ANDHRA PRADESH STATE COUNCIL OF HIGHER EDUCATION

Model Syllabus for 4-Year UG Honours in B.Sc. (Data Science) as Major in consonance with Curriculum framework w.e.f. AY 2025-26

Prepared by Adikavi Nannaya University, Rajahmundry

COURSE STRUCTURE (for Semester I to VI)

Year	Semester	Course	Title of the Course	No. of Hrs /Week	No. of Credits
I	I	1	Computer Fundamentals and Office Automation	3	3
			Computer Fundamentals and Office Automation Lab	2	1
		2	Problem Solving Using C	3	3
			Problem Solving Using C Lab	2	1
	II	3	Python Programming and Data Structures	3	3
			Python programming and Data Structures lab	2	1
		4	Statistical Foundations for Data Science	3	3
			Statistical Foundations for Data Science lab	2	1
II	III	5	Database Management Systems	3	3
			Database Management Systems Lab	2	1
		6	Data Science with R	3	3
			Data Science With R lab	2	1
		7	Web Technologies	3	3
			Web Technologies Lab	2	1
	IV	8	Data Mining	3	3
			Data Mining Lab	2	1
		9	Python for Data Analysis and Visualization	3	3
			Python for Data Analysis and Visualization lab	2	1
		10	Document Oriented Database	3	3
			Document oriented Database lab	2	1
III	V	11	Business Intelligence Tools	3	3
			Business Intelligence Tools Lab	2	1

Year	Semester	Course	Title of the Course	No. of Hrs /Week	No. of Credits
		12 A	Machine Learning	**3**	**3**
			Machine Learning Lab	**2**	**1**
		OR			
		12 B	Big Data Technologies	**3**	**3**
			Big Data Technologies Lab	**2**	**1**
		13 A	Artificial Intelligence	**3**	**3**
			Artificial Intelligence Lab	**2**	**1**
		OR			
		13 B	Cloud computing for Data Science	**3**	**3**
			Cloud computing for Data Science Lab	**2**	**1**
	VI	**14 A**	Neural networks and Deep Learning	**3**	**3**
			Neural networks and Deep Learning lab	**2**	**1**
		OR			
		14 B	Time Series Analysis and Forecasting	**3**	**3**
			Time Series Analysis and Forecasting Lab	**2**	**1**
		15 A	Natural Language Processing	**3**	**3**
			Natural Language Processing Lab	**2**	**1**
		OR			
		15 B	Data Engineering & MLOps	**3**	**3**
			Data Engineering & MLOps	**2**	**1**

Note: In the III Year (during the V and VI Semesters), students are required to select a pair of electives from one of the **TWO** specified domains. **For example: if set 'A' is chosen, courses 12 to 15 to be chosen as 12 A, 13 A, 14 A and 15 A.** To ensure in-depth understanding and skill development in the chosen domain, students must continue with the same domain electives in both the V and VI Semesters.

SEMESTER-I

COURSE 1: COMPUTER FUNDAMENTALS AND OFFICE AUTOMATION

Theory **Credits: 3** **3 hrs/week**

Course Objectives

1. **Understand foundational computing concepts**, including number systems, the evolution of computers, block diagrams, and generational progress.
2. **Develop knowledge of computer architecture**, focusing on system organization and networking fundamentals.
3. **Acquire practical skills in document creation**, formatting, and digital presentations using word processing tools.
4. **Gain proficiency in spreadsheet operations**, such as data entry, formulas, functions, and charting techniques.
5. **Introduce data visualization and basic modelling principles**, fostering analytical thinking in structuring and interpreting data sets.

Course Outcomes

1. At the End of the Course, The Students will be able to **explain different number systems**, the historical evolution of computers, and identify key components in a block diagram.
2. Learners will demonstrate **basic blocks of a computer and fundamental networking knowledge.**
3. Learners will create professional-level documents and **design visually appealing presentations** using word processing software and presentation software.
4. Learners will manipulate data within spreadsheets, apply formulas, and **generate accurate summaries and visualizations.**
5. Learners will apply data modelling techniques to **analyze, organize, and represent data effectively** in various scenarios.

Unit 1. Number Systems, Evolution , Block Diagram and Generations:

Number Systems: Binary, Decimal, Octal, Hexadecimal; conversions between number systems.

Evolution of Computers: History from early mechanical devices to modern-day systems.

Block Diagram of a Computer: Components like Input Unit, Output Unit, Memory, CPU (ALU + CU).

Generations of Computers: First to Fifth Generation – technologies, characteristics, examples.

Unit 2. Basic organization and N/W fundamentals:

Computer Organization: Functional components – Input/Output devices, Storage types, Memory hierarchy.

Types of Computers: Micro, Mini, Mainframe, and Supercomputers.

Networking Fundamentals: Definition, need for networks, types (LAN, WAN, MAN), topology (Star, Ring, Bus).

Internet Basics: IP Address, Domain Name, Web Browser, Email, WWW.

Unit 3. Word Processing and presentations:

Word Processing Basics: Using MS Word/Google Docs – formatting, styles, tables, mail merge.

Presentation Tools: Using PowerPoint/Google Slides – slide design, animations, transitions.

Applications: Creating resumes, reports, brochures, and presentations.

Keyboard Shortcuts

Unit 4. Spreadsheet Basics:

Spreadsheet Concepts: Understanding rows, columns, cells in tools like MS Excel/Google Sheets, cell referencing.

Functions and Formulae: SUM, AVERAGE, IF, COUNT.

Charts and Graphs: Creating visual representations

Data Handling: Sorting, filtering, conditional formatting.

Text Functions: LEFT, RIGHT, MID, LEN, TRIM, CONCAT, TEXTJOIN

Advanced Functions: **Logical**: IF, AND, OR, IFERROR, **Lookup**: VLOOKUP, HLOOKUP, XLOOKUP, INDEX, MATCH

Unit 5. Data Analysis and Visualization:

Conditional Formatting: Custom rules, Color scales, Icon sets, Data bars

Data Analysis Tools: Pivot Tables and Pivot Charts, Data Validation (Drop-downs, Input Messages, Error Alerts), What-If Analysis: Goal Seek, Scenario Manager, Data Tables

Charts and Dashboards: Creating Interactive Dashboards, Using slicers with Pivot Tables,Combo Charts and Sparklines

Productivity Tips: Using Named Ranges, Freeze Panes, Split View

Textbooks:

1. **Fundamentals of Computers,** Reema Thareja, Oxford University Press, Second Edition
2. **Fundamentals of Computers**, V. Rajaraman – PHI Learning
3. **Introduction to Computers** by Peter Norton – McGraw Hill
4. **Microsoft Office 365 In Practice** by Randy Nordell – McGraw Hill Education

References:

1. **Excel 2021 Bible** by Michael Alexander, Richard Kusleika – Wiley
2. **Networking All-in-One For Dummies** by Doug Lowe – Wiley
3. **Microsoft Official Docs and Training**: https://learn.microsoft.com
4. **Google Workspace Learning Center**: https://support.google.com/a/users/

Activities:

Outcome: At the End of the Course, The Students will be able to **explain different number systems**, the historical evolution of computers, and identify key components in a block diagram.

Activity: Create a digital poster or infographic comparing number systems (binary, decimal, octal, hexadecimal) and illustrating the timeline of computer generations with key innovations.

Evaluation Method: Rubric-based assessment of the poster presentation on a 10-point scale focusing on:

- Accuracy of number system conversions
- Correct identification of block diagram components
- Visual organization and creativity

Outcome: Learners will demonstrate **basic blocks of a computer and fundamental networking knowledge.**

Activity: Design a concept map showing the internal architecture of a computer and types of networks (LAN, WAN, MAN), including devices and topologies.

Evaluation Method: Checklist-based peer review and instructor validation:

- Completeness of the map
- Correctness of networking concepts

- Use of appropriate terminology
- Logical flow and structure of the map

Outcome: Learners will create professional-level documents and **design visually appealing presentations** using word processing software and presentation software.
Activity: Prepare a formal report (e.g., project proposal) in a word processor and present it using a slide deck with transitions, embedded media, and design elements.
Evaluation Method: Performance-based evaluation using a 10-point scoring scale:

- Formatting and structure of the document
- Presentation aesthetics and clarity
- Communication skills during presentation

Outcome: Learners will manipulate data within spreadsheets, apply formulas, and **generate accurate summaries and visualizations.**
Activity: Analyze a dataset (e.g., student scores or sales data) using spreadsheet software. Apply formulas (SUM, AVERAGE, IF, VLOOKUP) and create relevant charts.
Evaluation Method: Practical test with a rubric:

- Correct use of formulas
- Accuracy of data summaries

Outcome: Learners will apply data modelling techniques to **analyze, organize, and represent data effectively** in various scenarios.
Activity: Prepare an interactive dashboard for a given data set using EXCEL.
Evaluation Method: Evaluation of the dashboard on a 10-point scoring scale:

- Presentation aesthetics and clarity
- Interactiveness
- Communication skills during presentation

SEMESTER-I

COURSE 1: COMPUTER FUNDAMENTALS AND OFFICE AUTOMATION

Practical **Credits: 1** **2 hrs/week**

List of Experiments:

1. Demonstration of Assembling and Dessembling of Computer Systems.
2. Identify and prepare notes on the type of Network topology of your institution.
3. Prepare your resume in Word.
4. Using Word, write a letter to your higher official seeking 10-days leave.
5. Prepare a presentation that contains text, audio and video.
6. Using a spreadsheet, prepare your class Time Table.
7. Using a Spreadsheet, calculate the Gross and Net salary of employees(Min 5) considering all the allowances.
8. Generate the class-wise and subject-wise results for a class of 20 students. Also generate the highest and lowest marks in each subject.
9. Using IF, AND, OR, and IFERROR to Automate Grade Evaluation.
    a. Create a table of student scores in different subjects.
    b. Use IF to assign grades (A/B/C/Fail).
    c. Use IFERROR to handle missing scores or invalid data.
10. *Employee Database Search Using VLOOKUP, HLOOKUP, XLOOKUP, INDEX, and MATCH*
    a. Create a database of employees (Name, ID, Department, Salary).
    b. Implement VLOOKUP to search by employee ID.
    c. Use HLOOKUP to extract department heads by role.
    d. Apply XLOOKUP for more flexible searches.
    e. Use INDEX + MATCH as an alternative to VLOOKUP.
11. Sales Report Analysis Using Pivot Tables and Charts
    a. Use a dataset of product sales (Product, Region, Date, Quantity, Revenue).
    b. Create Pivot Tables to summarize data by region/product.
    c. Insert Pivot Charts for visual analysis (e.g., bar, line).
    d. Add slicers to make the dashboard interactive.
12. Designing a Data Entry Form with Drop-downs and Input Rules
    a. Create a student registration form.
    b. Add drop-down lists for course selection using Data Validation.

c. Add input messages to guide users.
d. Add error alerts for wrong entries.

13. Monthly Budget Planning using Goal Seek and Scenario Manager
    a. Create a simple personal budget (income, expenses, savings).
    b. Use Goal Seek to determine income needed to save a desired amount.
    c. Use Scenario Manager to compare different budgeting scenarios (best/ worst/ realistic case).
    d. Create a one-variable Data Table to analyze how different expenses affect savings.
14. Dashboard Creation Using Combo Charts, Sparklines & Slicers
    a. Use existing sales or attendance data.
    b. Insert combo charts (e.g., column + line).
    c. Add sparklines to show trends.
    d. Use slicers with Pivot Tables to control dashboard elements.
    e. Finalize and format for interactivity.

SEMESTER-I

COURSE 2: PROBLEM SOLVING USING C

Theory **Credits: 3** **3 hrs/week**

Course Objectives:

1. Understand the fundamentals of computer programming, Apply structured problem-solving approaches using algorithms, flowcharts, and C programming constructs.
2. Develop efficient logic using decision-making, loop, and jump control statements.
3. Utilize derived data types like arrays and strings for modular program design.
4. Design and implement modular solutions using functions, recursive logic, pointer operations, and dynamic memory management.
5. Handle complex data structures including structures, unions, and text file operations.

Course Outcomes:

At the end of the course, students will be able to:

1. Understand basic computing concepts, programming paradigms and write structured C programs.
2. Apply control flow statements to solve logical and repetitive tasks in C.
3. Implement arrays and string operations to manage and manipulate data efficiently.
4. Design modular code using functions, recursion, and appropriate parameter passing.
5. Utilize pointers and memory operations for effective data handling. Demonstrate competence in dynamic memory allocation and text file processing.

Unit 1. Introduction to computer programming:

Introduction, Types of software, Compiler and interpreter, Concepts of Machine level, Assembly level and high-level programming, Flowcharts and Algorithms, Fundamentals of C: History of C, Features of C, C Tokens-variables and keywords and identifiers, constants and Data types, Rules for constructing variable names, Operators, Structure of C program, Input /output statements in C-Formatted and Unformatted I/O

Unit 2. Control statements:

Decision making statements: if, if else, else if ladder, switch statements. Loop control statements: while loop, for loop and do-while loop. Jump Control statements: break,continue and goto.

Unit 3. Derived data types in C:

Arrays: One Dimensional arrays - Declaration, Initialization and Memory representation; Two Dimensional arrays -Declaration, Initialization and Memory representation. Strings: Declaring & Initializing string variables; String handling functions, Character handling functions

Unit 4. Functions:

Pointers: Pointer data type, Pointer declaration, initialization, accessing values using pointers. Pointer arithmetic, Pointers and arrays.

Function Prototype, definition and calling. Return statement. Nesting of functions. Categories of functions. Recursion (Basic Concept only). Parameter Passing by address & by value. Local and Global variables. Storage classes: automatic, external, static and register.

Unit 5. Dynamic Memory Management:

Introduction, Functions-malloc, calloc, realloc, free Structures: Basics of structure, structure members, accessing structure members, nested structures, array of structures, structure and functions, structures and pointers. Unions - Union definition; difference between Structures and Unions. Working with text files - modes: opening, reading, writing and closing text files.

Text Books:

1. Programming in ANSI C, E. Balagurusamy, Tata McGraw Hill, 6 th Edn,
2. Computer fundamentals and programming in C, Reema Theraja, Oxford University Press

Reference Books:

1. Let us C, Y Kanetkar, BPB publications
2. Head First C: A Brain-Friendly Guide, David Griffiths, Dawn Griffiths

Activities:

Outcome: Understand basic computing concepts, programming paradigms and write structured C programs.

