x

AC 21-03-2023

R. A. PODAR COLLEGE

OF

COMMERCE AND ECONOMICS

(AUTONOMOUS)

MATUNGA, MUMBAI

SYLLABUS

FOR UNDER GRADUATE PROGRAMME

S. Y. B.SC. (DATA SCIENCE AND ANALYTICS)

SEMESTER III & IV

CHOICE BASED CREDIT AND GRADING SYSTEM (CBCGS)

With effect from the academic year 2023-24

www.rapodar.ac.in

Course Code	Semester III	Credits	Course Code	Semester IV	Credits
	Elective Courses (EC)			**Elective Courses (EC)**	
	Discipline Specific Elective (DSE) Courses			**Discipline Specific Elective (DSE) Courses**	
60301	Research Methods and Ethics	**02**	60401	Testing of Hypothesis	**02**
60302	Research Methods and Ethics Practical	**02**	60402	Testing of Hypothesis Practical	**02**
60303	Data Structures and Algorithms using Python	**02**	60403	Big Data	**02**
60304	Data Structures and Algorithms using Python Practical	**02**	60404	Big Data Practical	**02**
	Ability Enhancement Courses (AEC)			**Ability Enhancement Courses (AEC)**	
	Ability Enhancement Compulsory Courses (AECC)			**Ability Enhancement Compulsory Courses (AECC)**	
60305	Economics	**02**	60405	Artificial Intelligence	**02**
60306	Economics Practical	**02**	60406	Artificial Intelligence Practical	**02**
	Skill Enhancement Courses (SEC)			**Skill Enhancement Courses (SEC)**	
60307	Data Warehousing and Mining	**02**	60407	Operating Systems	**02**
60308	Data Warehousing and Mining Practical	**02**	60408	Operating Systems Practical	**02**
	Core Courses (CC)			**Core Courses (CC)**	
60309	Linear Algebra and Discrete Mathematics	**02**	60409	Numerical Methods	**02**
60310	Linear Algebra and Discrete Mathematics Practical	**02**	60410	Numerical Methods Practical	**02**
	Total Credits	**20**		**Total Credits**	**20**

Semester III

Elective Courses (EC)
Discipline Specific Elective (DSE)Courses
Research Methods and Ethics

B. Sc. (Data Science and Analytics)		Semester – III	
Course Name: Research Methods and Ethics		Course Code: 60301	
Periods per week (1 Period is 48 minutes)		5	
Credits		2	
		Hours	Marks
Evaluation System	Theory Examination	2	60
	Internal	--	40

Course Objectives:

1. To impart analytical skill in solving complex problems.
2. To foster the ability to critically think in developing robust, extensible and highly maintainable solutions to simple and complex problems.
3. To explore the unknown and unlock new possibilities in different dimensions of the system.
4. To portray accurately the characteristics of a particular individual, situation or a group under study.

Unit	Details	Lectures
I	**Research Methodology-An Introduction:** Meaning of Research, Objectives of Research, Motivation in Research, Types of Research, Research Approaches, Significance of Research, Research Methods versus Methodology, Research and Scientific Method, Research Process, Criteria of Good Research, Problems Encountered by Researchers in India **Defining the Research Problem:** What is a Research Problem? Selecting the Problem, Necessity of Defining the Problem, Technique Involved in Defining a Problem, An Illustration **Review of literature-** Review of Literature: Need - sources to collect review - how to write review of literature survey using Internet.	**15**
II	**Research Design:** Concept and Importance in Research - Features of a good research design - Exploratory Research Design - concept, types and uses, Descriptive Research Designs - concept, types and uses. Experimental Design: Basic Principles of Experimental Designs. Concept of Independent & Dependent variables. **Measurement and Scaling:** Concept of measurement- what is measured? Problems in measurement in research- Validity and Reliability. Levels of measurement Nominal, Ordinal, Interval, Ratio. Technique of Developing Measurement Tools, Scaling, Meaning of Scaling, Scale Classification Bases, Important Scaling Techniques, Scale Construction Techniques **Method of Data Collection:** Collection of Primary Data, Observation Method, Interview Method, Collection of Data through Questionnaires, Collection of Data through Schedules, Difference between Questionnaires and Schedules, Some Other Methods of Data Collection, Collection of Secondary Data, Selection of Appropriate Method for Data Collection, Case Study Method, Pilot study	**15**
III	**Sampling Design:** Census and Sample Survey, Implications of a Sample Design, Steps in Sampling Design, Criteria of Selecting a Sampling Procedure, Characteristics of a Good Sample Design, Different Types of Sample Designs, How to Select a Random Sample? Random Sample from an Infinite Universe, Complex Random Sampling Designs **Sampling Fundamentals:** Need for Sampling, Some Fundamental Definitions, Important Sampling Distributions, Central Limit Theorem, Sampling Theory, Concept of Standard Error, Estimation, Estimating the Population Mean (m), Estimating Population Proportion, Sample Size and its Determination, Determination of Sample Size through the Approach Based on Precision Rate and Confidence Level, Determination of Sample Size through the Approach, **Processing and Analysis of Data:** Processing Operations, Some Problems in Processing, Elements/Types of Analysis, Statistics in Research	**15**

I V	**Testing of Hypotheses:** What is a Hypothesis? Basic Concepts concerning Testing of Hypotheses, Procedure for Hypothesis Testing, Flow Diagram for Hypothesis Testing, Measuring the Power of Hypothesis Test, Tests of Hypotheses, Important Parametric Tests, Hypothesis Testing of Means, Hypothesis Testing for Differences between Means, Limitations of the Tests of Hypotheses **Interpretation of Data and Paper Writing** – Report Writing - types - Format - Principles of Writing report - Documentation: Footnotes and Endnotes - Bibliography - Citation Model - APA Model - guidelines for writing references. **Ethical issues related to publishing-** Plagiarism and Self-Plagiarism**.**	**15**

Books and References:

Sr. No.	Title	Author/s	Publisher	Edition	Year
1.	Research Methodology – Methods and techniques	C. R. Kothari	New Age International (P) Ltd.,Publishers	---	---
2.	Business Research Methods	Donald R. Cooper Pamela Schindler	McGraw Hill/Irwin	12th Ed	
3.	Business Research Methods	Allan Bryman Emma Bell	OXFORD University Press	---	---
4.	RESEARCH METHODOLOGY - a step by step guide for beginners	Ranjit Kumar	SAGE Publication Ltd	---	---
5.	Research Methods for Business Students	Mark Saunders Philip Lewis Adrian Thornhill	Pearson Education Limited	---	---

Course Outcome:

CO 1: Learner understands the reasons for doing research, the applications of research, characteristics and requirements of the research process, types of research and Research paradigms.

CO 2: Learner is applying major approaches to information gathering, the relationship between attitudinal and measurement scales Methods for exploring attitudes in research.

CO 3: Learner is able to analyze data in qualitative and quantitative studies, application of IT in data analysis.

CO 4: Learner is able to write a research report and use Information Technology in Research

CO 5: Learners practicing ethical codes and practices of conduct research.

Elective Courses (EC)

Discipline Specific Elective (DSE)Courses

Research Methods and Ethics Practical

B. Sc (Data Science and Analytics)		Semester – III	
Course Name: **Research Methods and Ethics Practical**		Course Code: 60302	
Periods per week (1 Period of 48 minutes)		3	
Credits		2	
		Hours	Marks
Evaluation System	Practical Examination	2½	50
	Internal	--	--

List of Practical:	
1.	**Introduction to EndNote, Zotero or Mendeley:** Integration with Word and adding citation and creating bibliographies. Creating your own library. Creating references from website. Creating references manually.
2.	Visit the college library or nearby research center or from the internet collect 5 titles of research papers/theses and classify them according to types of research. Discuss how the problems are delineated, how they are relevant to scientific method etc
3.	Preparation of a review article and justify the significance of their study
4.	Preparation of questionnaire for micro-level educational survey. Preparation of a sampling design given the objectives and research questions/hypotheses of a research study.
5.	Prepare 1 proposal on an identified research problem. Checking and removing plagiarism using Plagiarism Detection Software

Evaluation scheme

I Continuous Internal Assessment (C.A.)– 40 Marks

1) Assessment 1 (20 Marks)
2) Assessment 2 (20 Marks**)**

II Semester End Examination (SEE) - 60 Marks

QUESTION PAPER PATTERN OF (SEE)

Maximum Marks: 60 Marks

Time: 2 Hours

Note: 1) Attempt all Questions

2) Any three sub questions to be attempted. All sub questions carry equal marks

Question No	Particulars	Marks
Q -1 (Unit I)	A) B) C) D)	15 Marks
Q -2 (Unit II)	A) B) C) D)	15 Marks
Q -3 (Unit III)	A) B) C) D)	15 Marks
Q4 (Unit IV)	A) B) C) D)	15 marks

Practical Exam: 50 marks
Certified copy journal is essential to appear for the practical examination.

