - Computing Changes to the Weights in Intermediate Layers - Synergies Between Neural Network and Knowledge-Based Methods - Statistical Learning : Using Statistical Decision Theory - Learning Belief Networks - Nearest-Neighbor Methods

Unit III: [12 Periods]

Decision Trees : Definitions - Supervised Learning of Univariate Decision Trees - Networks Equivalent to Decision Trees - Overfitting and Evaluation - Inductive Logic Programming : Notation and Definitions - A Generic ILP Algorithm - An Example - Inducing Recursive Programs - Choosing Literals to Add - Computational Learning Theory - Notation and Assumptions for PAC

Learning Theory - PAC Learning

Unit IV: [12 Periods]

Unsupervised Learning : What is Unsupervised Learning? - Clustering Methods - Hierarchical Clustering Methods - Temporal-Difference Learning : emporal Patterns and Prediction Problems - Supervised and Temporal-Difference Methods - Incremental Computation of the (ΔW) i - An Experiment with TD Methods - Theoretical Results - Intra-Sequence Weight Updating - An Example Application: TD-gammon

Unit V: [12 Periods]

Delayed-Reinforcement Learning : The General Problem- An Example – Temporal Discounting and Optimal Policies - Q-Learning - Discussion, Limitations, and Extensions of Q-Learning - Explanation-Based Learning : Deductive Learning - Domain Theories - An Example - Evaluable Predicates - Utility of EBL – Applications

Reference :

1. Introduction To Machine Learning, Nils J. Nilsson, Robotics Laboratory

2. Machine Learning, Tom M. Mitchell, Tata Mc Graw Hill
3. Introduction to Machine Learning, Ethem Alpaydın, MIT Press

Mapping of Course Outcomes with Program Outcomes:

Course Outcomes	Program Outcomes						
	PO1	**PO2**	**PO3**	**PO4**	**PO5**	**PO6**	**PO7**
CO1	M	H					
CO2			H	L			
CO3	L		M				
CO4			M			H	
CO5					H		L

H-High; M-Medium; L-Low

Semester : II

Subject Code	Subject Title	Credit	Lecture	Tutorial	Practical	Type
	Core - I – Information Storage and Management	**4**	**5**	**1**	**0**	**Theory**

Introduction :

Information storage is a central pillar of information technology. A large quantity of digital information is being created every moment by individual and corporate consumers of IT. This information needs to be stored, protected, optimized, and managed.

Course Outcome:

CO1	:	Provides detailed knowledge, practical training and insight into the implementation and management of various storage technologies with a focus towards applying these technologies in an information lifecycle paradigm.
CO2	:	Ability to understand components and the implementation of DAS.
CO3	:	Ability to understand components and the implementation of NAS.
CO4	:	To understand about backup and recovery techniques and technologies.
CO5	:	To monitor the storage infrastructure and management activities.

Unit I: [12 Periods]

Introduction to Information Storage and Management: Information Storage - Evolution of Storage Technology and Architecture - Data Center Infrastructure - Key Challenges in Managing Information . Storage System Environment: Components of a Storage System Environment - Disk Drive Components – Local components of the Host – Data Protection – RAID Implementation – Levels – Intelligent storage system.

Unit II: [12 Periods]

Direct-Attached storage : Introduction to SCSI : Types of DAS- Disk drive interface – Introduction to parallel SCSI - Storage Area Networks: Fiber Channel: Overview- Components of SAN - FC Connectivity - Zoning - FC Topologies.

Unit III: [12 Periods]

Networks attached storage : Benefits of NAS – NAS file I/O – Components of NAS – NAS Implementation – NFS – CIFS – NAS I/O Operations – IP SAN – FCIP – Content Address Storage – Storage Visualization.

Unit IV: [12 Periods]

Backup and Recovery : Backup Purpose - Methods – Process – Techniques – Data Consistency – Level Replication Technologies – Remote Replication technologies- Networks Infrastructure.

Unit V: [12 Periods]

Storage and Management : Securing the storage Infrastructure- Framework – Risk Triad – Storage Security Domain – BURA – SAN – NAS – IPSAN – Storage Management Architecture – Capacity – Performance – Security – Reporting – Developing an ideal solution.

Reference :

1. G. Somasundaram, Alok Shrivastava "Information Storage and Management - Storing, Managing, and Protecting Digital Information", Wiley Publishing, Inc.
2. Information Storage and Management, Wiley Publication ISBN: 978-81-265-2147-0
3. Marc Farley Osborne, "Building Storage Networks", Tata McGraw Hill
4. Robert Spalding, "Storage Networks: The Complete Reference", Tata McGraw Hill

Mapping of Course Outcomes with Program Outcomes:

Course Outcomes	Program Outcomes						
	PO1	**PO2**	**PO3**	**PO4**	**PO5**	**PO6**	**PO7**
CO1		M			H		L
CO2			M				H
CO3			L			H	
CO4	L		M	H			
CO5					H	H	

H-High; M-Medium; L-Low

Semester : II

Subject Code	Subject Title	Credit	Lecture	Tutorial	Practical	Type
	Core - I – Tcp / IP Protocol Suite	**4**	**5**	**1**	**0**	**Theory**

Introduction :

To provide students with an overview of the concepts and fundamentals networking protocols.