Activity: Create a concept map of computing fundamentals and programming paradigms (procedural, structured, object-oriented). Then, they write a structured C program (e.g., a calculator or student grade system) using proper syntax, indentation, and modular design.

Evaluation Method: Rubric-based Code Review & Viva to check the

- The correctness of the concept map
- Correct use of structure (main + functions)

- Identification of paradigm used
- Code readability and documentation

Outcome: Apply control flow statements to solve logical and repetitive tasks in C.

Activity: Implement a program that solves a logic puzzle (e.g., number guessing game, pattern generation, or prime number finder) using if, switch, for, while, and do-while.

Evaluation Method: Automated Test Cases + Peer Review to check the

- Correct use of control statements
- Logical correctness of output
- Efficiency and edge case handling
- Peer feedback on clarity and logic

Outcome: Implement arrays and string operations to manage and manipulate data efficiently.

Activity: Build a program that stores and arranges student marks in ascending and descending order using arrays and performs string operations like concatenation, comparing, and formatting names.

Evaluation Method: Functional Demonstration + Code Walkthrough to check the

- Correct array and string usage
- Memory efficiency
- Handling of invalid inputs
- Explanation of sorting/searching logic

Activity:

- **Recursive Problem Solver**

  Students write a modular program to solve a recursive problem (e.g., factorial, Fibonacci, or Tower of Hanoi) using functions with parameters and return values.

Evaluation Method:

- **Code Trace + Written Quiz**
  - Correct function decomposition
  - Proper parameter passing (by value/reference)
  - Recursion depth and base case handling
  - Quiz on tracing recursive calls

Outcome: Utilize pointers and memory operations for effective data handling. Demonstrate competence in dynamic memory allocation and text file processing.

Activity: Create a program that dynamically stores user input (e.g., survey responses) using pointers and writes/reads the data to/from a text file.

Evaluation Method: Memory Debugging + File I/O Assessment to check the

- Proper use of malloc, calloc, realloc, and free
- Pointer arithmetic and dereferencing
- File creation, reading, writing, and error handling
- Use of tools like Valgrind or manual memory trace (Optional for Unix flavours)

SEMESTER-I

COURSE 2: PROBLEM SOLVING USING C

Practical **Credits: 1** **2 hrs/week**

List of Experiments:

1. Write a program to check whether the given number is Armstrong or not.
2. Write a program to find the sum of individual digits of a positive integer.
3. Write a program to generate the first n terms of the Fibonacci sequence.
4. Write a program to find both the largest and smallest number in a list of integer values
5. Write a program to demonstrate change in parameter values while swapping two integer variables using Call by Value & Call by Address
6. Write a program to perform various string operations.
7. Write a program to search an element in a given list of values.
8. Write a program that uses functions to add two matrices.
9. Write a program to calculate factorial of given integer value using recursive functions
10. Write a program for multiplication of two N X N matrices.
11. Write a program to sort a given list of integers in ascending order.
12. Write a program to calculate the salaries of all employees using the Employee (ID, Name, Designation, Basic Pay, DA, HRA, Gross Salary, Deduction, Net Salary) structure.
    a. DA is 30 % of Basic Pay
    b. HRA is 15% of Basic Pay
    c. Deduction is 10% of (Basic Pay + DA)
    d. Gross Salary = Basic Pay + DA+ HRA
    e. Net Salary = Gross Salary - Deduction
13. Write a program to read / write the data from / to a file.
14. Write a program to reverse the contents of a file and store in another file.
15. Write a program to create Book (ISBN,Title, Author, Price, Pages, Publisher)structure and store book details in a file and perform the following operations
    a. Add book details
    b. Search a book details for a given ISBN and display book details, if available
    c. Update a book details using ISBN
    d. Delete book details for a given ISBN and display list of remaining Books

SEMESTER-II

COURSE 3: PYTHON PROGRAMMING AND DATA STRUCTURES

Theory **Credits: 3** **3 hrs/week**

Course Objectives

1. To introduce the fundamentals of Python programming, including environment setup, syntax, and core concepts.
2. To develop problem-solving skills using control flow, functions, and modules.
3. To provide knowledge of Python data structures, file handling, and exception handling for effective programming.
4. To impart object-oriented programming concepts and GUI development skills for building applications.

Course Outcomes (COs)

After successful completion of the course, students will be able to:

1. Explain the basic features, syntax, data types, and operators of Python programming.
2. Apply control flow constructs, functions, and modules to develop structured Python programs.
3. Demonstrate the use of sequences, sets, and dictionaries for effective data handling and manipulation.
4. Implement file handling techniques and apply exception handling mechanisms for robust applications.
5. Develop object-oriented and GUI-based applications using Python.

Unit 1. Basics of Python Programming:

Introduction to Python, Features of Python, Programming Modes - Interactive Mode & Script Mode, Identifiers, Naming Conventions, Keywords (Reserved Words), Built-in Data Types, Literals - Integer, Float, Complex, Boolean, String, Variables, Operators, Expressions, Assignment Statements, Input/Output Statements, Python Syntax (Lines, Comments, Indentation)

Operators & Operands, Classification of Operators - Arithmetic Operators, Relational Operators, Logical Operators, Bitwise Operators, Assignment, Augmented Assignment, Identity Operators, Expressions & Precedence Rules

Unit 2. Control Flow, Functions & Modules:

Control Flow - if Statement, if-else, if-elif-else. Iterative Statements – while, for, Nested Loops, Loop Control Statements – break, continue, pass; else with loops

Need for Functions, Defining & Invoking User-defined Functions, Return Statement, Function Input/Output Cases, Scope of Variables - Local, Global, Nested Functions, Function Arguments - Required, Positional, Default, Variable-length, main() Function, Documentation Strings, Recursive Functions, Anonymous Functions (Lambda), Library Functions

Modules - Import, from..import, Creating & Using Modules, Namespaces

Unit 3. Sequence, Set, Mapping Types:

Strings- Representation, Indexing, Slicing, Immutability, String Operators, Traversal, Accumulation, Formatting & Methods

Lists - Overview, Indexing, Slicing, Methods, Mutability, List Operations - Add, Update, Delete, Search, Copy, Traverse, Comprehension

Tuples - Operations, Immutability, Tuple Assignment, Arrays & Operations

Sets - Overview, Methods, Mathematical Operations, Frozenset, Comprehension

Dictionaries - Overview, Methods, Operations, Traversal, Comparison

Unit 4. File Handling, Exception Handling & Object Oriented Programming:

File Handling - Types, Paths, Basic Operations on Files - Open/Close, Read/Write, CSV Files, OS/Pathlib

Error & Exception Handling - Syntax Errors, Built-in Exceptions, Catching and Handling Exceptions: try-except, raise, User-defined Exceptions, Assertions

OOP Concepts: Classes, Objects, Attributes, Methods, Constructor and Destructors

Encapsulation: Private and Public Members

Inheritance: Single, Multilevel, Multiple, Method Overriding

Unit 5: Abstract Data Structures and GUI Programming

Abstract Data Structures (ADTs): Concepts and Importance

Linked List: Definition, Types- Singly, Doubly, Circular; Node Structure, Insertion, Deletion, Traversal (Single Linked list implementation only)

Stacks: LIFO Principle, Implementation using List, Applications

Queues: FIFO Principle, Implementation using List, Priority Queues

GUI Programming with Tkinter: Widgets (Label, Button, Entry, Menu, Listbox, Canvas etc.), Event Handling, Building Simple GUI Apps

Textbooks:

1. Python Programming-An Object Oriented approach, Anita Goel, Universities Press
2. Python Programming using Problem Solving Approach Reema Thareja Oxford University Press 2020
3. Exploring Python, Budd T A, McGraw-Hill Education, 1st Edition, 2011.

Reference Book:

1. Python: The Complete Reference, Martin C. Brown, Mc Graw-Hill, 2018
2. Fundamentals of Python, Kenneth A. Lambert. (2019), First Programs,2nd Edition, CENGAGE Publication.

Activities:

Outcome: Explain the basic features, syntax, data types, and operators of Python programming.

Activity: Conduct a "Python Basics Lab" where students write small programs to demonstrate literals, variables, data types, and operators (e.g., swapping numbers, simple calculator).

Evaluation Method:

- Lab performance checklist (execution of 3 mini tasks)
- Short quiz with multiple-choice and fill-in-the-blanks on syntax, data types, and operators

Outcome: Apply control flow constructs, functions, and modules to develop structured Python programs.

Activity: Group activity - "Python Problem Solving Challenge": Students solve real-life problems (e.g., finding prime numbers, grade calculator, menu-driven calculator) using control structures, functions, and importing standard modules.

Evaluation Method:

- Code submission with proper use of functions/modules (20%)
- Viva-voce to explain logic and flow of control (40%)
- Unit test with scenario-based programming questions (40%)

Outcome: Demonstrate the use of sequences, sets, and dictionaries for effective data handling and manipulation.

Activity: Hands-on mini project – "Student Data Manager": Students create a program using lists, tuples, sets, and dictionaries to store and manipulate student records (e.g., marks, courses, hobbies).

Evaluation Method:

- Practical demo of program with at least 5 data operations (add, search, delete, update, traverse)
- Evaluation rubric for correctness, efficiency, and use of appropriate data structure

Outcome: Implement file handling techniques and apply exception handling mechanisms for robust applications.

Activity: Individual assignment - "File-Based Address Book": Students create a program to store, update, and retrieve data from files, with exception handling for invalid inputs or missing files.

Evaluation Method:

- Assessment of program correctness (file read/write, append, delete, exception handling)
- Short quiz with error-tracing and debugging questions (given code with errors, students identify and correct)

Outcome: Develop object-oriented and GUI-based applications using Python.

Activity: Mini Project – "Student Information System with GUI": Students design a simple Tkinter-based application with classes/objects for handling student data, including basic GUI widgets (Entry, Button, Listbox).

Evaluation Method:

- Project demo and presentation (50%)
- Rubric-based evaluation for OOP concepts (classes, inheritance, encapsulation) and GUI design (widgets, event handling) (30%)
- Peer review/feedback on usability (20%)

SEMESTER-II

COURSE 3: PYTHON PROGRAMMING AND DATA STRUCTURES

Practical **Credits: 1** **2 hrs/week**

1. Basic Python Programs:
    a. Write a program to display basic details (name, roll number, department) using print() and demonstrate different literal types (int, float, string, boolean, complex).
    b. Write a program to perform arithmetic, relational, logical, bitwise, and assignment operations on given inputs.
2. Control Flow Practice
    a. Write a program to find the largest of three numbers using if-elif-else.
    b. Write a program to check whether a number is prime or not using loops.
    c. Write a program to illustrate the use of loop control statements (break, continue, pass).
3. Functions and Recursion
    a. Write a program to define a function to calculate factorial of a number (using recursion).
    b. Write a program to demonstrate different types of function arguments (default, positional, keyword, variable-length).
4. Write a program to illustrate string slicing, concatenation, repetition, and built-in methods.
5. Write a program to create a list of numbers, perform insertion, deletion, searching, sorting, and list comprehension.
6. Write a program to demonstrate tuple packing, unpacking, and immutability.
7. Write a program to implement set operations (union, intersection, difference, subset, superset).
8. Write a program to create a dictionary of student roll numbers and marks, and perform add, update, delete, and traversal operations.
9. Write a program to read and display count of vowels, consonants, digits, and spaces of a text file.
10. Write a program to copy the contents of one file into another file.
11. Write a program to read and process student marks from a CSV file (calculate average, highest, lowest).

12. Write a program to demonstrate exception handling using try-except-finally.
13. Write a program to create a class Student with attributes and methods to display details.
14. Write a program to demonstrate single and multilevel inheritance.
15. Implement stack (LIFO) and queue (FIFO) using lists and linked lists.
16. Implement singly linked lists: node creation, insertion, deletion, traversal.
17. Write a Tkinter program with Label, Entry, and Button widgets to take user input and display it.
18. Write a Tkinter program to create a simple calculator application.

SEMESTER-II

COURSE 4: STATISTICAL FOUNDATIONS FOR DATA SCIENCE

Theory **Credits: 3** **3 hrs/week**

Course Objectives

1. To introduce the fundamental concepts of probability and statistics for quantifying and analyzing uncertainty in real-world problems.
2. To develop an understanding of random variables, expectations, and common probability distributions (discrete and continuous).
3. To build the ability to summarize and describe data using measures of central tendency, dispersion, correlation, and visualization techniques.
4. To equip students with statistical tools for modeling relationships using correlation and regression analysis.
5. To provide knowledge of estimation and hypothesis testing for making valid inferences from sample data about populations.

Course Outcomes

At the end of the course, students will be able to:

1. **Apply** the basic rules of probability, conditional probability, and Bayes' theorem to solve problems involving uncertainty.
2. **Compute and interpret** descriptive statistics (mean, median, mode, variance, standard deviation, correlation, covariance) and represent data effectively using histograms, bar charts, and scatter plots.
3. **Analyze** random variables and probability distributions (Binomial, Poisson, Normal, Exponential, etc.) to model real-life situations.
4. **Perform** correlation and regression analysis to identify and interpret relationships between variables.
5. **Conduct** statistical inference through confidence intervals and hypothesis testing (z-test, t-test, chi-square, F-test) for decision-making.