1.	**Practical Question**	**40**
2.	**Journal**	**5**
3.	**Viva Voce**	**5**

Elective Courses (EC)

Discipline Specific Elective (DSE)Courses

Data Structures and Algorithms Using Python

B. Sc (Data Science and Analytics)		Semester – III	
Course Name: **Data Structures and Algorithms Using Python**		Course Code: 60303	
Periods per week (1 Period is 48 minutes)		5	
Credit		2	
		Hours	Marks
Evaluation System	Theory Examination	2	60
	Internal	--	40

Course Objectives:

1. To learn the essential Python data structures.
2. To learn the significant Python implementation of popular data structures
3. To learn about various data structure algorithms and design paradigms
4. To acquire knowledge of how to create complex data structures.
5. To acquire basic understanding of complex data structures such as trees and graphs and their applications.

Unit	Details	Lectures
I	Python Objects & Object-Oriented Programming: Goals, Principles, and Patterns, Overview of data types and objects, Classes and object programming, Class Definitions, Inheritance, Data encapsulation and properties, Namespaces and Object-Orientation, Shallow and Deep Copying Python Data Types and Structures:- Modules for data structures and algorithms- Collections, Deques, Chain Map objects Counter, Counter objects, Ordered dictionaries defaultdict, Learning about named tuples Arrays Principles of Algorithm Design: An introduction to algorithms, Algorithm design paradigms Recursion and backtracking, Backtracking, Divide and conquer - long multiplication The recursive approach Runtime analysis Asymptotic analysis Big O notation, Composing complexity classes Omega notation, Theta notation, Amortized analysis	**12**

II	Lists and Pointer Structures: Arrays-Pointer structures Singly linked lists-Singly linked list class, The append operation, A faster append operation, Getting the size of the list, Improving list traversal, Deleting nodes, List search, Clearing a list Doubly linked lists-A doubly linked list node Doubly linked list class Append operation The delete operation List search Circular lists-Appending elements, Deleting an element in a circular list, Iterating through a circular list Stacks: Stack implementation, Push operation, Pop operation, Peek operation, Bracket-matching application,	**12**
III	Queues:- List-based queues, Stack-based queues Node-based queues, Application of queues Media player queues Trees: Terminology, Tree nodes, Tree traversal, Depth-first traversal In-order traversal and infix notation, Pre-order traversal and prefix notation, Post-order traversal and postfix notation, Breadth-first traversal, Binary trees-Binary search trees, Binary search tree implementation, Binary search tree operations, Finding the minimum and maximum nodes Inserting nodes Deleting nodes, Searching the tree, Benefits of a binary search tree, Balancing trees, Expression trees, Parsing a reverse Polish expression, Heaps, Ternary search tree	**12**
IV	Hashing and Symbol Tables:- Hashing- Perfect hashing functions Hash tables-Storing elements in a hash table, Retrieving elements from the hash table, Testing the hash table, Using [] with the hash table, Non string keys, Growing a hash table, Open addressing, Chaining, Symbol tables Graphs and Other Algorithms:- Graphs-Directed and undirected graphs, Weighted graphs, Graph representations, Adjacency lists, Adjacency matrices, Graph traversals- Breadth-first traversal, Depth-first search, Other useful graph methods, Priority queues and heaps- Insert operation, Pop Operation, Selection Algorithm	**12**
V	Sorting: Sorting algorithms- Bubble sort algorithms, Insertion sort algorithms, Selection sort algorithms, Quick sort algorithms Selection Algorithms: Selection by sorting, Randomized selection Quick Select, Deterministic selection-Pivot selection Median of medians Partitioning step Pattern Matching Algorithms: The brute-force algorithm, The Rabin Karp algorithm	**12**

Books and References:					
Sr. No.	**Title**	**Author/s**	**Publisher**	**Edition**	**Year**
1.	Hands-On Data Structures and Algorithms with Python	Basant Agarwal, Benjamin Baka	Packt Publishing	2nd	2018
2.	Data Structure and algorithm Using Python	Rance D. Necaise	Wiley India Edition		2016
3.	Data Structure and Algorithm in Python	Michael T. Goodrich, RobertomTamassia	Wiley India Edition		2016
4.	Data Structure and Algorithmic Thinking with Python	Narasimha Karumanchi	Careermonk Publications		2015
5.	Fundamentals of Python: Data Structures	Kenneth Lambert	Delmar Cengage Learning		2018

Course Outcomes:

CO 1: Learner is capable of choosing appropriate data structure in Python for specified problems and algorithms.

CO 2: Learner is able to implement Linked list and Stack data structure in various domains.

CO 3: Learner is able to implement Tree and Queue data structures and use their operation.

CO 4: Learner has ability to apply of Hashing techniques, Symbol Table and Graph Algorithms appropriately.

CO 5: Learner has skills to handle sorting, searching and pattern matching on various data structures.

Data Structures and Algorithms Using Python Practical

B. Sc. (Data Science)		Semester – III	
Course Name: - Data Structure and Algorithm Using Python Practical		**Course Code:** 60304	
Periods per week (1 Period is 50 minutes)		**5**	
Credits		**2**	
		Hours	**Marks**
Evaluation System	**Theory Examination**	**2**	**60**
	Internal	**--**	**40**

List of Practical: (Any 5 practicals to be submitted in the Journal)	
1	**General Python Programs**
a	Write Python Program to demonstrate the use of various Python Data Types and Structures
b	Write Python Program to demonstrate OOP Concepts including Class, Objects, Inheritance and encapsulation.
c	Write Python Program to implement array and operations of arrays.
2	**List and Pointer Structure**
a	Write Python Program to create singly linked list and various operations on it.
b	Write Python Program to create doubly linked list and various operations on it.
c	Write Python Program to create circular linked list and various operations on it.
3	**Stacks and Queues**
a.	Write Python Program to implement stack and demonstrate push, pop and peek operations.
b.	Write Python Program to implement stack for Bracket-matching application
c.	Write Python Program to implement list based queues and demonstrate various operations on it.
d.	Write Python Program to implement stack based queues and demonstrate various operations on it.

e.	Write Python Program to implement Node based queues and demonstrate various operations on it.
f.	Write Python Program to implement queue data structure for simulating media player playlist queues.
4	**Trees**
a.	Write Python Program to implement tree data structure and demonstrate depth- first traversal
b.	Write Python Program to implement tree data structure and demonstrate breadth first traversal
c.	Write Python Program to implement binary search tree to find the minimum node.
d.	Write Python Program to implement binary search tree to find the minimum node.
e.	Write a Python implementation to demonstrate the insert and delete method to add/delete the nodes in the BST.
f.	Python implementation to search the node in the BST
g.	Write a python program build up a tree for an expression written in postfix notation and evaluate it.
g.	Write a python program build up a tree for an expression written in postfix notation and evaluate it.
5	**Hashing and Symbol Tables**
a.	Write a Python Program to demonstration of computing Hash for given strings.
b.	Write a Python program to implement hash table for storing and searching values from it.
c.	Write aa Python Program to create Symbol Table
6	**Graphs**
a.	Write a Python program to store and display Graph data structure using adjacency matrix.
b.	Write a Python Program to implement Graph traversal (BFS/DFS) based on above practical.
c.	Write a Python program to implement priority queue and heap operations
7	**Searching**
a.	Write a Python Program for implementation in Python for the linear search on an unordered list of items
b.	Write a Python Program for implementation in Python for the linear search on an ordered list

	of items
c.	Write a Python Program for implementation of the binary search algorithm on an ordered list of items
d.	Write a Python Program for implementation of implementation of the interpolation search algorithm
8	**Sorting**
a.	Write a Python Program for implementing Insertion Sort.
b.	Write a Python Program for implementing Bubble Sort.
c.	Write a Python Program for implementing Quick Sort.
d.	Write a Python Program for implementing Selection Sort.
9	**Selection Algorithms**
a.	Write a Python Program to implement Randomized Selection
b.	Write a Python Program to implement Deterministic Selection
10	**Application**
a.	Write a Python Program to create an application for storing Polynomial
b.	Write a Python Program to create an application for adding two Polynomials

Evaluation scheme

I Continuous Assessment (C.A.)– 40 Marks

1) Assessment 1 (20 Marks)

2) Assessment 2 (20 Marks**)**

II Semester End Examination (SEE) - 60 Marks

QUESTION PAPER PATTERN OF (SEE)

Maximum Marks: 60 Marks

Time: 2 Hours

Note: 1) Attempt all Questions

2) Any three sub questions to be attempted. All sub questions carry equal marks

Question No	Particulars	Marks
Q -1 (Unit I)	A) B) C) D)	12 Marks
Q -2 (Unit II)	A) B) C) D)	12 Marks
Q -3 (Unit III)	A) B) C) D)	12 Marks
Q4 (Unit IV)	A) B) C) D)	12 marks
Q5 (Unit V)	A) B) C) D)	12 marks

Practical Exam: 50 marks

Certified copy journal is essential to appear for the practical examination.