To familiarize with the basic taxonomy and terminology of computer networking area.

To experience the designing and managing of communication protocols while getting a good

exposure to the TCP/IP protocol suite

Course Outcome:

CO1	:	To master the terminology and concepts of the OSI reference model and the TCP-IP reference model.
CO2	:	To master the concepts of protocols and Standards, network interfaces, and design/performance issues in ethernet and wide area networks,
CO3	:	To be familiar with network layer services, features.
CO4	:	To be familiar with ARP & mobile IP
CO5	:	To be familiar with Multicasting and Multicast Routing Protocols

Unit I: [12 Periods]

Introduction and Underlying Technologies : Introduction : ARPANET - Birth of the Internet - Transmission Control Protocol/Internetworking Protocol (TCP/IP) - MILNET - CSNET - NSFNET – ANSNET – Protocols and Standards – Standards Organizations – Internet Administration : Internet Society (ISOC) - Internet Architecture Board (IAB) - Internet Engineering Task Force (IETF) - Internet Research Task Force (IRTF) - Internet Assigned Numbers Authority (IANA) and Internet Corporation for Assigned Names and Numbers (ICANN) - Network Information Center (NIC)

Unit II: [12 Periods]

The OSI Model and the TCP/IP Protocol Suite : Protocol layers - The OSI model - TCP/IP Protocol Suite – Addressing – Underlying Technologies : Wired Local Area Networks : IEEE

Standards - Frame Format - Addressing - Ethernet Evolution - Standard Ethernet - Fast Ethernet - Gigabit Ethernet - Ten-Gigabit Ethernet – Wireless LANS – Point to Point WANS – Switched WANS – Connected Devices

Unit III: [12 Periods]

Network Layer : Introduction to Network Layer : Introduction – Switching – Packet Switching at Network Layer - Network Layer Services – Other Network Layer Issues – IPV4 addresses – Classful Addressing – Classless Addressing – NAT – Delivery and Forwarding of IP Packets : Delivery – Forwarding – Structure of a Router – IPV4 : Introduction – Datagrams – Fragmentation

Unit IV: [12 Periods]

Address Resolution Protocol : Address Mapping – The ARP Protocol – ATM ARP – ARP Package – ICMPV4 : Introduction – Messages – Debugging Tools – ICMP Package – Mobile IP : Addressing – Agents – Three Phases – Inefficiency in Mobile IP - Unicast Routing Protocols (RIP, OSPF, and BGP) : Introduction – Intra and Inter domain Routing – Distance Vector Routing – RIP – Link State Routing – Path Vector Routing

Unit V: [12 Periods]

Multicasting and Multicast Routing Protocols : Introduction – Multi cast address – IGMP - Multicast Routing - Routing Protocols – MBONE – Transport Layer : Introduction to Transport Layer : Transport Layer Services - Transport Layer Protocols – User Datagram Protocol : Introduction – User Datagram – UDP Services – UDP Applications – UDP Package – Transmission Control Protocol : TCP Services – TCP Features – Segment – A TCP Connection – State Transition Diagram – Flow Control – Error Control – Congestion Control

Reference :

1. TCP / IP Protocol Suite, 4th Edition, Behrouz A.Forouzan, Mc Graw Hill Publication, 2010.
2. A Guide to the TCP / IP Protocol Suite, 2nd Edition, Floyd Wilder
3. Internetworking with TCP / IP Principles, Protocols and Architecture, Douglas E.Comer, 6th edition, 2014
4. Internetworking With TCP/IP Volume II: Design, Implementation, and Internals (with D. Stevens), Third Edition, 1999

Mapping of Course Outcomes with Program Outcomes:

Course Outcomes	Program Outcomes						
	PO1	**PO2**	**PO3**	**PO4**	**PO5**	**PO6**	**PO7**
CO1	H	M	M				
CO2			L	H			
CO3				M		M	
CO4		M	H				
CO5					L		H

H-High; M-Medium; L-Low

Semester: II

Subject Code	Subject Title	Credit	Lecture	Tutorial	Practical	Type
	Core - I -R PROGRAMMING FOR DATA ANALYTICS LAB	**4**	**5**	**1**	**0**	**Practical**

Course Outcome:

The focus of this lab is to introduce you to R and the R Commander (a graphical user interface to R). To use R to analyze data, you will need to become familiar with the technical components of this software package. This lab will help familiarize you with the R software, including how to access data files, the various base components and how to define new variables and how to enter data.

List of programs:

1. Introduction to R Defining and Downloading R and installing R Programming
2. Write a simple R program to perform some Built-in mathematical functions
3. Write a simple R program to perform vector operations
4. Write a simple R program to Reading data from csv files, inspection of data.
5. Write a simple R program to Read data from Excel files
6. Write a simple R program to Working with datasets: Inspection, using variables, attaching
7. Write a simple R program to Transformation of variables, subsets of datasets and Merging datasets
8. Write a simple R program to perform Graphs
9. Write a simple R program to perform Regression
10. Write a simple R program to perform Analysis of variance (ANOVA)