Unit 1: Fundamentals of Probability & Basic Statistics

Probability: Concept of Uncertainty, Axioms and rules of probability, Conditional probability and independence, Law of total probability and Bayes' theorem

Measures of central tendency: Mean, Median, Mode

Measures of dispersion: range, interquartile range, variance, standard deviation

Introduction to correlation and covariance

Data representation: histograms, bar charts, scatter plots

Unit 2: Random Variables, Expectation, and Variance

Random variables: definition, types (discrete & continuous), and properties, Probability mass function (PMF) and probability density function (PDF), Cumulative distribution function (CDF), Mathematical expectation (mean), variance, and standard deviation, Moments and moment-generating functions

Unit 3: Probability Distributions

Discrete distributions: Binomial, Poisson, Geometric, Negative Binomial distributions - definitions, properties, and examples

Continuous distributions: Uniform, Normal (Gaussian), Exponential, Gamma distributions - definitions, properties, and applications

Joint, marginal, and conditional distributions, Introduction to Central Limit Theorem

Unit 4: Correlation and Regression

Bivariate data and scatter plots

Correlation: Pearson and Spearman coefficients, interpretation

Simple linear regression: model, estimation, properties, and analysis of variance

Multiple linear regression basics (conceptual understanding)

Residuals and goodness of fit

Unit 5: Statistical Inference, Estimation, and Hypothesis Testing

Population and sample, parameters and statistics, Sampling distributions, Point and interval estimation (confidence intervals), Tests of significance: z-test, t-test, chi-square test, and F-test, p-values and errors (Type I & II), Power of a statistical test

Textbooks:

1. Probability and Statistics for Engineers and Scientists, Ronald E. Walpole, Wiley.
2. Sheldon M. Ross, Introduction to Probability and Statistics for Engineers and Scientists

3. Douglas C. Montgomery & George C. Runger, Applied Statistics and Probability for Engineers

Reference Books:

1. D.C. Agarwal, Statistics for Data Science and AI
2. Larry J. Stephens, Excel Data Analysis: Your visual blueprint for analyzing data, statistics, and AI

Activities:

Outcome: Apply probability rules, conditional probability, and Bayes' theorem

Activity: Classroom **Quiz** (MCQs & short problems on probability, conditional probability, Bayes).

Evaluation Method: Individual quiz marks (accuracy of solutions).

Outcome: Compute and interpret descriptive statistics & visualize data

Activity: Poster Presentation – students prepare posters with a dataset summary (mean, median, variance, histograms, scatter plot).

Evaluation Method: Rubric-based evaluation (clarity, correctness, creativity, interpretation).

Outcome: Analyze random variables and probability distributions

Activity: Seminar/Presentation – each student (or group) explains one distribution (Binomial, Poisson, Normal, Exponential, etc.) with real-life examples and graphs.

Evaluation Method: Seminar grading (content accuracy, explanation, use of examples, presentation skills).

Outcome: Perform correlation and regression analysis

Activity: Case Study Assignment – students are given real-world data (e.g., sales vs advertising, study hours vs marks) to compute correlation/regression and interpret.

Evaluation Method: Submission of case study report + viva (correctness of analysis and interpretation).

Outcome: Conduct statistical inference (estimation & hypothesis testing)

Activity: Group Discussion/Debate – teams test a given hypothesis (e.g., "Male and Female students score equally") using sample data and statistical tests, then defend their conclusion.

Evaluation Method: Marks based on correct application of test, interpretation of p-value, and clarity in argument.

SEMESTER-II

COURSE 4: STATISTICAL FOUNDATIONS FOR DATA SCIENCE

Practical **Credits: 1** **2 hrs/week**

Advanced Spreadsheets/Excel Lab/PSPP Open Source

1. Construct a contingency table from sales data and compute conditional probabilities. Verify independence of variables.
2. Apply Bayes' theorem on medical test data to compute the probability of disease given a positive result.
3. Calculate measures of central tendency (mean, median, mode) for student marks dataset.
4. Compute measures of dispersion (range, variance, standard deviation, IQR) for the same dataset and interpret variability.
5. Create a histogram of student marks and comment on the shape of the distribution.
6. Prepare bar charts of categorical data (e.g., Gender vs Section) and interpret group comparisons.
7. Generate scatter plots between Hours Studied and Exam Score, compute correlation and covariance, and interpret the relationship.
8. Random Variable Simulation: Simulate and visualize discrete/continuous random variables using Excel functions and Data Analysis Tool pack.
9. Expectation & Variance Calculation: Use Excel formulas to compute expected value, variance, and standard deviation from given datasets.
10. Modeling Discrete Probability Distributions: Generate and plot Binomial and Poisson distributions; analyze probabilities and mean/variance.
11. Modeling Continuous Distributions: Simulate Normal and Exponential distributions; use NORM.DIST, NORM.INV, EXPON.DIST functions.
12. Correlation Analysis: Calculate Pearson/Spearman correlation coefficients; visualize with scatter plots and trendlines.
13. Linear Regression in Excel: Fit a linear regression model, interpret coefficients, predict new values; use REGRESSION tool.
14. Statistical Inference & Estimation: Create confidence intervals using Excel formulas; visualize sampling distributions
15. Hypothesis Testing: Perform z-test, t-test, chi-square tests in Excel; interpret p-values and results.

SEMESTER-III

COURSE 5: DATABASE MANAGEMENT SYSTEMS

Theory **Credits: 3** **3 hrs/week**

Course Objectives:

1. To understand the fundamentals of data, information, and the evolution from file-based systems to modern database management systems.
2. To develop the ability to design conceptual data models using Entity-Relationship (ER) and Enhanced ER diagrams.
3. To explore relational model principles, such as keys, integrity constraints and normalization.
4. To perform data definition and manipulation using SQL commands including queries, joins, subqueries, views, and set operations.
5. To apply procedural logic using PL/SQL, incorporating control structures, functions, procedures, and database triggers.

Course Outcomes:

At the end of the course, students will be able to:

1. **Describe** the fundamentals of data, database systems, and the differences between file-based and database approaches. **Compare and classify** various DBMS architectures, data models, and their components, including the three-schema architecture.
2. **Design** conceptual data models using Entity-Relationship and Enhanced ER diagrams, applying generalization, specialization, and constraints.
3. **Apply** relational model concepts, including CODD rules and normalization techniques.
4. **Construct and execute** SQL queries for data definition, manipulation, aggregation, joining, and subqueries, including views and set operations.
5. **Develop** PL/SQL programs incorporating control structures, procedures, and functions to manage database behavior effectively.

Unit 1. Overview of Database Management System:

Introduction to data, information, database, database management systems, file-based system, Drawbacks of file-Based System, database approach, Classification of Database Management Systems, advantages of database approach, Various Data Models, Components of Database Management System, three schema architecture of data base, costs and risks of database approach.

Unit 2. Entity-Relationship Model:

Introduction, the building blocks of an entity relationship diagram, classification of entity sets, attribute classification, relationship degree, relationship classification, reducing ER diagram to tables, enhanced entity-relationship model (EER model), generalization and specialization, IS A relationship and attribute inheritance, multiple inheritance, constraints on specialization and generalization, advantages of ER modeling.

Unit 3. Relational Model:

Introduction, CODD Rules, relational data model, concept of key, relational integrity, relational algebra, relational algebra operations, advantages of relational algebra, limitations of relational algebra, Functional dependencies and normal forms.

Unit 4. Structured Query Language:

Introduction, Commands in SQL, Data Types in SQL, Data Definition Language, Selection Operation, Projection Operation, Aggregate functions, Data Manipulation Language, Table Modification Commands, Join Operation, Set Operations, View, Sub Query.

Unit 5. PL/SQL:

Introduction, Shortcomings of SQL, Structure of PL/SQL, PL/SQL Language Elements,Data Types, Operators Precedence, Control Structures, Steps to Create a PL/SQL, Program, Iterative Control, Procedures, Functions.

Textbooks:

1. Database System Concepts, Avi Silberschatz, Henry F. Korth,S. Sudarshan, Seventh Edition, McGraw-Hill
2. Database Management Systems by Raghu Ramakrishnan, McGrawhill

Reference Books:

1. Fundamentals of Database Systems, Elmasri Navathe Pearson Education
2. An Introduction to Database systems, C.J. Date, A.Kannan, S.Swami Nadhan, Pearson

Activities:

Outcome: Describe the fundamentals of data, database systems, and the differences between file-based and database approaches. Compare and classify various DBMS architectures, data models, and their components, including the three-schema architecture.

Activity: Create a comparative presentation or infographic illustrating:

- File-based vs. DBMS approaches
- Types of DBMS architectures (1-tier, 2-tier, 3-tier)
- Data models and the three-schema architecture

Evaluation Method: Rubric-based assessment of the presentation covering clarity, accuracy, and depth of comparison. Include a short quiz to test conceptual understanding.

Outcome: Design conceptual data models using Entity-Relationship and Enhanced ER diagrams, applying generalization, specialization, and constraints.

Activity: Model a university or hospital database using ER and Enhanced ER diagrams that shows:

- Entity sets, relationships
- Generalization/specialization
- Participation and cardinality constraints

Evaluation Method: Diagram submission with peer review and instructor feedback. Use a checklist to assess completeness, correctness, and notation usage.

Outcome: Apply relational model concepts, including CODD rules, and normalization techniques.

Activity: Normalize a given Database upto 3NF.

Evaluation Method: Written assignment graded on:

- Correctness of normalization steps
- Short-answer questions on CODD rules

Outcome: Construct and execute SQL queries for data definition, manipulation, aggregation, joining, and subqueries, including views and set operations.

Activity: Implement a mini-project (e.g., Library or Inventory DB) using SQL. Include:

- Table creation (DDL)
- Data manipulation (DML)
- Aggregation, joins, subqueries, views, and set operations

Evaluation Method: Lab-based practical test with query execution and output validation. Include a viva to explain logic and optimization.

Outcome: Develop PL/SQL programs incorporating control structures, procedures and functions to manage database behaviour effectively.

Activity: Build a PL/SQL-based payroll or student grading system using:

- Procedures and functions
- Control structures (IF, LOOP)
- Triggers for automated updates

Evaluation Method: Code review and demonstration. Evaluate based on:

- Syntax correctness
- Logical flow

SEMESTER-III

COURSE 5: DATABASE MANAGEMENT SYSTEMS

Practical **Credits: 1** **2 hrs/week**

Experiment 1 : Database: Inventory Management

Table 1: Products

Structure:

Column Name	Data Type	Constraints
product_id	INT	PRIMARY KEY
product_name	VARCHAR(50)	NOT NULL
price	DECIMAL(10,2)	CHECK(price > 0)
stock_qty	INT	CHECK(stock_qty >= 0)

Sample Data:

product_id	product_name	price	stock_qty
1	Pen	10.00	100
2	Notebook	50.00	200
3	Stapler	120.00	50
4	Marker	25.00	80
5	File Folder	60.00	150

Table 2: Suppliers

Structure:

Column Name	Data Type	Constraints
supplier_id	INT	PRIMARY KEY
supplier_name	VARCHAR(50)	NOT NULL
contact_no	VARCHAR(20)	UNIQUE
product_id	INT	FOREIGN KEY REFERENCES Products(product_id)

Sample Data:

supplier_id	supplier_name	contact_no	product_id
101	StationeryMart	9876543210	1
102	PaperWorld	9876500000	2
103	OfficeSupplies	9876512345	3
104	MarkerHub	9876522222	4
105	FileDepot	9876533333	5

Section A: DDL (Data Definition Language)

1. Create a database called InventoryDB.
2. Create a table Products and table Suppliers with the specified columns and constraints:

Section B: DML (Data Manipulation Language)

4. Insert at least 5 rows into the Products table.
5. Insert at least 5 rows into the Suppliers table.
6. Update the stock quantity of product 'Pen' to 120.
7. Delete a supplier with a specific supplier_id.
8. Write a query to rename 'Notebook' to 'NoteBook A4'

Section C: DQL (SELECT Queries)

9. Display all records from the Products table.
10. Display only product_name and price of all products.
11. List all products that have a stock quantity less than 100.
12. Show all products between 20 and 100 price range.
13. Find all suppliers whose contact number starts with '98765'.
14. Find the average price of products.
15. Display the total number of products in the inventory.
16. Show the maximum and minimum stock quantities.
17. Count how many suppliers supply each product.
18. Show all products where price > 50 AND stock_qty > 100.
19. Show all products where price < 20 OR stock_qty < 80.
20. Display suppliers whose supplier_name contains the word 'Mart'
21. List all suppliers along with the product they supply (use INNER JOIN).
22. Display suppliers whose name starts with 'S'.
23. Find products whose name has exactly 5 characters
24. Find suppliers who supply products costing more than 100.

Experiment 2 : ONLINE BOOKSTORE DB

An online book store wants to implement a BOOKSTORE DB for managing their online transactions by using the following tables.

Authors Table

Column Name	Data Type	Constraints
author_id	INTEGER	PRIMARY KEY
first_name	VARCHAR	NOT NULL
last_name	VARCHAR	NOT NULL
nationality	VARCHAR	NULL allowed

Books Table

Column Name	Data Type	Constraints
book_id	INTEGER	PRIMARY KEY

Title	VARCHAR	NOT NULL
author_id	INTEGER	FOREIGN KEY REFERENCES Authors
publication_year	INTEGER	
Price	DECIMAL	

Customers Table

Column Name	Data Type	Constraints
customer_id	INTEGER	PRIMARY KEY
first_name	VARCHAR	NOT NULL
last_name	VARCHAR	NOT NULL
Email	VARCHAR	UNIQUE, NOT NULL
Address	VARCHAR	NOT NULL

Orders Table

Column Name	Data Type	Constraints
order_id	INTEGER	PRIMARY KEY
customer_id	INTEGER	FOREIGN KEY REFERENCES Customers
book_id	INTEGER	FOREIGN KEY REFERENCES Books
order_date	DATE	NOT NULL
quantity	INTEGER	NOT NULL

SAMPLE DATA SET for BOOKSTORE DB

Authors Table

author_id	first_name	last_name	nationality
1	Jane	Austen	British
2	George	Orwell	British
3	Gabriel	Garcia Marquez	Colombian
4	Toni	Morrison	American
5	Mark	Twain	American
6	Harper	Lee	American
7	Fyodor	Dostoevsky	Russian

Books Table

book_id	Title	author_id	publication_year	price

101	Pride and Prejudice	1	1813	12.99
102	1984	2	1949	9.50
103	One Hundred Years of Solitude	3	1967	15.00
104	Beloved	4	1987	11.25
105	Animal Farm	2	1945	8.75
106	Adventures of Huckleberry Finn	5	1884	10.50
107	To Kill a Mockingbird	6	1960	14.00

Customers Table

customer_id	first_name	last_name	Email	address
201	Alice	Smith	alice.s@example.com	12 Oak St, London
202	Bob	Johnson	bob.j@example.com	45 Pine Ave, Oxford
203	Charlie	Brown	charlie.b@example.com	78 Maple Rd, Bristol
204	Diana	Prince	diana.p@example.com	34 Queen St, York
205	Edward	Norton	edward.n@example.com	22 River Ln, Leeds
206	Fiona	Hall	fiona.h@example.com	56 Lake Dr, Bath
207	Greg	Miller	greg.m@example.com	89 Park Ave, Glasgow

Orders Table

order_id	customer_id	book_id	order_date	Quantity
301	201	101	2025-07-20	1
302	202	102	2025-07-21	2
303	201	105	2025-07-22	1
304	203	103	2025-07-23	1
305	204	106	2025-07-24	1
306	205	107	2025-07-25	3
307	206	104	2025-07-26	2

Section A: DDL (Schema Design & Constraints)

1. Write SQL statements to create all 4 tables (Authors, Books, Customers, Orders) with:
    - Primary Keys
    - Foreign Keys
    - Appropriate data types
    - NOT NULL constraints where necessary.
2. Alter the Books table to add a constraint that price must be greater than 0.