1.	**Practical Question**	**40**
2.	**Journal**	**5**
3.	**Viva Voce**	**5**

x

Economics

B. Sc. (Data Science)		Semester – III	
Course Name:- Economics		**Course Code:** 60305	
Periods per week (1 Period is 50 minutes)		**5**	
Credits		**2**	
		Hours	**Marks**
Evaluation System	**Theory Examination**	**2**	**60**
	Internal	**--**	**40**

Course Objectives:

1. To understand Fundamental economic ideas and the operation of the economy on a national scale.

2. Basic Understanding of production, distribution and consumption of goods and services, the exchange process, the role of government, the national income and its distribution, GDP, consumption function, savings function, investment spending and the multiplier principle

3. Acquire basic knowledge of the influence of government spending on income and output.
4. Develop ability to analyze monetary policy, including the banking system and the Federal Reserve System.

Unit	Details	Lectures
I	**First Principles:** · What is the difference between macro and micro economics? The central choices of economic decision making: what, how and for whom to produce? · The participants in the market economy Key concepts used in economic analysis: Scarcity, choice, opportunity cost Marginal analysis and choice Ceteris Paribus or _everything else held constant. · Positive and normative economics and using theories and models to measure economic events · Criteria for evaluation of economic policy and policy proposals Economic systems – the market economy, mixed economies & command economies · Review of expressing relationships between economic variables using graphs Economic Models: · Trade-offs and Trade: Defining the resources used in the production of goods and services · The production possibility frontier applied to the concept of opportunity cost/tradeoffs and to marginal costs and benefits; increasing marginal opportunity costs. · Productive efficiency; inefficient choices and unattainable choices.	**12**

x

	· Using the frontier to illustrate economic growth, attainment of new resources, technological change, and more efficient production.	
II	**Supply and Demand: Product and Resource Markets –** · Role of households (consumers) and firms What is a market? · Consumer demand and the ―Law of Demand ·Change in quantity demanded vs. Variations in demand, Causes of a shift in demand. ·Law of Supply: The positive relationship between price and quantity supplied. ·Change in quantity supplied vs. Variation in supply, Causes of a shift in Supply ·Applications (examples) of Demand and Supply graphs; Market demand, market supply and market equilibrium Government price controls: price ceilings, price floors. **Macroeconomics:** · Theory and Policy: The Business Cycle in Market Economies; · Components of aggregate demand and aggregate supply, shifts in the AD and AS curves. · The roots of macroeconomics: John Maynard Keynes and the Great Depression Classical vs. Keynesian economics; the short-run vs. long run model of macroeconomic equilibrium. · The Keynesian short-run model and the classical economists 'long-run model Keynes 'challenge to Say 's Law. · The role of Government Concerns of Inflation (boom times) and deflation (severe economic downturns) The impact of recession on trade imbalances	**12**
III	**Unemployment and Inflation:** · Measuring employment and unemployment. Who is not counted in the Government 's official count of the unemployed? Types of unemployment; cyclical unemployment and the business cycle. · The difference between the _household survey '(the civilian labor force) and the establishment survey '(number of payroll jobs added by employers). · The labor force participation rate Unemployment and the changes in the global economy **Gross Domestic Product:** · Measuring the economy 's output of goods and services; Government Sector: federal state and local government in the economy the financial sector; the international sector, The three markets: goods and service, labor market, money market · Nominal and real GDP; The difference between GNP and GDP Expenditure Measure of GDP: consumption by households, businesses, government and the rest of the world (Net exports) Income · Measure of GDP: Income from labor, rent, interest, proprietors 'income, profit Value added approach vs. measure of final goods and services produced What GDP Does Not Include; alternative measures of GDP	**12**

x

IV	**Measuring inflation, the consumer price index:** · What does it say about the state of the economy? Real vs. nominal income and earnings Real and Nominal rates of interest Costs and causes of inflation · Fiscal policy: Defining fiscal policy: taxation and spending to achieve macroeconomic goals ·The role of government in the U.S. economy Fiscal policy and the Recession of 2007 – 2009 · The Employment Act of 1946 A history of U.S. fiscal policy since the early 20th century ·The multiplier effect Government spending and taxation Automatic stabilizers: the income tax, unemployment insurance · Discretionary tax and spending policy Progressive, proportional and regressive taxes and their impacts ·Fiscal Policy Lags The circular flow diagram with government spending and taxation ·Budget deficits and surpluses; Government debt and deficits: Are they the same thing?	**12**
V	**Money, Banking and the Federal Reserve System:** · What is money? Commodity and fiat monies; the barter system Money as a medium of exchange; · Money supply defined: M1 and M2 Gold and the money supply: Why was the gold standard adopted (1873) and why was it later eliminated (1971)? Monetary role of banks; · Establishment of bank reserves; The T-account (assets and liabilities) Bank regulation: the FDIC deposit insurance; capital requirements; the discount window at the Fed. **Monetary Policy:** ·The structure of the Federal Reserve System How the Fed regulates the money supply: reserve requirements, the discount rate, open market operations; the goals of monetary policy · The federal funds rate; fed funds market Banking legislation and deregulation since the 1980‘s Growth of the ―Shadow Banking System‖ and the financial crisis of 2007-2009 ·The role of credit, debit cards and electronic money in the money supply · Role of financial intermediaries – modern depository institutions Savings and Loan crisis of the late 1980‘s ·The financial crisis of 2008 and the Federal Reserve ‘s policy response, how the banking system creates and expands money in circulation ·The difference between treasury bonds and bonds issued by the Fed Policies during the 2007 – 2009 Recession	**12**

Books and References:					
Sr. No.	Title	Author/s	Publisher	Edition	Year
01	Macroeconomics	Krugman and Wells, Eds.,	Worth Publishers	3 rd	2012
02	Macroeconomics	Leeds, Michael A.,von Allmen, Peter and Schiming,Richard C	Pearson Education	1st	2006
03	Lectures in Quantitative Economics with Python	Thomas J. Sargent and John Stachurski	----	----	2019

Course Outcome

CO1:Learner understands the basic economic decisions that underline the economic process: What and how to produce goods and services and how they are distributed.

CO2:Learner is able to apply of the concepts of scarcity, choice and opportunity cost to analyze the workings of a market economy.

CO3:Learneris able to demonstrate a firm knowledge of the interrelationships among consumers, government, business and the rest of the world in the U.S. macroeconomy.

CO4:Learner is able to identify the process of how the nation‘s output of goods and services is measured through the national income and product accounts; clearly comprehend the income and expenditure approaches to measuring national output and national income.

CO5:Learner is capable to clearly illustrate the specific roles and functions of monetary and fiscal policy in the economy and explain how these are applied to the process of shaping economic policy and stabilizing the economy, specifically regarding controlling inflation, promoting full employment and facilitating economic growth.

x

Economics Practical

B. Sc. (Data Science)		Semester – III	
Course Name: Economics Practical		**Course Code: 60306**	
Periods per week (1 Period is 50 minutes)		**3**	
Credits		**2**	
		Hours	**Marks**
Evaluation System	**Practical Examination**	**2½**	**50**
	Internal	**--**	

List of Practical:	
1	Relationship between Economic variable using tableau/
2	Application of demand and supply graph
3	Visualization of Keynesian short run model in Tableau
4	Understanding Macroeconomics using Tableau
5	Understanding Gross Domestic Product using Tableau

Evaluation scheme

I Continuous Assessment (C.A.)– 40 Marks

1) Assessment 1 (20 Marks)

2) Assessment 2 (20 Marks**)**

II Semester End Examination (SEE) - 60 Marks

QUESTION PAPER PATTERN OF (SEE)

Maximum Marks: 60 Marks

Time: 2 Hours

Note: 1) Attempt all Questions

2) Any three sub questions to be attempted. All sub questions carry equal marks

Question No	Particulars	Marks
Q -1 (Unit I)	A) B) C) D)	12 Marks
Q -2 (Unit II)	A) B) C) D)	12 Marks
Q -3 (Unit III)	A) B) C) D)	12 Marks
Q4 (Unit IV)	A) B) C) D)	12 marks
Q5 (Unit V)	A) B) C) D)	12 marks

Practical Exam: 50 marks

Certified copy journal is essential to appear for the practical examination.