3. Add a new column phone_number to the Customers table (VARCHAR(15)) and ensure it is unique.
4. Drop the phone_number column from the Customers table.

Section B: DML (Data Manipulation)

5. Insert at least 7 records for each table (use sample dataset above).
6. Update the price of the book titled *Animal Farm* by increasing it by 10%.
7. Delete all orders made before 2025-07-21.
8. Change the nationality of Gabriel Garcia Marquez to "Latino-American".

Section C: SELECT Queries (Data Querying)

9. List all books published between 1900 and 2000.
10. Find all customers whose email contains "example.com".
11. Retrieve books whose price is between 10 and 15 and published before 1950.
12. Show authors who are either ‘British’ or ‘American’.
13. Find books that have a price less than 10 or are published after 1980.
14. Display all orders placed after 2025-07-22.
15. List all books written by author with author_id = 2.
16. Find customers whose last name starts with B.
17. Show all books with a price NOT between 9 and 13.
18. Display books whose publication_year is in (1813, 1945, 1987).
19. Find authors whose nationality is NOT ‘British’.
20. List customers whose address contains the word Park.
21. Show all books sorted by price in descending order.
22. List authors in alphabetical order by last_name.
23. Display orders sorted by order_date (latest first).

Use of Date Functions

24. Show all orders placed in July 2025.
25. Show all orders with an estimated delivery date (5 days after order date).
26. Show customers who placed an order on a weekend.
27. Calculate how many days have passed since the last order was placed.

Aggregate Functions (COUNT, SUM, AVG, MIN, MAX)

28. Count the total number of books in the database.
29. Find the average price of all books.
30. Show the highest-priced book.
31. Count how many orders each customer has placed.
32. Calculate the total sales (price × quantity) for each customer.

GROUP BY and HAVING

33. Count how many books are written by each author.
34. Group orders by customer_id and display total quantity ordered.
35. Show customers who have ordered more than 2 books in total (use HAVING).
36. Find the total number of books sold per author (GROUP BY author).

Experiment 3: EMPLOYEE DB

An enterprise wants to automate its employee management process by implementing an Employee Database.The goal is to replace manual record-keeping with a centralized system that stores employee, department, and project details. Use the following table structures and data set to implement Employee DB.

EmployeeDB - Table Structures

1. Departments Table

Column	**Type**	**Constraints**
dept_id	INT	PRIMARY KEY
dept_name	VARCHAR	UNIQUE, NOT NULL
location	VARCHAR	NOT NULL

2. Employees Table

Column	Type	Constraints
emp_id	INT	PRIMARY KEY
first_name	VARCHAR	NOT NULL
last_name	VARCHAR	NOT NULL
email	VARCHAR	UNIQUE, NOT NULL
phone	VARCHAR	CHECK (phone LIKE '--____')
hire_date	DATE	NOT NULL
job_title	VARCHAR	NOT NULL
salary	DECIMAL	CHECK (salary > 0)
dept_id	INT	FOREIGN KEY REFERENCES Departments(dept_id)
manager_id	INT	FOREIGN KEY REFERENCES Employees(emp_id) (self-referential)

3. Projects Table

Column	**Type**	**Constraints**
project_id	INT	PRIMARY KEY
project_name	VARCHAR	NOT NULL
start_date	DATE	NOT NULL
end_date	DATE	NULL
dept_id	INT	FOREIGN KEY REFERENCES Departments(dept_id)

4. Employee_Project Table (Many-to-Many)

Column	Type	Constraints
emp_id	INT	FOREIGN KEY REFERENCES Employees(emp_id), PRIMARY KEY(emp_id, project_id)
project_id	INT	FOREIGN KEY REFERENCES Projects(project_id)
hours_allocated	INT	CHECK (hours_allocated > 0)

Sample Data Set

Departments Table

dept_id	dept_name	Location
1	HR	New York
2	IT	San Francisco
3	Finance	Chicago
4	Marketing	Boston
5	Operations	Seattle
6	Legal	Washington D.C.
7	Sales	Dallas
8	R&D	Austin
9	Procurement	Denver
10	Customer Care	Miami

2. Employees Table

emp_id	first_name	last_name	Email	phone	hire_date	job_title	salary	dept_id	manager_id
101	Alice	Johnson	alice.j@corp.com	123-456-7890	2020-03-15	HR Manager	75000	1	NULL
102	Bob	Smith	bob.s@corp.com	234-567-8901	2019-05-20	IT Analyst	65000	2	104
103	Charlie	Brown	charlie.b@corp.com	345-678-9012	2021-01-10	Finance Executive	58000	3	106

104	Diana	Prince	diana.p@corp.com	456-789-0123	2018-07-12	IT Manager	90000	2	NULL
105	Ethan	Hunt	ethan.h@corp.com	567-890-1234	2022-02-25	Marketing Lead	62000	4	NULL
106	Fiona	Hall	fiona.h@corp.com	678-901-2345	2017-11-01	Finance Manager	85000	3	NULL
107	Greg	Miles	greg.m@corp.com	789-012-3456	2023-04-15	IT Support	45000	2	104
108	Hannah	White	hannah.w@corp.com	890-123-4567	2021-09-05	HR Executive	50000	1	101
109	Ian	Scott	ian.s@corp.com	901-234-5678	2020-11-20	Operations Analyst	56000	5	NULL
110	Julia	Adams	julia.a@corp.com	012-345-6789	2019-12-18	Legal Advisor	70000	6	NULL

3. Projects Table

project_id	**project_name**	**start_date**	**end_date**	**dept_id**
201	Payroll System	2023-01-01	NULL	3
202	Website Upgrade	2023-02-10	NULL	2
203	Recruitment Drive	2023-03-05	NULL	1
204	Ad Campaign	2023-05-20	NULL	4
205	New CRM Tool	2023-04-15	NULL	7
206	Compliance Portal	2023-06-10	NULL	6
207	Inventory System	2023-07-01	NULL	5
208	AI Research	2023-08-05	NULL	8
209	Customer Feedback	2023-09-10	NULL	10
210	Procurement System	2023-10-01	NULL	9

4. Employee_Project Table

emp_id	project_id	hours_allocated
102	202	120

104	202	80
103	201	100
106	201	150
101	203	50
105	204	70
107	202	60
109	207	90
110	206	110
108	203	40

Section A: DDL (Schema Creation & Modification)

1. Write SQL statements to create the above tables with the specified constraints
2. Alter the Employees table to add a column bonus DECIMAL(8,2) with default value0.
3. Drop the column bonus from Employees.

Section B: DML (Insert, Update, Delete)

4. Insert at least 10 rows into Departments, Employees, Projects, and Employee_Project.(use the above data set)
5. Try inserting an employee with a negative salary (should fail due to CHECK constraint).
6. Update the salary of the employee with emp_id = 103 by 15%.
7. Delete an employee record who has resigned (choose any emp_id).
8. Increase all employees' salaries in the IT department by 5%.
9. Change the department of an employee to "Research".(should fail due to FK constraint)

Section C: DQL (Select Queries)

10. List all employees and their details.
11. Show all employees in the "HR" department.
12. Find employees with salaries between 50,000 and 80,000.
13. Retrieve employees hired after 2020.
14. Show employees who are in either the IT or Finance department.
15. Find employees whose email ends with "@corp.com".
16. List all employees with salary > 60,000 AND located in "New York".
17. Display employees in descending order of salary.
18. Count the number of employees in each department.
19. Show the average salary of employees department-wise.
20. Display departments where the average salary is greater than 70,000.
21. Find the number of employees in each project.
22. Display departments with more than 3 employees.
23. Show the sum of all salaries department-wise.
24. List all distinct department IDs from the Employees table.

25. Show employee names with the year they were hired.
26. Show employees grouped by the year of hire.
27. List employees hired in the last 90 days.
28. List the no of years of experience of all the employees

Section D: Joins

29. List all employees with their department names (INNER JOIN).
30. Display all departments along with employees, including those departments without employees (LEFT JOIN).
31. Show employees and the projects they are working on (JOIN 3 tables: Employees, Employee_Project, Projects).
32. List projects along with total hours allocated by employees.
33. Write a query to find employees who are working on more than one project.
34. Show all projects handled by the 'Finance' department.

Section E: PL/SQL Programming

1. Write a procedure GetEmpInfo that takes emp_id as input and displays name, salary, and department.
2. Write a PL/SQL block that checks if an employee's salary is above 50,000.
   If yes, print "High Salary" ;Otherwise print "Standard Salary".
3. Write a PL/SQL program to display the top 10 rows in the Emp table based on their job and salary
4. Write a stored procedure GiveBonus that takes department ID and a designation as input, along with a bonus amount, and updates the salary of all employees in that department who have the specified designation by adding the bonus amount to their current salary.
5. Create a trigger to prevent inserting employees with a salary less than 30,000.
6. Create a trigger to avoid any transactions(insert, update, delete) on the EMP table on Saturday & Sunday.

SEMESTER-III

COURSE 6: DATA SCIENCE WITH R

Theory **Credits: 3** **3 hrs/week**

Course Objectives

1. Introduce the data science process, lifecycle, and applications in real-world domains.
2. Build proficiency in R programming for data manipulation, exploration, and visualization.
3. Train students in handling structured, unstructured, and time-based data effectively.
4. Familiarize with basic machine learning and statistical modeling using R.
5. Develop awareness of ethical, interpretability, and responsible use of data science.

Course Outcomes

At the end of the course, students will be able to:

1. Explain the Data Science process and perform EDA (Exploratory Data Analysis).
2. Write R programs using variables, functions, loops, and packages for basic analytics.
3. Perform data wrangling, cleaning, and visualization with R libraries (dplyr, tidyr, ggplot2).
4. Build and evaluate basic machine learning models such as regression and clustering.
5. Apply data science techniques to practical case studies.

Unit 1. Introduction to Data Science Process:

Introduction- Definition - Data Science in various fields - Examples - Impact of Data Science - Data Analytics Life Cycle - Data Science Toolkit - Data Scientist - Data Science Team, Exploratory Data Analysis (EDA), Feature Engineering & Data Transformation

Unit 2. Basics of R Programming:

Introduction to R and RStudio, Data Types, Variables, Operators, Control Structures (if, loops, apply), Functions and Packages, Data Input/Output (CSV, Excel, XML, JSON).

Unit 3. Data Handling & Visualization in R:

Data Frames, Lists, Matrices, Data Wrangling with dplyr and tidyr, Handling Missing Data, Working with Date/Time in R. Visualization with ggplot2: grammar of graphics, aesthetics, geometries, scales.

Faceting and layering techniques,Visualizing categorical and numerical data, Customizing and exporting plots

Unit 4. Applications & Case Studies in Data Science:
Simple Linear Regression, Multiple Regression
Model Evaluation Method: Accuracy, Confusion Matrix, ROC.
K-Means Clustering, Text Mining & Word Clouds, Recommender Systems Basics, Ethical Issues in Data Science

Unit 5. Advanced Topics in Data Science with R :
Introduction to Time Series Analysis in R (ARIMA basics)- Concept of time series (trend, seasonality, noise), Time series objects in R (ts, zoo, xts), Plotting and decomposing time series, Stationarity and differencing, Autocorrelation & Partial Autocorrelation (ACF/PACF), AR, MA, ARIMA model basics, Forecasting using forecast package
Creating interactive visualizations with plotly packages-Converting ggplot2 plots to interactive plots
Animations and sliders in plotly

R Shiny: Building interactive web applications-Introduction to Shiny framework, UI and server functions, Reactive expressions and reactivity in Shiny, Input and output widgets (sliders, dropdowns, text), Layouts and dashboard design

Textbooks

1. An Introduction to Statistical Learning with Applications in R, Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani, Springer, 2nd Edition, 2021
2. R for Data Science, Hadley Wickham and Garrett Grolemund, O'Reilly Media, 2017.

Reference Books

1. The Art of R Programming, Norman Matloff,, No Starch Press, 2011.
2. Modern Applied Statistics with S, W.N. Venables & B.D. Ripley, Springer, 2002.
3. Introduction to Data Science: Data Analysis and Prediction Algorithms with R, Rafael A. Irizarry, CRC Press, 2020.
4. Data Science from Scratch: First Principles with Python (for conceptual clarity only), Joel Grus,

Activities:

Outcome: Explain the Data Science process and perform EDA (Exploratory Data Analysis).

Activity: Use a real-world dataset (e.g., Titanic or COVID data) to:

- Outline the steps of the Data Science workflow
- Perform EDA using summary statistics and visualizations (histograms, boxplots, scatterplots)

Evaluation Method: Presentation and checklist (10-point scale):

- Clear explanation of workflow stages
- Quality of EDA insights
- Use of appropriate plots and summaries

Outcome: Write R programs using variables, functions, loops, and packages for basic analytics.

Activity: Write an R script that:

- Reads a CSV file
- Uses if, for, and while loops
- Defines and calls custom functions with arguments and return values

Evaluation Method: Code review and execution test to verify (10-point scale):

- Correctness of the syntax and logic
- Functionality of control structures
- Output accuracy and modularity

Outcome: Perform data wrangling, cleaning, and visualization with R libraries (dplyr, tidyr, ggplot2).

Activity: Clean a messy dataset using:

- dplyr for filtering, selecting, and mutating
- tidyr for reshaping and handling missing values
- Time-based operations (e.g., filling gaps, formatting dates)

Evaluation Method: Before-and-after comparison (10 point score):

- Completeness of cleaning steps
- Use of appropriate functions
- Handling of missing/time data

Outcome: Implement basic machine learning models and evaluate performance using appropriate metrics and visual tools.

Activity: Build a simple classification model (e.g., logistic regression or decision tree) using R:

- Train/test split
- Predict outcomes
- Evaluate using confusion matrix, accuracy, precision, recall

Evaluation Method: Model report and demo (10 point scale):

- Correct implementation of model
- Use of evaluation metrics

SEMESTER-III

COURSE 6: DATA SCIENCE WITH R

Practical **Credits: 1** **2 hrs/week**

List of Practicals:

1. Compute Mean, Median, Mode, Variance, and Standard Deviation
2. Visualize Binomial, Normal, and Poisson Distributions
3. Perform t-test and Chi-Square Test in R
4. Calculate Correlation and Build a Simple Linear Regression Model
5. Conduct Exploratory Data Analysis (EDA) on a Real-World Dataset
6. Apply Feature Engineering: Scaling, Normalization, and Encoding
7. Practice R Programming: Variables, Control Structures, and Functions
8. Read and Write Data from CSV, Excel, JSON, and XML Files
9. Use dplyr and tidyr for Data Wrangling Tasks
10. Handle Missing Data and Detect Outliers
11. Work with Dates and Times in R
12. Visualize Data Using ggplot2 (Bar, Scatter, Histogram, Boxplot)
13. Perform K-Means Clustering and Visualize Clusters
14. Evaluate Models Using Confusion Matrix, Accuracy, and ROC Curve
15. Perform Text Mining and Create a Word Cloud
16. Time Series Forecasting with ARIMA on a real dateset (e.g., monthly airline passengers, stock prices, or temperature data).
17. Create interactive bar, line, and scatter plots using plotly. On a real dataset (e.g., COVID-19 cases, sales data, or student marks).
18. Develop a Shiny app that lets users upload a CSV file.

SEMESTER-III

COURSE 7: WEB TECHNOLOGIES

Theory **Credits: 3** **3 hrs/week**

Course Objectives

1. Understand the principles of web design and distinguish between web and desktop application architectures.
2. Develop static web pages using HTML elements, attributes, and multimedia integration techniques.
3. Style web pages effectively using CSS, including layout control, responsive design, and UI enhancements.
4. Implement dynamic behaviors and form validations using JavaScript and the Document Object Model (DOM).
5. Explore JSON and jQuery for handling structured data and simplifying client-side scripting in web development.

Course Outcomes

At the end of the course, students will be able to:

1. Design and structure HTML-based webpages incorporating text, images, tables, forms, and multimedia content.
2. Apply CSS styling rules to manage layout aesthetics, interactivity, and responsiveness across devices.
3. Use JavaScript for string manipulation, event handling, arrays, object operations, and basic validation.
4. Employ client-side scripting to enhance form functionality, create dialog interactions, and add animations via events.
5. Parse JSON data and use jQuery to simplify DOM manipulation, AJAX calls, and build dynamic, data-driven web applications.

Unit 1.HTML:

Introduction to web designing, difference between web applications and desktop applications, introduction to HTML, HTML structure, elements, attributes, headings, paragraphs, images, tables, lists, blocks, symbols, embedding multi-media components in HTML, HTML forms

Unit 2.CSS:

CSS home, introduction, syntax, CSS combinators, colors, background, borders, margins, padding, height/width, text, fonts, tables, lists, position, overflow, float, pseudo class, pseudo elements, opacity, tool tips, image gallery, CSS forms, CSS counters.

Unit 3.Java Script:

What is DHTML, JavaScript, basics, variables, operators, statements, string manipulations, mathematical functions, arrays, functions. objects, regular expressions, exception handling.

Unit 4. Client-Side Scripting:

Accessing HTML form elements using Java Script object model, basic data validations, data format validations, generating responsive messages, opening windows using java script, different kinds of dialog boxes, accessing status bar using java script, embedding basic animative features using different keyboard and mouse events.

Unit 5. JSON and jQuery

Introduction to JSON: Need for data exchange formats, JSON syntax, JSON vs XML, parsing JSON, creating JSON objects and arrays, accessing nested JSON data, reading/writing JSON in JavaScript.

Working with jQuery: Introduction to jQuery, selectors, filters, DOM manipulation, event handling, animations, effects, and chaining.

Text Book(s)

1. Web Programming: Building Internet Applications, Chris Bates, Wiley, Second Edition.
2. An Introduction to Web Design plus Programming, Paul S. Wang, Sanda S. Katila, Thomson.
3. Learning jQuery – Jonathan Chaffer, Karl Swedberg, Packt Publishing.
4. JSON at Work – Tom Marrs, O'Reilly Media.

Reference Books

1. Head First HTML and CSS, Elisabeth Robson, Eric Freeman, O'Reilly Media Inc.
2. An Introduction to HTML and JavaScript: for Scientists and Engineers, David R. Brooks, Springer.

3. Schaum's Easy Outline: HTML, David Mercer, McGraw Hill Professional.
4. jQuery in Action, Bear Bibeault, Yehuda Katz, Manning Publications.
5. Beginning JSON, Ben Smith, Apress.

Activities:

Outcome: Design and structure HTML webpages with text, images, tables, forms, and multimedia.
Activity: Create a personal profile webpage with all these elements.
Evaluation Method: Checklist (10 pts) - correct tags, structure, working forms/media.

Outcome: Apply CSS rules for layout, aesthetics, interactivity, and responsiveness.
Activity: Style the profile page using colors, fonts, Flexbox/Grid, hover effects, and media queries.
Evaluation Method: Rubric (10 pts) - visual appeal, responsiveness, selector/layout usage, clean code.

Outcome: Use JavaScript for string manipulation, events, arrays/objects, and validation.
Activity: Add form validation, greeting message, array/object display, and button click handling.
Evaluation Method: Demo & testing (10 pts) - correct syntax, validation, event handling, output behavior.

Outcome: Employ client-side scripting for dialogs, animations, and event-based interactions.
Activity: Add show/hide sections, confirmation dialog on submit, and animations via mouse/keyboard events.
Evaluation Method: Live demo (10 pts) - smooth interaction, correct event listeners, proper animations, good UX.

Outcome: Parse JSON and use jQuery for DOM manipulation, AJAX, and dynamic data display.
Activity: Fetch JSON via jQuery AJAX, render as table/list, add filter/sort, and compute summaries.
Evaluation Method: Demo & checklist (10 pts) - JSON parsing, jQuery usage, filter/sort functionality, error handling.

SEMESTER-III

COURSE 7: WEB TECHNOLOGIES

Practical **Credits: 1** **2 hrs/week**

List of Experiments:

1. Create an HTML document with the following formatting options:

   (a) Bold, (b) Italics, (c) Underline, (d) Headings (Using H1 to H6 heading styles), (e) Font (Type, Size and Color), (f) Background (Colored background/Image in background), (g) Paragraph, (h) Line Break, (i) Horizontal Rule, (j) Pre tag
2. Create an HTML document which consists of:

   (a) Ordered List (b) Unordered List (c) Nested List (d) Image
3. Collect any ten images of your choice. Using table tag, align the images as follows:

4. Create a form using HTML which has the following types of controls:

   (a) Text Box (b) Option/radio buttons (c) Check boxes (d) Reset and Submit buttons
5. Embed a calendar object in your web page.
6. Create a form that accepts the information from the subscriber of a mailing system.
7. Apply CSS to design a student registration form (use different selectors, colors, borders, spacing).
8. Create a responsive webpage using CSS Flexbox/Grid.
9. Add hover effects and transitions on images and buttons using CSS.
10. Write a JavaScript program to perform string operations (reverse, substring, count vowels).
11. Create a JavaScript form validation program (check email format, password length, required fields).

12. Develop a webpage that displays greetings based on the current time (morning, afternoon, evening).
13. Use JavaScript to manipulate arrays and objects (add, delete, sort, search).
14. Fetch and display student information stored in a JSON object on a webpage.
15. Fetch real-time weather data from an open API (e.g., OpenWeatherMap) in JSON format and display temperature, humidity, and conditions dynamically on a webpage.
16. Use jQuery to simplify DOM manipulation (hide, show, fade, slide, toggle).

SEMESTER-IV

COURSE 8: DATA MINING

Theory **Credits: 3** **3 hrs/week**

Course Objectives:

- Provide an understanding of data warehousing concepts, architecture, and OLAP operations for effective storage, modeling, and analysis.
- Develop knowledge of data mining fundamentals, tasks, and preprocessing techniques to prepare data for mining.
- Introduce students to association rule mining algorithms for discovering hidden patterns and relationships in large datasets.
- Enable learners to apply classification techniques (decision trees, Bayesian, nearest neighbor, rule-based) for predictive modeling.
- Equip students with knowledge of clustering paradigms and algorithms (partitioning, hierarchical, density-based, categorical) for data grouping and pattern discovery.

Course Outcomes:

Upon successful completion of the course, the student will be able to:

1. Explain the architecture, schemas, and OLAP operations of data warehousing and distinguish it from traditional database systems.
2. Apply preprocessing techniques (data cleaning, dimensionality reduction, feature selection, transformation, and similarity measures) to prepare raw data for analysis.
3. Implement various association rule mining algorithms (Apriori, Partition, FP-Growth, etc.) to uncover meaningful relationships within large datasets.
4. Build and evaluate classification models using decision tree algorithms (ID3, C4.5, CART), rule-based classifiers, Bayesian classifiers, and nearest-neighbor methods.
5. Analyze and implement clustering techniques such as K-Means, K-Medoid, DBSCAN, BIRCH, and categorical clustering methods (STIRR, ROCK, CACTUS) for grouping and pattern discovery in different types of datasets.

Unit-1: Data Warehousing:

Introduction to Data Ware House, Differences between Database systems and Data Ware House, Data Ware House characteristics, Data Ware House Architecture and its components,

Data Modeling, Schema Design, star and snow-Flake Schema, Fact Constellation, Fact Table, OLAP cube, OLAP Operations.

Unit-2: Data Mining:

What is Data Mining? Data Mining: Definitions, KDD vs Data Mining, Data Mining Tasks, Data Preprocessing- Data Cleaning, Missing Data, Dimensionality Reduction, Feature Subset Selection, Discretization and Binarization, Data Transformation; Measures of similarity and Dissimilarity-Basics.

Issues and Challenges in DM, DM Applications- Case Studies

Unit-3: Association Analysis:

Association Rules: What is an Association Rule?, Methods to Discover Association Rules, A Priori Algorithm, Partition Algorithm, Pincer-Search Algorithm, Dynamic Itemset Counting Algorithms, FP-Tree Growth Algorithm, Generalized Association Rule, Association Rules with Item Constraints

Unit-4: Classification:

Definition, What is Decision Tree?, Tree Construction Principle, Best Split, Splitting Indices, Splitting Criteria, Decision Tree Construction Algorithms: CART, ID3, C4.5, Method for Comparing Classifiers, Rule Based Classifiers, Nearest Neighbor Classifiers, Bayesian Classifiers.

Unit-5: Clustering Techniques:

Clustering Paradigms, Partitioning Algorithms (K-Means), k-Medoid Algorithms, Hierarchical Clustering: DBSCAN, BIRCH, Categorical Clustering Algorithms: STIRR, ROCK, CACTUS

Textbooks:

1. Data Mining Techniques, Arun K Pujari, 3rd Edition, Universities Press
2. Data Mining: Concepts and Techniques, Jiawei Han, Micheline Kamber, Jian Pei, 3rd Edition, Morgan Kaufmann Publishers

Reference Books:

1. K.P. Soman , Shyam Diwakar, V.Ajay ,2006, Insight into Data Mining Theory and Practice, Prentice Hall of India Pvt. Ltd - New Delhi.

2. Introduction to Data Mining, Pang-Ning Tan, Michael Steinbach, Anuj Karpatne, Vipin Kumar, 2nd edition,

Activities:

Outcome: Explain the architecture, schemas, and OLAP operations of data warehousing and distinguish it from traditional database systems.

Activity: Students will design a **conceptual schema** for a data warehouse using **star or snowflake schema** for a given business case (e.g., sales, hospital, or university records). They will also **demonstrate OLAP operations** (roll-up, drill-down, slice, dice) on a sample dataset using a spreadsheet or OLAP tool.

Evaluation Method: Assessment will be based on correctness and completeness of schema design, appropriateness of fact/dimension tables, and accuracy of OLAP operation outputs. A short viva/quiz will be used to test conceptual understanding.

Outcome 2: Apply preprocessing techniques (data cleaning, dimensionality reduction, feature selection, transformation, and similarity measures) to prepare raw data for analysis.

Activity: Students will be given a **real-world noisy dataset** (with missing values, outliers, redundant features). They will perform:

- Data cleaning (handling missing values, outliers)
- Dimensionality reduction (e.g., PCA)
- Feature selection (e.g., filter/wrapper methods)
- Similarity/dissimilarity measure calculations.

Evaluation Method: Evaluation will consider correctness of preprocessing steps, justification of chosen methods, and clarity of intermediate outputs. Students will submit a short report with before/after comparisons.

Outcome 3: Implement various association rule mining algorithms (Apriori, Partition, FP-Growth, etc.) to uncover meaningful relationships within large datasets.

Activity: Using a **transaction dataset** (e.g., market basket data), students will implement Apriori/FP-Growth in **Python (mlxtend library) or Weka**. They will generate frequent itemsets, derive association rules, and interpret them under given **support, confidence, and lift thresholds**.

Evaluation Method: Submissions will be graded on correctness of generated rules, clarity of code/parameters, quality of interpretation, and ability to link discovered rules to practical business insights.

Outcome 4: Build and evaluate classification models using decision tree algorithms (ID3, C4.5, CART), rule-based classifiers, Bayesian classifiers, and nearest-neighbor methods.
Activity: Students will implement at least **two classification algorithms** (e.g., Decision Tree + Naïve Bayes or KNN) on a dataset like **Iris, Titanic, or Student Performance**. They will evaluate models using **confusion matrix, accuracy, precision, recall, and F1-score** and compare results.
Evaluation Method: Assessment will consider correctness of implementation, clarity in performance comparison, and visualization of decision trees/rules. Students must explain why certain models performed better.