1.	**Practical Question**	**40**
2.	**Journal**	**5**
3.	**Viva Voce**	**5**

Data Warehousing and Mining

B. Sc. (Data Science)		Semester – III	
Course Name: Data Warehousing and Mining		**Course Code: 60307**	
Periods per week (1 Period is 50 minutes)		**5**	
Credits		**2**	
		Hours	**Marks**
Evaluation System	**Theory Examination**	2	60
	Internal	--	40

Course Objectives:

1. To understand business intelligence for an enterprise and review data warehouse with architectural types and architectural building blocks
2. To discuss and understand changing dimensions and learn about aggregate tables and determine their usage.
3. To learn the basics of data mining, understand the need and the process of data mining in contrast with machine learning.
4. To study the use of classification and clustering techniques for Data Mining.
5. To appreciate the use of various data mining algorithms and learn about their specific applications.

Unit	Details	Lectures
I	**THE COMPELLING NEED FOR DATA WAREHOUSING**: Escalating Need for Strategic Information, Failures of Past Decision-Support Systems, Operational Versus Decision-Support Systems, Data Warehousing—The Only Viable Solution, Data Warehouse Defined, The Data Warehousing Movement, Evolution of Business Intelligence **DATA WAREHOUSE:** The Building Blocks: Defining Features, Data Warehouses and Data Marts, Architectural Types, Overview of The Components, Metadata in The Data Warehouse **TRENDS IN DATA WAREHOUSING:** Continued Growth in Data Warehousing, Significant Trends, Emergence of Standards, Web-Enabled Data Warehouse **ARCHITECTURAL COMPONENTS:** Understanding Data Warehouse Architecture, Distinguishing Characteristics, Architectural Framework, Technical Architecture, Architectural Types **THE SIGNIFICANT ROLE OF METADATA:** Why Metadata Is Important, Metadata Types By Functional Areas, Business Metadata, Technical Metadata, How To Provide Metadata	12
II	**PRINCIPLES OF DIMENSIONAL MODELING**: From Requirements to Data Design, The Star Schema, Star Schema Keys, Advantages of The Star Schema, Star Schema: Examples **DIMENSIONAL MODELING: ADVANCED TOPICS**: Updates to The Dimension Tables, Miscellaneous Dimensions, The Snowflake Schema, Aggregate Fact Tables, Families of Stars **DATA EXTRACTION, TRANSFORMATION, AND LOADING:** ETL Overview, ETL Requirements and Steps, Data Extraction, Data Transformation, Data Loading, ETL Summary, Other Integration Approaches	12
III	**INTRODUCTION TO DATA MINING**: Introduction to Data Mining, Need of Data Mining, What Can Data Mining Do and Not Do? Data Mining Applications, Data Mining Process, Data Mining Techniques, Difference between Data Mining and Machine Learning **BEGINNING WITH IRIS DATASET IN R:** Understanding Fisher's Iris Flower Dataset, Preparing the Dataset, Understanding A RFF, Working with a Dataset in Weka, Working with the Iris dataset in R Data Preprocessing: Need for Data Preprocessing, Data Preprocessing Methods **CLASSIFICATION:** Introduction to Classification, Types of Classification, Input and Output Attributes, Guidelines for Size and Quality of the Training Dataset, Introduction to the Decision Tree Classifier, Naive Bayes Method, Understanding Metrics to Assess the Quality of Classifiers	12
IV	**CLUSTER ANALYSIS:** Introduction to Cluster Analysis, Applications of Cluster Analysis, Desired Features of Clustering, Distance Metrics, Major Clustering Methods/Algorithms, Partitioning Clustering, **HIERARCHICAL CLUSTERING ALGORITHMS:** Web Mining and Search Engines: Introduction, Web Content Mining, Web Usage Mining, Web Structure Mining, Hyperlink Induced Topic Search algorithm, Introduction to Modern Search Engines, Working of a Search Engine, PageRank Algorithm, Precision and Recall	12

V	INTRODUCTION TO ASSOCIATION RULE MINING: Defining Association Rule Mining, Representations of Items for Association Mining, The Metrics to Evaluate the Strength of Association Rules, The Naive Algorithm for Finding Association Rules, Approaches for Transaction Database Storage THE APRIORI ALGORITHM, Closed and Maximal Item sets, The A priori–TID Algorithm for Generating Association Mining Rules, Direct Hashing and Pruning (DHP), Dynamic Itemset Counting (DIC), Mining Frequent Patterns without Candidate Generation (FP Growth)	12

Books and References:

Sr. No.	Title	Author/s	Publisher	Edition	Year
1.	DATA WAREHOUSING FUNDAMENTALS FOR IT PROFESSIONALS	PAULRAJ PONNIAH	Wiley	Second	2010
2.	Data Mining and Data Warehousing : Principles and Practical Techniques	Parteek Bhatia	Cambridge University Press	First	2019
3.	The Data Warehouse Toolkit	Ralph Kimball Margy Ross	Wiley	Third	2013
4.	Encyclopedia of Data Warehousing and Mining	John Wang	Information Science Reference	Second	2008
5.	Data Mining and Data Warehousing	S.K.Mourya Shalu Gupta	Alpha Science International Ltd	First	2013

Course Outcomes:

CO1: Learner is able to demonstrate knowledge of business intelligence, data warehouse with clear understanding of architectural types and will be able to establish the relationship between architectural building blocks.

CO2: Learner is able to elaborate changing dimensions with respect to current trends & using aggregate tables.

CO3: Learner is able to handle the processes of data preprocessing, data transformation and data reduction.

CO4: Learner has knowledge of using various Data Mining techniques for classification and clustering.

CO5: Learneris able to align the Data Mining techniques for analyzing the datasets using tools like Weka, R or Python

x

Data Warehousing and Mining Practical

B. Sc. (Data Science)		Semester – III	
Course Name: Data Warehousing and Mining Practical		**Course Code: 60308**	
Periods per week (1 Period is 50 minutes)		**3**	
Credits		**2**	
		Hours	**Marks**
Evaluation System	**Practical Examination**	**2½**	**50**
	Internal	**--**	

List of Practical: (Any 5 To be submitted for evaluation)	
1.	**Data warehouse design**
a.	Design dimension tables.
b.	Design fact tables.
c.	Create an indexed view and rebuild column store indexes.
2.	**Data Warehouse with Azure**
a.	Create an Azure SQL Data Warehouse Project.
b.	Develop tables in Azure SQL Data Warehouse.
c.	Migrate Data Warehouse to Azure.
d.	Pause and remove Azure data warehouse.
3.	**Data Warehouse implementation and use**
a.	Cleanse data with SQL Server Data Quality Services.
b.	Create a custom knowledge base.
c.	Install Master Data Services and IIS.
d.	Configure MDS and deploy sample MDS models.
e.	Install MDS excel add-in and Update master data in excel.
f.	Consume the data from the warehouse.
4.	**Working with Data and Data Preprocessing**

a.	Demonstrate the use of ARFF files taking input and display the output of the files.
b.	Create your own excel file. Convert the excel file to .csv format and prepare it as ARFF files.
c.	Preprocess and classify Customer dataset. http://archive.ics.uci.edu/ml/
d.	Perform Preprocessing, Classification techniques on Agriculture dataset. (http://archive.ics.uci.edu/ml/)
e.	Preprocess and classify Weather dataset. http://archive.ics.uci.edu/ml/
f.	Perform data Cleansing of customer dataset. http://archive.ics.uci.edu/ml/ www.kdnuggets.com/datasets/
5.	**Performing classification on data sets**
a.	Applying Naïve Bayes on Dataset for classification
b.	Creating the Testing Dataset
c.	Decision Tree Operation with R
d.	Naïve Bayes Operation using R
e.	Classify the dataset using a decision tree. www.kdnuggets.com/datasets/
6.	**Simple Clustering**
a.	Perform Clustering technique on Customer dataset. http://archive.ics.uci.edu/ml/
b.	Perform Clustering technique on Agriculture dataset. http://archive.ics.uci.edu/ml/
c.	Perform Clustering technique on Weather dataset.http://archive.ics.uci.edu/ml/
7.	**Implementing Clustering with R**
h.	Clustering Fisher‘s Iris Dataset with the Simple k-Means Algorithm
i.	Handling Missing Values
j.	Results Analysis after Applying Clustering
k.	Classification of Unlabeled Data
l.	Clustering in R using Simple k-Means
8.	**Implementing Apriori Algorithm with R**

a.	Applying Predictive Apriori in Weka
b.	Applying the Apriori Algorithm in Weka on a Real World Dataset
c.	Applying the Apriori Algorithm in Weka on a Real World Larger Dataset
d.	Applying the Apriori Algorithm on a Numeric Dataset
9.	**Implementing Association Mining with R**
a.	Applying Association Mining in R
b.	Application of Association Mining on Numeric Data in R
c.	Perform Association technique on Agriculture dataset. http://archive.ics.uci.edu/ml/,www.kdnuggets.com/datasets/
d.	Perform Association technique on Agriculture dataset. http://archive.ics.uci.edu/ml/ www.kdnuggets.com/datasets/
e.	Perform Association technique on Weather dataset.
10.	**Web Mining**
a.	Implement Hyperlink Induced Topic Search (HITS) algorithm
b.	Implement PageRank Algorithm