Outcome: Analyze and implement clustering techniques such as K-Means, K-Medoid, DBSCAN, BIRCH, and categorical clustering methods (STIRR, ROCK, CACTUS).
Activity: Students will implement at least two clustering algorithms (e.g., K-Means and DBSCAN or BIRCH) on real datasets (e.g., customer segmentation, text data, student groups). They will compare clusters using Silhouette Score, Davies-Bouldin Index, and visualize results using scatter plots/dendrograms.
Evaluation Method: Evaluation will be based on accuracy of clustering implementation, choice of parameters (e.g., k in K-Means, eps in DBSCAN), visualization quality, and clarity in interpretation of clusters.

SEMESTER-IV

COURSE 8: DATA MINING

Practical **Credits: 1** **2 hrs/week**

List of Experiments:

Recommended datasets: **weather.arff**, **iris.arff**, **supermarket.arff**, **vote.arff**, **contact-lenses.arff**, or custom CSV datasets.

1. Load datasets in WEKA and explore data formats (ARFF/CSV)
2. Perform data cleaning and handle missing values using filters
3. Apply normalization and discretization on numeric attributes
4. Reduce data using attribute selection and PCA
5. Summarize and visualize data using statistical tools and class-wise comparison in WEKA.
6. Generate association rules using the Apriori algorithm
7. Apply multilevel association rule mining using hierarchical attributes
8. Apply K-means clustering and interpret the cluster outputs.
9. Perform hierarchical clustering and visualize results using dendrograms.
10. Apply Expectation-Maximization (EM) clustering and analyze cluster summaries.
11. Build a decision tree classifier using J48 and evaluate its performance.
12. Perform Naive Bayes classification and compare with decision tree results.
13. Apply rule-based classification using PART or JRip algorithms.
14. Compare classifiers using confusion matrix, accuracy, and ROC curves.
15. Perform basic text preprocessing and clustering using TF-IDF and K-means.

SEMESTER-IV

COURSE 9: PYTHON FOR DATA ANALYSIS AND VISUALIZATION

Theory **Credits: 3** **3 hrs/week**

Course Objectives:

1. Introduce foundational concepts of NumPy arrays and array operations for efficient numerical computing.
2. Teach key data structures and manipulation techniques using Pandas.
3. Enable students to perform data input/output operations and implement basic data cleaning workflows.
4. Explore string processing methods and feature engineering strategies in Pandas.
5. Guide learners in advanced data wrangling tasks including merging, reshaping, hierarchical indexing and visualization.

Course Outcomes:

1. Demonstrate proficiency in creating and manipulating NumPy arrays for mathematical operations and simulations.
2. Apply Pandas Series and DataFrame operations for structured data handling and analysis.
3. Read, write, and clean diverse data formats using Python tools, addressing missing values and outliers.
4. Implement vectorized string operations and create derived features for enhanced model readiness.
5. Perform complex data wrangling tasks such as merging datasets, reshaping data structures, generating group-level statistics, Visualize the data.

Unit 1. NumPy Essentials:

NumPy ndarray: A Multidimensional Array Object, Creating ndarrays, Data Types for ndarrays, Arithmetic with Arrays, Basic Indexing and Slicing, Boolean Indexing, Fancy Indexing, Transposing Arrays, Swapping Axes, Universal Functions: Element-wise Operations, Basic Mathematical and Statistical Functions, Random Number Generation (basic use)

Unit 2. Pandas Basics and Data Structures:

Series, DataFrame, Index objects, Indexing and Selection, Filtering and Boolean Indexing, Arithmetic and Data Alignment, Sorting and Ranking, Dropping Entries, Handling Duplicate Indexes

Unit 3. Data Input, Output, and Cleaning:

Reading and Writing Data in Text Format (CSV, TXT), Working with JSON, Reading Microsoft Excel Files, Handling Missing Data, Dropping and Filling Missing Values, Replacing Values, Renaming Axis Indexes, Removing Duplicates, Filtering Outliers, Transforming Data Using Mapping or Functions

Unit 4. String Operations and Feature Engineering:

String Methods in pandas, Basic Regular Expressions, Vectorized String Functions, Creating Dummy/Indicator Variables, Permutation and Random Sampling.

Unit 5. Data Wrangling, Reshaping & Visualization:

Merging and Joining Datasets, Concatenating Along an Axis, Combining Data with Overlap, Reshaping with Pivot, Stack, and Unstack, Basic Hierarchical Indexing, Summary Statistics by Group or Level

Introduction to matplotlib: plots, customization, styling, Seaborn for statistical data , visualization, Plotly for interactive charts and dashboards.

Textbooks

1. Python for Data Analysis: Data Wrangling with pandas, NumPy, and Jupyter, Wes McKinney, 3rd Edition, O'Reilly Media, 2022.
2. Python Programming-An Object Oriented Approach, Anita Goel, Universities Press
3. Python for Data Science For Dummies, Yuli Vasiliev, 2nd Edition, Wiley, 2022.

Reference Books

1. Python Data Science Handbook: Essential Tools for Working with Data, Jake VanderPlas, 2nd Edition, O'Reilly, 2022.
2. Introduction to Machine Learning with Python, Andreas Müller & Sarah Guido, O'Reilly Media, Reprint Edition, 2023.

3. Foundations for Analytics with Python: From Non-programmer to Hacker, Clinton Brownley, 2nd Edition, Pearson, 2020.

Activities:

Outcome: Demonstrate proficiency in creating and manipulating NumPy arrays for mathematical operations and simulations.

Activity: Create a NumPy-based simulation:

- Generate a 2D array representing temperature data over time
- Apply mathematical operations (mean, std, element-wise addition)
- Simulate random noise and visualize the effect

Evaluation Method: Code-based assessment (10-point scale):

- Correct use of np.array, np.random, and math functions
- Accuracy of simulation logic
- Output clarity and reproducibility

Outcome: Apply Pandas Series and DataFrame operations for structured data handling and analysis.

Activity: Analyze a CSV dataset (e.g., sales or COVID data):

- Load into a DataFrame
- Perform Series operations (filter, map, value_counts)
- Apply DataFrame methods (groupby, sort, describe)

Evaluation Method: 10-point scale checklist and peer review to verify:

- Proper use of Series vs DataFrame
- Logical data manipulation
- Insightful summary statistics

Outcome: Read, write, and clean diverse data formats using Python tools, addressing missing values and outliers.

Activity: Work with multiple formats:

- Read CSV, Excel, and JSON files
- Identify and handle missing values (dropna, fillna)
- Detect and treat outliers using IQR or Z-score

Evaluation Method: Rubric-based evaluation to check (10-point scale):

- File handling accuracy

- Cleaning completeness
- Outlier detection logic

Outcome: Implement vectorized string operations and create derived features for enhanced model readiness.

Activity: Prepare text data for modelling to:

- Use str methods to clean and standardize strings
- Extract features (e.g., domain from email, length of name)
- Encode categorical variables (e.g., get_dummies, LabelEncoder)

Evaluation Method: Feature report to check (10-point scale):

- Efficiency of vectorized operations
- Relevance of derived features
- Readiness for ML input

Outcome: Perform complex data wrangling tasks such as merging datasets, reshaping data structures, generating group-level statistics and Visualization.

Activity: Integrate and reshape datasets:

- Merge two datasets on a common key
- Reshape using pivot, melt, stack, unstack
- Generate group-level stats (e.g., mean sales per region)
- Visualize the Datasets

Evaluation Method: Before-and-after comparison to validate:

- Accuracy of merge and reshape
- Correct use of aggregation
- Final structure suitability for analysis

SEMESTER-IV

COURSE 9: PYTHON FOR DATA ANALYSIS AND VISUALIZATION

Practical **Credits: 1** **2 hrs/week**

C9P: Python for Data Analysis and Visualization Lab

List of Practicals:

1. Create and Manipulate NumPy ndarrays; Explore Data Types
2. Perform Arithmetic Operations and Element-wise Calculations on Arrays
3. Practice Indexing, Slicing, Boolean, and Fancy Indexing on ndarrays
4. Use Universal Functions and Compute Basic Mathematical/Statistical Functions with NumPy
5. Create and Manipulate Pandas Series and DataFrames
6. Perform Indexing, Selection, Filtering, and Boolean Indexing in Pandas
7. Conduct Arithmetic Operations and Data Alignment in DataFrames
8. Sort, Rank, Drop Entries and Handle Duplicate Indexes in Pandas
9. Read and Write Data in CSV, TXT, JSON, and Excel Formats
10. Handle Missing Data: Detect, Drop, Fill, and Replace Missing Values
11. Rename Axis Indexes, Remove Duplicates, and Filter Outliers
12. Transform Data Using Mapping Functions and Apply String Operations
13. Perform String Operations and Use Regular Expressions on DataFrames
14. Create Dummy Variables and Perform Permutations and Random Sampling
15. Merge, Join, and Concatenate Datasets Using Pandas
16. Reshape Data Using Pivot, Stack, Unstack, and Perform Hierarchical Indexing
17. Compute Summary Statistics Grouped by Levels or Categories
18. Basic Visualizations using Matplotlib

SEMESTER-IV

COURSE 10: DOCUMENT ORIENTED DATABASE

Theory **Credits: 3** **3 hrs/week**

Course Objectives

1. To introduce students to the concepts of NoSQL databases and their significance compared to traditional relational databases.
2. To provide hands-on experience with MongoDB for performing CRUD operations, querying, and advanced data handling.
3. To develop skills in schema design, data modeling, and working with embedded and referenced documents.
4. To explore MongoDB's advanced features such as indexing, aggregation framework, replication, and transactions.
5. To prepare students for real-world applications of MongoDB in scalable, high-performance data-driven applications.

Course Outcomes

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

1. Differentiate between SQL and NoSQL databases, and explain the architecture and features of MongoDB.
2. Perform CRUD operations and construct queries using MongoDB Query Language (MQL).
3. Apply schema design strategies and use appropriate data modeling techniques for different application scenarios.
4. Utilize advanced features like indexing, aggregation, GridFS, and transactions to optimize data handling.
5. Implement replication concepts and ensure high availability, fault tolerance, and scalability in MongoDB-based applications.

Unit 1. Introduction to NoSQL & Fundamentals of MongoDB:

What is NoSQL DB? History & evolution of NoSQL, Features of NoSQL databases, CAP theorem & BASE properties, Types of NoSQL (Key-Value, Document, Column, Graph), Difference between RDBMS & NoSQL, Why and when to use NoSQL?, NoSQL Database misconceptions, Benefits & real-world use cases of NoSQL, Comparison of popular NoSQL systems (Redis, Cassandra, CouchDB, Neo4j), Introduction to JSON & BSON

Installation & Setup: Installing MongoDB locally and using Atlas (MongoDB's cloud service), connecting via Mongo shell or GUI.

Unit 2. MongoDB Architecture, Data Modeling and Basics:

MongoDB Architecture: Database, Collection, Document concepts, BSON format, Advantages of MongoDB over RDBMS,

MongoDB Datatypes (String, Number, Date, Boolean, Array, ObjectId, Embedded Documents, Null)

Data Modeling in MongoDB: Schema design strategies, Embedded vs Referenced documents

Database & Collection Management: Create & Drop Database, Create & Drop Collection

Unit 3. CRUD Operations and Querying in MongoDB:

CRUD Operations: Insert Documents (insertOne, insertMany), Query Documents (find, operators, conditions), Update Documents (updateOne, updateMany, replaceOne), Delete Documents (deleteOne, deleteMany)

Query operators ($gt, $lt, $in, $nin, $and, $or, $not), Regular expression queries, Bulk operations

Working with Arrays.

Unit 4. Data Modelling and Aggregation:

Data Modelling and Aggregation: Data Models: Introduction to embedded vs normalized models, advantages and trade-offs

Embedded Data Models: Use cases, benefits, and limitations

Normalized Data Models: References between documents, when to normalize data Relationships Between Documents, Data Model Using an Embedded Document, Data Model Using Document References

Aggregation Basics: Introduction to MongoDB Aggregation Framework, simple pipelines and operators

Unit 5. Advanced Query Processing and Optimization in MongoDB:

Query Optimization & Operations: Projection, Limiting & Skipping Records, Sorting Records

Indexing in MongoDB (single field, compound, multikey, text index), Aggregation Framework (pipelines, stages, operators), Replication Concepts: Replica sets, failover, consistency

Textbooks:

1. MongoDB: The Definitive Guide, Shannon Bradshaw, Eoin Brazil, Kristina Chodorow, 3rd Edition, O'Reilly Media, 2019.
2. MongoDB Recipes: With Data Modeling and Query Building Strategies, Subhashini Chellappan, Dharanitharan Ganesan , Apress

Reference Books:

1. MongoDB in Action, Kyle Banker, 2nd Edition, Manning Publications, 2016.
2. MongoDB Applied Design Patterns, Steve Francia, O'Reilly Media, 2013.
3. MongoDB for Developers, Rick Copeland, O'Reilly Media, 2013.

Web Resources:

1. Official MongoDB Documentation: https://www.mongodb.com/docs/
2. MongoDB University Free Courses: https://learn.mongodb.com/
3. W3Schools, Tutorialspoint, geeksforgeeks

Activities:

Outcome: Differentiate between SQL and NoSQL databases, and explain the architecture and features of MongoDB

Activity: Students create a comparative chart (SQL vs NoSQL) and draw MongoDB architecture diagram.

Evaluation Method: Assess based on accuracy, clarity of explanation, and presentation.

Outcome: Perform CRUD operations and construct queries using MongoDB Query Language (MQL)

Activity: Implement CRUD operations and write 5 different queries on a sample dataset (e.g., Library or E-commerce).

Evaluation Method: Marks for correct execution, query correctness, and output validation.

Outcome: Apply schema design strategies and use appropriate data modeling techniques for different application scenarios

Activity: Design schema for a university management system (students, courses, faculty) using MongoDB data modeling techniques.

Evaluation Method: Evaluate on schema correctness, use of embedding/referencing, and justification of design.

Outcome: Utilize advanced features like indexing, aggregation, GridFS, and transactions to optimize data handling

Activity: Create indexes and implement an aggregation pipeline to generate sales report from a dataset.