Evaluation scheme

I Continuous Assessment (C.A.)– 40 Marks

1) Assessment 1 (20 Marks)

2) Assessment 2 (20 Marks**)**

II Semester End Examination (SEE) - 60 Marks

QUESTION PAPER PATTERN OF (SEE)

Maximum Marks: 60 Marks

Time: 2 Hours

Note: 1) Attempt all Questions

2) Any three sub questions to be attempted. All sub questions carry equal marks

Question No	Particulars	Marks
Q -1 (Unit I)	A) B) C) D)	12 Marks
Q -2 (Unit II)	A) B) C) D)	12 Marks
Q -3 (Unit III)	A) B) C) D)	12 Marks
Q4 (Unit IV)	A) B) C) D)	12 marks
Q5 (Unit V)	A) B) C) D)	12 marks

Practical Exam: 50 marks

Certified copy journal is essential to appear for the practical examination.

1.	**Practical Question**	**40**
2.	**Journal**	**5**
3.	**Viva Voce**	**5**

Linear Algebra and Discrete Mathematics

B. Sc. (Data Science)		**Semester – III**	
Course Name: Linear Algebra and Discrete Mathematics		**Course Code: 60309**	
Periods per week (1 Period is 50 minutes)		**5**	
Credits		**2**	
		Hours	**Marks**
Evaluation System	**Theory Examination**	2	60
	Internal	--	40

Course Objectives

1. To analyze the solution set of a system of linear equations.
2. To interpret existence and uniqueness of solutions geometrically.
3. To formulate, solve, apply, and interpret properties of linear systems.

Unit	Details	Lectures
I	**Matrices and Gaussian Elimination:** Introduction, The Geometry of Linear Equations, An Example of Gaussian Elimination, Matrix Notation and Matrix Multiplication, Triangular Factors and Row Exchanges, Inverses and Transposes, Special Matrices and Applications **Vector Spaces:** Vector Spaces and Subspaces, Solving $Ax= 0$ and $Ax= b$, Linear Independence, Basis, and Dimension, The Four Fundamental Subspaces, Graphs and Networks, Linear Transformations	**12**
II	**Orthogonality:** Orthogonal Vectors and Subspaces, Cosines and Projections onto Lines, Projections and Least Squares, Orthogonal Bases and Gram-Schmidt **Determinants:** Introduction, Properties of the Determinant, Formulas for the Determinant, Applications of Determinants	**12**
III	**Eigenvalues and Eigenvectors:** Introduction, Diagonalization of a Matrix, Difference Equations and Powers Ak, Differential Equations and eAt, Complex Matrices, Similarity Transformations	**12**
IV	**Positive Definite Matrices:** Minima, Maxima, and Saddle Points, Tests for Positive Definiteness, Singular Value Decomposition, Minimum Principles, The Finite Element Method. **Computations with Matrices:** Introduction, Matrix Norm and Condition Number, Computation of Eigenvalues, Iterative Methods for $Ax= b$	**12**

V	**Linear Programming and Game Theory:** Linear Inequalities, The Simplex Method, The Dual Problem, Network Models, Game Theory	**12**

Books and References:					
Sr. No.	**Title**	**Author/s**	**Publisher**	**Edition**	**Year**
1	Linear Algebra and Its Applications	Gilbert Strang	**Cengage Publication**	Fourth Edition	**----**
2.	Advanced Linear Algebra	David Surowski			
3.	Linear Algebra, Theory and Applications	Kenneth Kuttlet			

Course Outcome:

CO 1: Learner is able to perform common matrix operations such as addition, scalar multiplication, multiplication, and transposition.

CO 2: Learner is able to describe how the determinant of a product of matrices relates to the determinant of the individual matrices.

CO 3: Learner expresses clear understanding of the concept of a solution to a game' and also the limitations on the applicability of the theory.

Linear Algebra and Discrete Mathematics Practical

B. Sc. (Data Science)		**Semester – III**	
Course Name: Linear Algebra and Discrete Mathematics Practical		**Course Code: 60310**	
Periods per week (1 Period is 50 minutes)		**3**	
Credits		**2**	
		Hours	**Marks**
Evaluation System	**Practical Examination**	**2½**	**50**
	Internal	**--**	

List of Practical:	
1	**Matrices and Gaussian Elimination**
a	Multiplication and transpose of matrix using R/python/scilab/matlab.
b	Inverses of matrix inR/python/scilab/matlab without using any inbuilt package.
c	Inverses of matrix inR/python/scilab/matlab using any inbuilt package like numpy.
d	Linear equation with n unknowns using Gauss Elimination Method using R/python/scilab/matlab.
2	**Vector**
a	Addition, subtraction, multiplication anddivision of vector using R/python/scilab/matlab.
b	dot product & cross product of vector using R/python/scilab/matlab
c	Visualising vector Linear Transformations using R/python/scilab/matlab
3	**Orthogonality**
a	Computes the orthonormal vectors using the GS algorithmusing R/python/scilab/matlab.
b	Projections and Least Squaresusing R/python/scilab/matlab.
c	Fast Fourier Transform using R/python/scilab/matlab.

d.	Finding determinant of matrix inR/python/scilab/matlab without using any inbuilt package.
4	**Eigen Value & Eigen Vector**
a.	Compute the eigenvalues and right eigenvectors of a given square array using R/python/scilab/matlab.
b	Program to test diagonalizable matrix using R/python/scilab/matlab.
5	**Positive Definite matrices**
a	Tests for Positive Definiteness using R/python/scilab/matlab.
b	Singular Value Decomposition using R/python/scilab/matlab.

SEMESTER-IV

x

Testing of Hypothesis

B. Sc. (Data Science)		Semester – IV	
Course Name: Testing of Hypothesis		Course Code: 60401	
Periods per week (1 Period is 50 minutes)		5	
Credits		2	
		Hours	Marks
Evaluation System	Theory Examination	2	60
	Internal	--	40

Course Objectives:

1. To impart statistical significance in solving complex problems.
2. To critically test in developing robust, extensible and highly maintainable solutions to simple and complex problems.
3. To implement various statistical functions using suitable programming languages and packages.
4. To scientifically test the unknown and unlock possibilities in different dimensions of the system.
5. To write the reports of analytical results generated by the system.

Unit	Details	Lectures
I	**Introduction to Hypothesis Testing:** Hypothesis Tests, Stating a Hypothesis, Types of Errors and Level of Significance, Statistical Tests and P-Values, Strategies for Hypothesis Testing, Characteristics of a good hypothesis. **Hypothesis Testing for the Mean (σ Known):** Using P-Values to Make Decisions, Using P-Values for a z-Test, Using Rejection Regions for a z-Test The t-Test for a Mean μ, Using P-Values with t-Tests,	15
II	**Goodness of fit tests :-** Chi-square test, Kolmogorov-Smirnov, Shapiro-Wilk. **Variance tests:** Chi-square test of a single variance, F-tests of two variances, Tests of homogeneity **Wilcoxon rank-sum/Mann-Whitney U test, Sign test** **Contingency tables:** Chi-square contingency table test, G contingency table test, Fisher's exact test, Measures of association.	15
III	**Analysis of variance and covariance :-** ANOVA**,** Single factor or one way ANOVA, Two factor or two-way ANOVA . **Non-Parametric ANOVA:** Kruskal-Wallis ANOVA, Friedman ANOVA test.	15
IV	**Regression and smoothing :-** Least squares, Ridge regression, Simple and multiple linear regression, Polynomial regression, Generalized Linear Models (GLIM), Logistic regression for proportion data, Poisson regression for count data, Non-linear regression, Smoothing and Generalized Additive Models (GAM). **Time series analysis and temporal autoregression:** Moving averages, Trend Analysis, ARMA and ARIMA (Box-Jenkins) models.	15

x

Books and References:					
Sr. No.	**Title**	**Author/s**	**Publisher**	**Edition**	**Year**
1	Hypothesis Testing	---	Pearson Higher Education	---	---
2	Statistical Analysis Handbook	Dr. Michael J de Smith	The Winchelsea Press, Drumlin Security Ltd,	2018 Ed	2018
3	An Introduction to Statistical Methods and Data Analysis	R. Lyman Ott& Michael Longnecker	Thomson Learning	---	---

Course Outcome:

CO 1: Learner is developing null and alternative hypotheses to test for a given situation.
CO 2: Learner is able to differentiate one- and two-tailed hypothesis tests.
CO 3: Learner is able to do sampling a normal distribution and random sampling.
CO 4: Learner is using statistical models and their associations in performing hypothesis testing.
CO 5: Lerner is writing the reports and interpreting the data using the various programming languages and packages.