Evaluation Method: Marks for index implementation, aggregation correctness, and efficiency of results.

Outcome: Implement replication concepts and ensure high availability, fault tolerance, and scalability in MongoDB-based applications

Activity: Configure a replica set with primary and secondary nodes, and demonstrate failover.

Evaluation Method: Assess on setup correctness, successful failover demonstration, and report/documentation.

SEMESTER-IV

COURSE 10: DOCUMENT ORIENTED DATABASE

Practical **Credits: 1** **2 hrs/week**

1. Installation and setup of MongoDB, connecting to Mongo Shell and Compass.
2. Creating and using databases, creating collections, inserting documents.
3. Basic queries using find(), filtering with comparison operators.
4. Using logical operators ($and, $or, $not, $nor) for complex queries.
5. Updating documents with $set, $unset, $inc, $rename.
6. Deleting documents using deleteOne() and deleteMany().
7. Using projection to display selective fields.
8. Sorting documents, limiting output, skipping records.
9. Designing an Embedded Data Model for a student-course enrollment system.
10. Designing a Normalized Data Model using document references.
11. Modeling relationships: One-to-One, One-to-Many, Many-to-Many in MongoDB.
12. Implementing schema validation using JSON Schema in MongoDB.
13. Creating and testing single-field and compound indexes.
14. Using text search and multikey indexes.
15. Building aggregation pipelines with $match, $group, $project, $sort.
16. Advanced aggregation operators: $lookup, $unwind, $bucket.
17. Configuring and testing replication with a replica set (minimum 3 nodes).
18. Storing and retrieving large files using GridFS.
19. Using MongoDB Transactions for multi-document consistency.
20. Case Study: Developing a mini-application (e.g., Library Management / E-commerce Cart) using MongoDB CRUD, Aggregation, Indexing, and Replication.

SEMESTER-V

COURSE 11: BUSINESS INTELLIGENCE TOOLS

Theory **Credits: 3** **3 hrs/week**

Course Objectives:

1. Introduce foundational concepts of Business Intelligence (BI) and Decision Support Systems (DSS), including their scope, evolution, and organizational relevance.
2. Familiarize students with leading BI tools such as Power BI and Tableau, highlighting their ecosystems, interfaces, and comparative strengths.
3. Develop skills in data preparation and transformation, using Power Query and Tableau's data connection features to clean and model datasets.
4. Enable effective data visualization and storytelling, leveraging charts, dashboards, and advanced features to communicate insights.
5. Equip learners with data modeling techniques, including dimensional modeling, relationships, joins, and governance principles for robust BI solutions.

Course Outcomes:

At the end of the course, the students will be able to:

1. Differentiate between BI, Data Analytics, and Data Science, and explain the BI lifecycle and its applications across functional domains.
2. Use Power BI and Tableau to prepare, transform, and visualize data, applying basic DAX functions and calculated fields for analysis.
3. Design and implement dimensional data models, including star and snowflake schemas, and apply relationships and joins in BI tools.
4. Create interactive dashboards and visualizations, incorporating parameters, slicers, filters, and drilldowns to enhance decision-making.
5. Build and publish complete BI dashboards, and effectively communicate business insights through storytelling and visualization best practices.

Unit-I: Introduction to Business Intelligence and Decision Support Systems

Business Intelligence: Definition, Scope, and Evolution, Business IntelligenceI vs. Data Analytics vs. Data Science, BI Lifecycle, **Applications of BI in Functional Domains:** Finance, HR, Marketing, Retail, Education, Healthcare, etc., BI Maturity Models &

Organizational Readiness to BI adoption Decision Support Systems (DSS): Concepts, Components, and Architecture.

BI Tools Overview: Power BI, Tableau, and other tools, Comparison and suitability of BI Tools.

Case Study: Retail Chain's BI Strategy to Optimize Inventory

Unit-II: Data Preparation and Visualization with Power BI

Introduction to Power BI, Power BI Ecosystem: Desktop, Service, Mobile; Power BI Interface; Data Sources: Excel, CSV, SQL Server, Web APIs, Power Query: Data Preparation, Cleaning & Transformation - Connect, transform, and model a dataset; Basic DAX Functions: SUM, COUNT, AVERAGE, CALCULATE, IF; Creating Simple Visualizations: Charts, Tables, Cards; Sharing Reports via Power BI Service.

Case Study: Student performance analysis in Higher Education , Analyze Finance Dataset

Unit-III: Data Preparation, Visualization and Storytelling with Tableau

Introduction to Tableau; Characteristics of Tableau; Tableau Architecture and components - Tableau Public, Desktop, Reader, Online, Server; Tableau Interface: Shelves, Marks Card, Views; Tableau extensions, Data Connection and Preparation: Cleaning, Pivoting, Filtering; Calculated Fields and LOD Expressions; Basic Visualizations: Bar, Line, Tree, Geo Maps, Scatter Plots; Storytelling with Tableau, Creating a Tableau story.

Case Study: HR Analytics

Unit-IV: Data Modeling and Relationships in BI Tools

Dimensional Modeling: Dimension, Dimension table, fact, fact table, schema, Star and Snowflake Schemas

Power BI: Relationships, Cardinality, Cross-filtering; Tableau: Joins (Inner, Left, Full), Blending; Data Governance: Metadata, Hierarchies, Quality; Data Model Design Best Practices - Design and implement data model in Power BI and Tableau; HR or Retail data model design and insights

Case Study: Retail BI for Sales Optimization

Unit-V: Dashboard Design and Business Insights

Introduction to Dashboard, when to use dashboards, Dashboard components, Principles of Effective Visualization & Dashboarding, Advanced Visualizations: Parameters, Slicers, Filters, Drilldowns, Graphs and Maps, Dashboard Design: Layout, Alignment, Accessibility

Publishing Dashboards: Power BI Service, Tableau Public; Storytelling and Insight Communication, Build a complete BI dashboard using either tool.

Case Study: Business decision-making scenario (e.g., Sales Forecasting, Budgeting)

Text Books:

1. Decision Support and Business Intelligence Systems (9th ed.). Turban, E., Sharda, R., & Delen, D. (2014), Pearson Education.
2. Learning Tableau 2022: Create effective data visualizations, build interactive dashboards, and transform your data into insights (6th Edition), Milligan, J. N. (2022)., Packt Publishing.
3. Expert Data Modeling with Power BI: Enrich and optimize your data models for reporting and business needs (2nd Edition). Bakhshi, S. (2023), Packt Publishing.

Reference Books:

1. Visual Analytics with Tableau, Loth, A. (2019), Addison-Wesley Professional.
2. The Definitive Guide to DAX: Business intelligence for Microsoft Power BI, SQL Server Analysis Services, and Excel (2nd Edition), Russo, M., & Ferrari, A. (2020), Microsoft Press.

Online Resources:

1. Decision Support Systems Courses – Class Central
2. Advanced Business Decision Support Systems – NPTEL (IIT Kanpur)
3. Power BI Learning Paths – Microsoft Learn
4. Power BI Courses – Coursera
5. Tableau Learning Hub – Official Site
6. Free Tableau Course – Simplilearn
7. Dashboard Design Concepts – DataCamp
8. Business Intelligence with Power BI – Swayam

Activities:

1. Differentiate BI, Data Analytics, and Data Science + BI Lifecycle

Activity: Case Study Analysis: Provide students with a business scenario and ask them to identify how BI, analytics, and data science each contribute. Then, have them map the BI lifecycle stages to that scenario and explain its impact across departments (e.g., finance, marketing).

Evaluation Method: Evaluate on a 10-point scale based on Conceptual differentiation clarity (30%), BI lifecycle accuracy (30%), Application to domains (20%) and Written summary or presentation quality (20%)

2. Prepare, Transform & Visualize Data in Power BI/Tableau (with DAX & Calculated Fields)

Activity: Hands-on Data Challenge: Supply students with a raw dataset (e.g. sales or customer data). Task them with cleaning, transforming, and building visuals in Power BI/Tableau using basic DAX and calculated fields (e.g. profit margins, year-over-year growth).

Evaluation Method: Evaluate for 10-points based on Effective transformation workflow (25%), Correct DAX/formula usage (25%), Visualization relevance & clarity (25%) and Documentation of process (25%)

3. Dimensional Modelling with Star/Snowflake Schemas + Joins/Relationships

Activity: Model Building Lab: Students design a star or snowflake schema based on a retail or HR database. Then, implement the schema in Power BI or Tableau and create relationships between tables to ensure correct joins and data flow.

Evaluation Method: Evaluate on a 10-point scale based on Schema design accuracy (30%), Appropriate schema selection (star vs. snowflake) (20%), Implementation of joins/relationships (30%) and Functional model validation (20%)

4. Create Interactive Dashboards with Advanced Features

Activity: Dashboard Design Workshop: Students build an interactive dashboard using parameters, slicers, filters, and drilldowns to simulate real-time decision-making (e.g., tracking regional product sales or employee performance across departments).

Evaluation Method: Evaluate on a 10-point scale on the basis of Integration of interactive features (30%), Usability and navigation experience (30%), Data-driven insights extracted (20%) and Design polish and layout consistency (20%).

5. Build and Publish BI Dashboards + Business Storytelling

Activity: BI Capstone Project: Students design and publish a complete dashboard solving a real or simulated business problem (e.g. customer churn, supply chain bottlenecks). Include visual storytelling techniques-titles, annotations, color themes, and narratives.

Evaluation Method: Evaluate on a 10-point scale on the basis of Clarity of business insights communicated (30%), Storytelling elements (25%), Dashboard completeness and polish (25%) and Peer or instructor presentation (20%)

SEMESTER-V

COURSE 11: BUSINESS INTELLIGENCE TOOLS

Practical **Credits: 1** **2 hrs/week**

List of Experiments:

1. Exploring BI Tools - Power BI vs Tableau.
2. Create a simple retail dashboard using both tools.
3. **Connecting to Different Data Sources in Power BI**
    a. Learn to connect Excel, CSV, and Web data.
    b. Load a dataset from different formats.
    c. Demonstrate dataset loading and view schema.
4. **Data Cleaning and Transformation using Power Query**
    a. Apply Power Query for cleaning.
    b. Remove duplicates, fill nulls, filter rows.
    c. Submit Power Query steps and cleaned dataset.
5. **Prepare and clean a higher education student performance dataset using Power Query and visualize key academic metrics.**
    a. Connect to a dataset and perform data cleaning, transformation, and reshaping using Power Query.
    b. Visualize academic performance indicators such as GPA trends, pass rates, or subject-wise scores.
    c. Upload file containing the cleaned dataset and relevant academic metric visualizations.
6. **Implementing DAX Functions**
    a. Use DAX to perform basic calculations.
    b. Create calculated columns with SUM, AVERAGE, COUNT, CALCULATE, IF.
    c. Submit DAX expressions with visual output.
7. **Creating Basic Visualizations in Power BI**
    a. Develop simple charts and cards.
    b. Use sales or HR data to create bar, pie charts, and KPIs.
    c. Display Dashboard showing at least 3 chart types.
8. **Tableau Basics and Connecting to Data**
    a. Connect and explore data in Tableau Public.

b. Load a student performance or any other dataset and preview data.
c. Upload dataset and share working link.

9. **Visualize employee turnover patterns using Tableau and apply LOD expressions to uncover retention drivers.**
    a. Import and clean HR data for turnover analysis in Tableau.
    b. Apply LOD expressions to identify patterns across departments, roles, and tenure.
    c. Upload cleaned dataset and interactive visualizations highlighting retention insights.
10. **Data Cleaning, Pivoting & Filtering in Tableau**
    a. Prepare data inside Tableau.
    b. Pivot columns, apply filters, and rename headers.
    c. Upload cleaned worksheet.
11. **Creating Visualizations in Tableau**
    a. Use Marks Card, Shelves, and Views for visualization.
    b. Build bar chart, map view, scatter plot.
    c. One dashboard with 3 visuals.
12. **Creating a Tableau Story - (Eg HR or Student Performance or any other)**
    a. Build a narrative with visualizations.
    b. Combine charts into a Tableau story with captions.
    c. Publish and present story with link.
13. **Designing Data Models in Power BI**
    a. Create dimensional models (Star/Snowflake schema).
    b. Use a retail dataset with multiple tables and define relationships.
    c. Submit relationship diagram and schema explanation.
14. **Joins and Blending in Tableau**
    a. Implement data joins and blending techniques.
    b. Combine multiple data tables using inner/left/full joins.
    c. Demonstrate join types with visuals.
15. **Dashboard with Drill-downs, Filters, and Slicers**
    a. Use interactivity in dashboards.
    b. Build a multi-level dashboard in Power BI with drill-through.
    c. Submit .pbix file and link to published dashboard.

SEMESTER-V

COURSE 12 A: MACHINE LEARNING

Theory **Credits: 3** **3 hrs/week**

Course Objectives:

1. Understand fundamental concepts, types, and applications of machine learning.
2. Develop, evaluate, and optimize machine learning models through preprocessing, training, and feature engineering techniques.
3. Apply supervised and unsupervised learning algorithms to real-world problems using appropriate tools and methods.

Course Outcomes:

Upon successful completion of this course, students will be able to:

1. Describe various machine learning paradigms, data types, and the overall structure of a machine learning pipeline.
2. Perform data preprocessing, feature engineering, and evaluate models using appropriate metrics.
3. Implement and analyze supervised learning algorithms for regression and classification tasks.
4. Apply unsupervised learning techniques for clustering and identify suitable machine learning approaches for specific application domains

Unit 1. Introduction to Machine Learning:

Introduction to Machine Learning: Types of human learning, What is machine learning?, Types of machine learning: supervised, unsupervised, semi-supervised and reinforcement learning, machine learning activities, applications of machine learning. Types of data in machine learning, Structure of data

Unit 2. Model Preparation, Evaluation and feature engineering:

Data pre-processing, Model selection and training (for supervised learning), Model representation and interpretability, Evaluating machine learning algorithms and performance enhancement of models. What is feature engineering?, Feature transformation, Feature subset selection. Principal component analysis.

Unit 3. Supervised Learning-Regression:

Regression: Introduction of regression, Regression algorithms: Simple linear regression, Multiple linear regression, Polynomial regression model, Logistic regression, Maximum likelihood estimation.