Testing of Hypothesis Practical

B. Sc. (Data Science)		**Semester – IV**	
Course Name: Testing of Hypothesis Practical		**Course Code: 60402**	
Periods per week (1 Period is 50 minutes)		**3**	
Credits		**2**	
		Hours	**Marks**
Evaluation System	**Practical Examination**	**2½**	**50**
	Internal	**--**	

List of Practical:	
Practical can be performed using R / Python / scilab / matlab / SPSS / MS Excel	
1	**Hypothesis Testing for the Mean**

a	Perform testing of hypothesis using one sample t-test.
b	Perform testing of hypothesis using two sample t-test.
c	Perform testing of hypothesis using paired t-test.
d	Perform testing of hypothesis using Z-test.
2	**Goodness-of-fit test**
a	Perform goodness-of-fit test using chi-squared test.
b	Perform goodness-of-fit test using KS-test.
c	Perform testing of hypothesis using chi-squared Test of Independence
3	**Variance Testing**
a.	Using Chi-square test of a single variance
b.	Using F-tests of two variances
c.	Testing of homogeneity
4	**Analysis of variance and covariance**
a.	Perform testing of hypothesis using one-way ANOVA.
b.	Perform testing of hypothesis using two-way ANOVA.
5	**Regression & Time series**
a.	Perform simple linear regression
b.	Perform multiple linear regression
c.	Perform polynomial regression
d.	Perform time series analysis using Moving averages
e.	Perform time series analysis using Trend Analysis

Evaluation scheme

I Continuous Internal Assessment (C.A.)– 40 Marks

1) Assessment 1 (20 Marks)
2) Assessment 2 (20 Marks**)**

II Semester End Examination (SEE) - 60 Marks

QUESTION PAPER PATTERN OF (SEE)

Maximum Marks: 60 Marks

Time: 2 Hours

Note: 1)Attempt all Questions

2) Any three sub questions to be attempted. All sub questions carry equal marks

Question No	Particulars	Marks
Q -1 (Unit I)	A) B) C) D)	15 Marks
Q -2 (Unit II)	A) B) C) D)	15 Marks
Q -3 (Unit III)	A) B) C) D)	15 Marks
Q4 (Unit IV)	A) B) C) D)	15 marks

Practical Exam: 50 marks

Certified copy journal is essential to appear for the practical examination.

1.	**Practical Question**	**40**
2.	**Journal**	**5**
3.	**Viva Voce**	**5**

x

Big Data

B. Sc. (Data Science)		Semester – IV	
Course Name: Big Data		**Course Code: 60403**	
Periods per week (1 Period is 50 minutes)		**5**	
Credits		**2**	
		Hours	**Marks**
Evaluation System	**Theory Examination**	2	60
	Internal	--	40

Course Objectives

1. To develop core abilities to make data-driven decisions through big data.
2. To provide an overview of an exciting growing field of big data analytics.
3. To introduce the tools required to manage and analyze big data like Hadoop, NoSql MapReduce.
4. To teach the fundamental techniques and principles in achieving big data analytics with scalability and streaming capability.

Unit	Details	Lectures
I	**Introduction to Big Data:** What is Big Data? , Evolution of Big Data, Different sources of Big Data, Important characteristics of Big Data (5V's , Volume, Variety, Velocity, Varacity, Volue) Challenges of Big Data. Introduction to different Big Data architectures: Hadoop, Spark, Databricks **Detail Architecture of Hadoop:** Components of Hadoop Ecosystem, HDFS architecture, HDFS features Name node and data node, Secondary name node Job tracker Task tracker Client nodes **Explain master-slaves ,** How does a file read and write work? Local file system and HDFS Checkpointing in Hadoop, HDFS commands , HDFS web interface	12 Lectures
II	**Defining Big Data analytics:** Discovering value from large data sets, Exploiting data to optimize decision-making Planning your analytics life cycle project: Outlining steps in the life cycle, Contrasting traditional analytics with Big Data analytics **Preparing the data**: Loading data for knowledge discovery, Spotting outliers in the data, Transforming and summarizing data Visualizing data characteristics: Revealing changes over time, Displaying proportions within your data,	12 Lectures

	Leveraging charts to display relationships,Displaying relationships across categories	
III	**Planning and Implementing a Complete Data Analytics Solution** Transforming business objectives to analytic projects · Arguing your business case for analytics, Mapping analytics models to business objectives, Identifying performance metrics targets **Implementing the analytics life cycle** · Finding core data sets, Preparing the data for analysis, Modeling the data, Executing the model, Communicating results **Ensuring a Successful Data Analytics Solution** · Identifying barriers to Big Data analytics, Managing and mitigating risks, Employing an implementation checklist	12 Lectures
IV	**Leveraging Analytics with RHadoop** Map-Reduce programming paradigm, How Map-Reduce works? Creating and executing Hadoop Map Reduce jobs, Integrating R and Hadoop with RHadoop, Examining the components of RHadoop, Creating modules for RHadoop jobs, Executing RHadoop jobs, **Monitoring job execution flow.** Word-count example using RHadoop, Building a Recommendation Framework: Streamlining business decisions · Considering motivations for a recommender engine, Leveraging recommendations based on collaborative filtering, Exploring the architecture of the recommendation framework, Building programming components, Executing the recommendation model, Performing tradeoff analysis	12 Lectures
V	**Mining Unstructured Data** Investigating business value within unstructured data · Making a business case for unstructured data mining, Extending business intelligence with mining tools Implementing text mining and social network analysis · Analyzing the structure of text mining, Evaluating mining approaches, Building a text mining framework, Inspecting social network interactions	12 Lectures

Reference Books:

1. Vignesh Prajapati, "Big Data Analytics with R and Hadoop", Packt Publishing House.
2. Bart Baesens , Analytics in a Big Data World: The Essential Guide to Data Science and its Applications (Wiley and SAS Business Series), Wiley
3. Arvind Sathi, Big Data Analytics Disruptive Technologies For Changing The Game, MC Press LLC
4. Viktor Mayer Schonberger, Kenneth Cukier, Big Data,
5. Adam Jorgensen, James Rowland-Jones, John Welch and Dan Clark, "Microsoft Big Data Solutions", Wiley
6. http://www.bigdatauniversity.com

x

Course Outcome:

CO 1: Learner understands the key issues in big data management and its associated applications in intelligent business and scientific computing.

CO 2: Lerner is acquiring fundamental techniques and algorithms like Hadoop, Map Reduce and NO SQL in big data analytics.

CO 3: Learner is able to interpret business models and scientific computing paradigms, and apply software tools for big data analytics.

CO 4: Learner understands adequate perspectives of big data analytics in various applications like recommender systems, social media applications etc.

Big Data Practical

B. Sc. (Data Science)		**Semester – IV**	
Course Name: Big Data Practical		**Course Code: 60404**	
Periods per week (1 Period is 50 minutes)		**3**	
Credits		**2**	
		Hours	**Marks**
Evaluation System	**Practical Examination**	**2½**	**50**
	Internal	**--**	

List of Practical:	
1 a	Install, configure and run Hadoop and HDFS
b	Implement word count/ frequency program using MapReduce
2	Implement a Map reduce program that process a weather dataset
3	Exploring Hadoop Distributed File System (HDFS)
4	Implement an application that store big data in Hbase/ Mongodb/ Pig using Hadoop/R
5	Implement a program in Pig
6	Configure the Hive and implement the application in Hive

Evaluation scheme

I Continuous Assessment (C.A.)– 40 Marks

1) Assessment 1 (20 Marks)

2) Assessment 2 (20 Marks**)**

II Semester End Examination (SEE) - 60 Marks

QUESTION PAPER PATTERN OF (SEE)

Maximum Marks: 60 Marks

Time: 2 Hours

Note: 1) Attempt all Questions

2) Any three sub questions to be attempted. All sub questions carry equal marks

Question No	Particulars	Marks
Q -1 (Unit I)	A) B) C) D)	12 Marks
Q -2 (Unit II)	A) B) C) D)	12 Marks
Q -3 (Unit III)	A) B) C) D)	12 Marks
Q4 (Unit IV)	A) B) C) D)	12 marks
Q5 (Unit V)	A) B) C) D)	12 marks

Practical Exam: 50 marks

Certified copy journal is essential to appear for the practical examination.

1.	**Practical Question**	**40**
2.	**Journal**	**5**
3.	**Viva Voce**	**5**

x

Artificial Intelligence

B. Sc. (Data Science)		Semester – IV	
Course Name: Artificial Intelligence		Course Code: 60405	
Periods per week (1 Period is 50 minutes)		5	
Credits		2	
		Hours	Marks
Evaluation System	Theory Examination	2	60
	Internal	--	40

Course Objectives :-

1. To introduce and appreciate use of AI and the theory underlying for solving problems.
2. To Learn about Rational Intelligent Agent and Agent types to solve problems
3. To learn about representing difficult real life problems as state space representation and solving them using AI techniques.
4. To understand the basic issues of knowledge representation and develop skills for reasoning and handling uncertainty
5. To introduce advanced topics of AI for solving complex problems

Unit	Details	Lectures
I	**Intelligent Systems and Intelligent Agents:** Introduction to AI, AI Problems and AI techniques, Solving problems by searching, Problem Formulation. State Space Representation Structure of Intelligent agents, Types of Agents, Agent Environments PEAS representation for an Agent.	12
II	**Searching Techniques:** Uninformed Search: DFS, BFS, Uniform cost search, Depth Limited Search, Iterative Deepening. Informed Search: Heuristic functions, Hill Climbing, Simulated Annealing, Best First Search, A* algorithm Constraint Satisfaction Programming: Crypto Arithmetic, Map Coloring, N-Queens. **Adversarial Search:** Game Playing, Min-Max Search, Alpha Beta Pruning	12
III	**Knowledge and Reasoning:** Knowledge Based Agent, Overview of Propositional Logic, First Order Predicate Logic, Inference in First Order Predicate Logic: Forward and Backward Chaining, Resolution.	12
IV	**Uncertainity and Reasoning:** Uncertainly, Representing Knowledge in an Uncertain Domain, Bayesian Network, Conditional Probability, Joint	12

	Probability, Bayes‘ theorem, Belief Networks, Simple Inference in Belief Networks. Sequential decision problems.	
V	**Machine Learning:** Forms of Learning, Supervised Learning, Learning Decision Trees, Evaluating and Choosing the Best Hypothesis, Theory of Learning, Regression and Classification with Linear Models, Artificial Neural Networks, Nonparametric Models, Support Vector Machines, Ensemble Learning, Statistical Learning, Introduction to deep learning concepts	**12**

Books and References:

Sr. No.	Title	Author/s	Publisher	Edition	Year
1.	Artificial Intelligence: A Modern Approach	Stuart J. Russell and PeterNorvig	Pearson	Fourth Edition	2020
2.	Artificial Intelligence: Foundations of Computational Agents	David L Poole, Alan K. Mackworth	Cambridge University Press	Second Edition	2017
3.	Artificial Intelligence	Kevin Knight and Elaine Rich	McGraw Hill	3rd Edition	2017
4.	The Elements of Statistical Learning	Trevor Hastie, Robert Tibshirani and Jerome Friedman	Springer		2013
5.	A First Course in Artificial Intelligence	Deepak Khemani	TMH	1st Edition	2017

Course Outcome:

CO 1: Leaner understands building blocks of AI.

CO 2: Learner is analyzing problem and solving it by implementing suitable techniques.

CO 3: Learner is applying logics based techniques to solve examples.

CO 4: Learner is able to implement Bayesian approaches.

CO 5: Learner is using machine learning concepts for solving problems

Artificial Intelligence Practical

B. Sc. (Data Science)		Semester – IV	
Course Name: Artificial Intelligence Practical		**Course Code: 60406**	
Periods per week (1 Period is 50 minutes)		**3**	
Credits		**2**	
		Hours	**Marks**
Evaluation System	**Practical Examination**	**2½**	**50**
	Internal	**--**	

List of Practical:	
1.	**Write the program to compute the following Uninformed Search.**
a.	Depth First Search
b.	Breadth First Search
2.	**Write the program to compute the following Informed Search.**
a.	Hill Climbing
b.	Simulated Annealing
c.	A* Algorithm
d.	simulation of tic – tac – toe game using Min-Max Search
e.	Alpha Beta Pruning
3.	**Write the program to compute the following Algorithms for suitable problem**
a.	Simulate solution for 4-Queen / N-Queen problem
b.	Constraint satisfaction problem: Map Coloring
c.	Missionaries and Cannibals
d.	Simple Inferencing
4.	**Write the program to implement decision tree for suitable problem.**

a.	Two Class decision
b.	Multi class decision
5.	**Write the program to compute the following Algorithms for suitable problem**
a.	Linear Regression
b.	Classification Problems

Evaluation scheme

I Continuous Assessment (C.A.)– 40 Marks

1) Assessment 1 (20 Marks)

2) Assessment 2 (20 Marks**)**

II Semester End Examination (SEE) - 60 Marks

QUESTION PAPER PATTERN OF (SEE)

Maximum Marks: 60 Marks

Time: 2 Hours

Note: 1) Attempt all Questions

2) Any three sub questions to be attempted. All sub questions carry equal marks

Question No	Particulars	Marks
Q -1 (Unit I)	A) B) C) D)	12 Marks
Q -2 (Unit II)	A) B) C) D)	12 Marks
Q -3 (Unit III)	A) B) C) D)	12 Marks
Q4 (Unit IV)	A) B) C) D)	12 marks
Q5 (Unit V)	A) B) C) D)	12 marks

Practical Exam: 50 marks

Certified copy journal is essential to appear for the practical examination.

1.	**Practical Question**	**40**
2.	**Journal**	**5**
3.	**Viva Voce**	**5**

Operating System

B. Sc. (Data Science)		Semester – IV	
Course Name: Operating System		**Course Code: 60407**	
Periods per week (1 Period is 50 minutes)		**5**	
Credits		**2**	
		Hours	**Marks**
Evaluation System	**Theory Examination**	2	60
	Internal	--	40

Course Objectives:

1. To introduce basic concepts and functions of modern operating systems.
2. To understand the concept of process, thread management and scheduling.
3. To learn the concept of concurrency control.
4. To study various Memory Management techniques.
5. To know the concept of I/O and File management

Unit	Details	Lectures
I	**OVERVIEW OFOPERATING SYSTEM:** Operating System Objectives and Functions, The Evolution of Operating Systems, Developments Leading to Modern Operating Systems, Virtual Machines, Introduction to Linux OS, BASH Shell scripting: Basic shell commands.	12
II	**PROCESS MANAGEMENT:** Process: Concept of a Process, Process States, Process Description, Process Control Threads: Processes and Threads, Concept of Multithreading, Types of Threads, Thread programming Using Pthreads. Scheduling: Types of Scheduling, Scheduling Algorithms, First Come First Served, Shortest Job First, Priority, Round Robin	12
III	**CONCURRENCY CONTROL:** Process/thread Synchronization and Mutual Exclusion: Principles of Concurrency, Requirements for Mutual Exclusion, Mutual Exclusion: Operating System Support (Semaphores and Mutex). Classical synchronization problems: Readers/Writers Problem, Producer and	12

	Consumer problem, Inter-process communication (Pipes, Shared Memory). Deadlock: Principles of Deadlock, Deadlock Modeling, and Strategies to deal with deadlock: Prevention, Avoidance, Detection and Recovery. Example: Dining Philosophers Problem / Banker's Algorithm.	
IV	**MEMORY MANAGEMENT:** Memory Management: Memory Management Requirements, Memory Partitioning: Fixed Partitioning, Dynamic Partitioning, Buddy System, Relocation, Paging, Page table structure, Segmentation Virtual Memory: Background, Demand Paging, Page Replacement (FIFO, LRU, Optimal), Allocation of frames, Thrashing	**12**
V	**INPUT/OUTPUT AND FILE MANAGEMENT:** I/O Management and Disk Scheduling: I/O Devices, Organization of the I/O Function, I/O Buffering, Disk Scheduling (FIFO, SSTF, SCAN, C-SCAN, LOOK, C-LOOK). File Management: Overview-Files and File Systems, File structure. File Organization and Access, File Directories, File Sharing, Record Blocking, Secondary Storage Management.	**12**

Books and References:

Sr. No.	Title	Author/s	Publisher	Edition	Year
1	Operating System: Internals and Design Principles	William Stallings	Prentice Hall	Eighth Edition	2014
2	Operating System Concepts	Abraham Silberschatz, Peter Baer Galvin, Greg Gagne	John Wiley & Sons	Ninth Edition	2012
3	Linux System Administration	Tom Adelstein and Bill Lubanovic	O'Reilly Media		
4	Operating Systems	Harvey M. Deitel	Prentice Hall		
5	Advanced Shell Scripting	Mendel Cooper	Linux Documentation Project		

Course Outcome:

CO1: Explain the role of Modern Operating Systems.
CO2: Apply the concepts of process and thread scheduling.
CO3: Illustrate the concept of process synchronization, mutual exclusion and the deadlock.
CO4: Implement the concepts of various memory management techniques.

CO5: Make use of concept of I/O management and File system.

Operating System Practical

B. Sc. (Data Science)		**Semester – IV**	
Course Name: Operating System Practical		**Course Code: 60408**	
Periods per week (1 Period is 50 minutes)		**3**	
Credits		**2**	
		Hours	**Marks**
Evaluation System	**Practical Examination**	**2½**	**50**
	Internal	**--**	

List of Practical:-	
1.	Study of Basic Linux Commands: echo, ls, read, cat, touch, test, loops, arithmetic comparison, conditional loops, grep, sed etc.
2.	Write a program to implement an address book with options given below: a) Create address book. b) View address book. c) Insert a record. d) Delete a record. e) Modify a record.
3	Implement the C program for CPU Scheduling Algorithms: Shortest Job First (Preemptive) and Round Robin with different arrival time.
4.	Implement the C program for Page Replacement Algorithms: FCFS, LRU, and Optimal for frame size as minimum three.
5.	Implement the C program for Disk Scheduling Algorithms: SSTF, SCAN, C-Look considering the initial head position moving away from the spindle.

Evaluation scheme

I Continuous Assessment (C.A.)– 40 Marks

1) Assessment 1 (20 Marks)

2) Assessment 2 (20 Marks**)**

II Semester End Examination (SEE) - 60 Marks

QUESTION PAPER PATTERN OF (SEE)

Maximum Marks: 60 Marks

Time: 2 Hours

Note: 1) Attempt all Questions

2) Any three sub questions to be attempted. All sub questions carry equal marks

Question No	Particulars	Marks
Q -1 (Unit I)	A) B) C) D)	12 Marks
Q -2 (Unit II)	A) B) C) D)	12 Marks
Q -3 (Unit III)	A) B) C) D)	12 Marks
Q4 (Unit IV)	A) B) C) D)	12 marks
Q5 (Unit V)	A) B) C) D)	12 marks

Practical Exam: 50 marks

Certified copy journal is essential to appear for the practical examination.

1.	**Practical Question**	**40**
2.	**Journal**	**5**
3.	**Viva Voce**	**5**

Numerical Methods

B. Sc. (Data Science)		Semester – IV	
Course Name: Numerical Methods		Course Code: 60409	
Periods per week (1 Period is 50 minutes)		5	
Credits		2	
		Hours	Marks
Evaluation System	Theory Examination	2	60
	Internal	--	40

Course Objectives

1. To be able to precisely solve problems using mathematical modeling.
2. To be able to find solutions for unsolvable problems.
3. To find an answer or solution close to answer, without even knowing what the answer is.

Unit	Details	Lectures
I	**Background and Introduction:** Differential Equations, Matrix Analysis, Matrix Eigenvalue Problem, Errors and Approximations, Iterative Methods **Numerical Solution of Equations of a Single Variable:-** Numerical Solution of Equations, Bisection Method, Regular Falsi Method, Fixed Point Method, Newton's Method, Secant Method, Equations with Several Roots	12
II	**Numerical Solution of Systems of Equations:** Linear Systems of Equations, Numerical Solution of Linear Systems, Gauss Elimination Method, LU Factorization Methods, Iterative Solution of Linear Systems, Ill-Conditioning and Error Analysis	12
III	**Curve Fitting and Interpolation:-** Least-Squares Regression, Linear Regression, Linearization of Nonlinear Data, Polynomial Regression, Polynomial Interpolation, Spline Interpolation, Fourier Approximation and Interpolation **Numerical Differentiation and Integration:** Numerical Differentiation, Finite-Difference Formulas for Numerical Differentiation, Numerical Integration: Newton–Cotes Formulas, Numerical Integration of Analytical Functions: Romberg Integration, Gaussian Quadrature, Improper Integrals	12
IV	**Numerical Solution of Initial-Value Problems:** Introduction, One-Step Methods, Euler's Method, Runge–Kutta Methods, Multistep Methods, Systems of Ordinary Differential Equations, Stability, Stiff Differential Equations	12

V	**Numerical Solution of Boundary-Value Problems:** Second-Order BVP, Boundary Conditions, Higher-Order BVP, Shooting Method,Finite-Difference Method **Matrix Eigenvalue Problem:-** Matrix Eigenvalue Problem, Power Method: Estimation of the Dominant Eigenvalue, Inverse Power Method: Estimation of the Smallest Eigenvalue, Shifted Inverse Power Method: Estimation of the Eigen value Nearest a Specified Value, Shifted Power Method, Deflation Methods, Householder Tridiagonalization and QR Factorization Methods, A Note on the Terminating Condition Used in Householder QR, Transformation to Hessenberg Form (Nonsymmetric Matrices)	**12**

Books and References:

Sr. No.	Title	Author/s	Publisher	Edition	Year
1. 1	**Numerical Methods for Engineers and Scientists Using MATLAB**	**Ramin S. Esfandiari**	CRC Press	**Second Edition**	2017

Course Outcome:

CO 1: Learner implementing Numerical Methods to solve the problems.

CO 2: Learner is computing the numerical results using raw data.

CO 3: Learner will learn numerical different and integration.

CO 4: Learner will learn Numerical Solution of Initial-Value

CO 5: Learner will learn Matrix Eigenvalue

Numerical Methods Practical

B. Sc. (Data Science)		Semester – IV	
Course Name:- Numerical Methods Practical		Course Code: 60410	
Periods per week (1 Period is 50 minutes)		3	
Credits		2	
		Hours	Marks
Evaluation System	Practical Examination	2½	50
	Internal	--	

List of Practical:	
1	**Write a program using R/Python/Scilab/Matlab**
a	Bisection Method
b	Regula Falsi Method
c	Newton‘s Method
2	**Write a program using R/Python/Scilab/Matlab**
a	Gauss Elimination Method
b	LU Factorization Methods
3	**Write a program using R/Python/Scilab/Matlab**
a	Numerical Differentiation
b	Newton–Cotes Formulas
c	Romberg Integration
d	Gaussian Quadrature
4	**Write a program using R/Python/Scilab/Matlab**
a	Euler‘s Method
b	Runge–Kutta Methods
5	**Write a program using R/Python/Scilab/Matlab**

a	Second-Order BVP
b	Higher-Order BVP
c.	Finite difference method
d.	Estimation of the Eigenvalue Nearest a Specified Value

Evaluation scheme

I Continuous Assessment (C.A.)– 40 Marks

1) Assessment 1 (20 Marks)

2) Assessment 2 (20 Marks**)**

II Semester End Examination (SEE) - 60 Marks

QUESTION PAPER PATTERN OF (SEE)

Maximum Marks: 60 Marks

Time: 2 Hours

Note: 1) Attempt all Questions

2) Any three sub questions to be attempted. All sub questions carry equal marks

Question No	Particulars	Marks
Q -1 (Unit I)	A) B) C) D)	12 Marks
Q -2 (Unit II)	A) B) C) D)	12 Marks
Q -3 (Unit III)	A) B) C) D)	12 Marks
Q4 (Unit IV)	A) B) C) D)	12 marks
Q5 (Unit V)	A) B) C) D)	12 marks

Practical Exam: 50 marks

Certified copy journal is essential to appear for the practical examination.

1.	**Practical Question**	**40**
2.	**Journal**	**5**
3.	**Viva Voce**	**5**

x