Unit 4. Supervised Learning- Classification:

Introduction of supervised learning, Classification model and learning steps, Classification algorithms: Naïve Bayes classifier, k-Nearest Neighbour (kNN), Decision tree, Support vector machines, Random Forest.

Unit 5. Unsupervised Learning:

Introduction of unsupervised learning, Unsupervised vs supervised learning, Application of unsupervised learning, Clustering and its types, Partitioning method: k-Means and KMedoids, Hierarchical clustering, Density-based methods - DBSCAN.

Case-study of ML applications: Image recognition, speech recognition, Email spam filtering, Online fraud detection and other.

Textbooks:

1. Introduction to Machine Learning, Ethem Alpaydin, MIT Press, Fourth Edition, 2020.
2. Machine Learning: Theory and Practice, M N Murthy, V.S Ananthanarayana, Universities press
3. Machine Learning, S. Sridhar, M. Vijayalakshmi, Second Edition, Oxford University Press

Reference Books:

1. Machine Learning: An Algorithmic Perspective, Second Edition, Stephen Marsland, CRC Press, 2014
2. Machine Learning, Tom Mitchell, McGraw Hill, 3rd Edition.
3. Python Machine Learning, Sebastain Raschka, Vahid Mirjalili , Packt publishing 3rd Edition, 2019.

Activities:

Outcome: Describe various machine learning paradigms, data types, and the overall structure of a machine learning pipeline.

Activity: Prepare a detailed comparative report/chart explaining supervised, unsupervised, and reinforcement learning paradigms, data types, and step-by-step machine learning pipeline stages.

Evaluation Method: Rubric-based assessment evaluating completeness, clarity, correctness, and presentation quality - scored on a 10-point scale.

Outcome: Perform data preprocessing, feature engineering, and evaluate models using appropriate metrics.
Activity: Conduct a hands-on lab exercise using a real dataset to perform data cleaning, normalization, feature extraction/selection, and evaluate model performance using metrics like accuracy, precision, recall, and F1-score.
Evaluation Method: Practical assessment including code correctness, applied techniques, and interpretation of evaluation metrics; assessed with a rubric out of 10.

Outcome: Implement and analyze supervised learning algorithms for regression and classification tasks.
Activity: Implement at least two supervised learning algorithms (e.g., Linear Regression and Decision Trees) to solve prediction tasks, followed by comparative analysis of their performance on test datasets.
Evaluation Method: Code and report evaluation focusing on implementation accuracy, performance comparison, and analysis depth; scored on a 10-point rubric.

Outcome: Apply unsupervised learning techniques for clustering and identify suitable machine learning approaches for specific application domains.
Activity: Perform clustering (e.g., K-Means, Hierarchical) on a given dataset and prepare a case study selecting and justifying machine learning methods suited for different application scenarios.
Evaluation Method: Lab practical combined with a written case study; assessed for correct algorithm application, cluster interpretation, and justification of approach - evaluated on a 10-point rubric.

SEMESTER-V

COURSE 12 A: MACHINE LEARNING

Practical **Credits: 1** **2 hrs/week**

Lab Experiments:

1. Write a python program to import and export data using Pandas library functions.
2. Demonstrate various data pre-processing techniques for a given dataset
3. Implement Dimensionality reduction using the Principal Component Analysis (PCA) method.
4. Write a Python program to demonstrate various Data Visualization Techniques.
5. Implement MLE on a Dataset
6. Implement Simple and Multiple Linear Regression Models.
7. Develop Logistic Regression Model for a given dataset.
8. Develop Decision Tree Classification model for a given dataset and use it to classify a new sample.
9. Implement Naïve Bayes Classification in Python.
10. Develop K-Means for a Given Dataset
11. Build KNN Classification model for a given dataset.
12. Develop DBSCAN on a given Dataset

SEMESTER-V

COURSE 12 B: BIG DATA TECHNOLOGIES

Theory **Credits: 3** **3 hrs/week**

Course Objectives:

1. Introduce students to the concepts, characteristics, and challenges of Big Data.
2. Familiarize students with the Hadoop ecosystem and its core components (HDFS, YARN, MapReduce).
3. Develop practical knowledge of distributed storage and parallel processing in Hadoop.
4. Provide hands-on exposure to data ingestion tools (Sqoop, Flume) and serialization techniques.
5. Enable students to explore NoSQL databases (HBase), coordination services (ZooKeeper), and Hadoop–Spark integration for large-scale data analysis.

Course Outcomes:

At the end of this course, students will be able to:

1. Explain Big Data concepts and challenges along with the role of the Hadoop ecosystem.
2. Demonstrate understanding of HDFS and YARN architectures and their functions in distributed data management.
3. Apply MapReduce and high-level tools (Hive, Pig, Spark) to process and analyze large datasets.
4. Design and implement data ingestion workflows using Sqoop, Flume, and serialization formats like Avro and Parquet.
5. Utilize NoSQL databases and ecosystem enhancements such as HBase, ZooKeeper, and Hadoop–Spark integration for scalable big data solutions.

Unit 1. Foundations of Big Data & Hadoop Ecosystem

Introduction to Big Data: characteristics (volume, variety, velocity, veracity, value), Hadoop Ecosystem Overview: HDFS, MapReduce, YARN, Hadoop Common, Hadoop architecture and use cases

Unit 2. Hadoop Distributed File System (HDFS) & YARN:

Deep dive into HDFS architecture: blocks, NameNode, DataNodes, HDFS file operations, fault tolerance, replication

YARN architecture: ResourceManager, NodeManager, application scheduling

Unit 3. MapReduce & High-Level Tools

MapReduce programming model: map, shuffle, reduce phases, Writing MapReduce applications in Hadoop

High-level abstractions: Hive, Pig, Crunch, and introduction to Spark integration

Unit 4. Data Ingestion & Serialization

Data ingestion pipelines: Sqoop (for RDBMS), Flume (streaming), Data formats & serialization: Avro, Parquet, SequenceFile, Practical ingestion workflows-batch and streaming

Unit 5. NoSQL & Ecosystem Enhancements

Overview of NoSQL within Hadoop ecosystem: HBase, Configuration and usage of ZooKeeper for coordination, Hadoop integration with Spark for data processing

Textbooks

1. Hadoop: The Definitive Guide, Tom White, 4th Edition, O'Reilly
   Free Resource available at piazza-resources.s3.amazonaws.comO'Reilly Media
2. Learning Spark, 2nd Edition, Jules S. Damji, Brooke Wenig, Tathagata Das, Denny Lee, O'Reilly

Reference Books

3. BIG DATA, Black Book TM, DreamTech Press, 2016 Edition.
4. BIG DATA and ANALYTICS, Seema Acharya, SubhasniChellappan , Wiley publications, 2016

Activities:

Outcome: Explain Big Data concepts and Hadoop ecosystem

Activity: Short seminar / presentation on Big Data applications

Evaluation Method: Oral presentation + concept quiz

Outcome: Demonstrate HDFS and YARN architectures

Activity: Hands-on lab to configure HDFS & analyze NameNode/DataNode logs

Evaluation Method: Lab performance + viva

Outcome: Apply MapReduce and high-level tools

Activity: Mini-project implementing MapReduce job and Hive queries

Evaluation Method: Project report + execution demo

Outcome: Design and implement data ingestion workflows

Activity: Lab task to ingest data using Sqoop/Flume and serialize with Avro/Parquet

Evaluation Method: Lab record + output validation

Outcome: Utilize NoSQL and ecosystem enhancements

Activity: Case study on HBase–Spark integration with example dataset

Evaluation Method: Case study report + written test

SEMESTER-V

COURSE 12 B: BIG DATA TECHNOLOGIES

Practical **Credits: 1** **2 hrs/week**

1. Installation & setup of Hadoop single-node cluster
2. Explore Hadoop directory structure and basic commands (hadoop fs operations)
3. Demonstration of Hadoop architecture components (HDFS, YARN, MapReduce) using sample logs
4. Store and retrieve large files from HDFS (block distribution, replication factor demo)
5. Simulate NameNode/DataNode failure and observe fault tolerance & recovery
6. Configure YARN and run sample applications, observe ResourceManager and NodeManager roles
7. Write a simple MapReduce program for word count
8. Develop a MapReduce job for inverted index creation
9. Perform data analysis using Pig Latin scripts
10. Execute Hive queries for structured data analysis (tables, partitions)
11. Import data from RDBMS into Hadoop using Sqoop
12. Capture and store log/streaming data using Flume
13. Serialize and store datasets in Avro and Parquet formats
14. Build an end-to-end ingestion workflow combining batch (Sqoop) and streaming (Flume)
15. Create and manage tables in HBase (CRUD operations)
16. Demonstrate coordination with ZooKeeper
17. Process HBase datasets using Spark integration with Hadoop

SEMESTER-V

COURSE 13 A: ARTIFICIAL INTELLIGENCE

Theory **Credits: 3** **3 hrs/week**

Course Objectives

1. Understand the fundamental concepts, history, types, and applications of Artificial Intelligence.
2. Develop problem-solving skills using state-space representations and search strategies for AI applications.
3. Apply informed and advanced search techniques including heuristics, local search, genetic algorithms, and constraint satisfaction problems.
4. Learn knowledge representation methods and reasoning techniques using propositional and first-order logic for intelligent agents.
5. Explore expert systems, probabilistic reasoning, fuzzy logic, and emerging AI technologies including NLP, robotics, and ethical considerations.

Course Outcomes

At the end of the course, students will be able to:

1. Explain the concepts, scope, types of AI, and structure of intelligent agents and their environments.
2. Formulate problems using state-space representation and solve them using uninformed search strategies like BFS, DFS, and Uniform Cost Search.
3. Apply informed search, local search, genetic algorithms, and constraint satisfaction techniques to solve complex AI problems.
4. Represent knowledge using propositional and first-order logic, perform reasoning, and design knowledge-based agents.
5. Design simple expert systems, implement probabilistic reasoning and fuzzy logic, and demonstrate awareness of emerging AI technologies and ethical issues.

Unit 1. Introduction to Artificial Intelligence & Intelligent Agents:

Definition and scope of AI, history and evolution of AI, Turing Test, Applications of AI in real world.

Types of AI: Weak AI vs Strong AI, Narrow AI vs General AI. Intelligent Agents: Structure of agents, Rationality, Agent types. Environments: Deterministic vs Stochastic, Static vs Dynamic, Discrete vs Continuous. PEAS representation (Performance measure, Environment, Actuators, Sensors).

Unit 2. Problem Solving: State Space & Uninformed Search:

State space representation: Components (State, Actions, Goal test, Path cost). Problem formulation and examples (8-puzzle, water jug, vacuum cleaner world).

Uninformed search strategies: Breadth First Search (BFS), Depth First Search (DFS), Uniform Cost Search- Properties: Completeness, Optimality, Time & Space complexity.

Unit 3. Informed & Advanced Search Strategies:

Informed search strategies: Heuristics (concept, admissibility, consistency), Greedy Best First Search, A* Algorithm

Local Search: Hill Climbing, Simulated Annealing. Genetic Algorithms

Constraint Satisfaction Problems (CSP): Definition, Backtracking search.

Unit 4. Knowledge Representation & Reasoning:

Knowledge Representation: Issues, Approaches.

Propositional Logic: Syntax, Semantics, Truth tables, Inference rules.

First Order Logic (FOL): Syntax, Semantics, Quantifiers, Substitution, Unification.

Inference in Logic: Forward Chaining, Backward Chaining, Resolution.

Knowledge-based agents.

Unit 5. Expert Systems, Probabilistic & Emerging AI:

Expert Systems: Architecture, Knowledge base, Inference engine, Explanation facility.

Probabilistic Reasoning: Bayes Theorem, Bayesian Belief Networks (concepts & examples)

Fuzzy Logic and Uncertainty handling.

Emerging topics: NLP basics, Robotics, AI Ethics & societal impact.

Textbooks:

1. Artificial Intelligence: A Modern Approach, Stuart Russell & Peter Norvig, 4th Edition, Pearson
2. Artificial Intelligence**,** Elaine Rich & Kevin Knight, 3rd Edition, McGraw-Hill

Reference Books:

1. The Art of Prolog, Leon Sterling & Ehud Shapiro, MIT Press
2. Learn Prolog Now, Patrick Blackburn, Johan Bos, Kristina Striegnitz (Free online book)

Activities:

Outcome: Explain the concepts, scope, types of AI, and structure of intelligent agents and their environments.

Activity: Divide the class into small groups and ask each group to create a poster or digital infographic comparing Weak AI vs Strong AI, and Narrow AI vs General AI, including real-world examples of intelligent agents and their environments.

Evaluation Method: Peer and instructor review rubric (10 points) assessing correctness of concepts, clarity of illustrations, examples provided, and creativity in presentation.

Outcome: Formulate problems using state-space representation and solve them using uninformed search strategies like BFS, DFS, and Uniform Cost Search.

Activity: Provide students with a simple problem like the 8-puzzle or water jug problem. Students model the state space, draw the state tree, and manually perform BFS and DFS traversal to reach the goal state.

Evaluation Method: Submission of state-space diagrams and solution steps, graded on accuracy, completeness, and correct application of search strategies.

Outcome 3: Apply informed search, local search, genetic algorithms, and constraint satisfaction techniques to solve complex AI problems.

Activity: Conduct a mini-coding session in Python where students implement A* search for a maze-solving problem or hill climbing for a simple optimization task. They can also experiment with a genetic algorithm for a simple fitness function problem.

Evaluation Method: Code demonstration and correctness check, including explanation of heuristics, search path, and results.

Outcome: Represent knowledge using propositional and first-order logic, perform reasoning, and design knowledge-based agents.

Activity: Give students a logic puzzle (e.g., Sudoku rules or “who owns which pet” problem). Students write propositional and/or FOL statements, draw inference chains, and perform forward/backward chaining to solve the puzzle.

Evaluation Method: Solution submission and in-class demonstration of inference process, graded on logical correctness, clarity of reasoning, and completeness.

Outcome: Design simple expert systems, implement probabilistic reasoning and fuzzy logic, and demonstrate awareness of emerging AI technologies and ethical issues.
Activity: Ask students to design a rule-based expert system for a small domain (e.g., medical symptom checker), represent a simple Bayesian network for probabilistic reasoning, and prepare a short presentation on AI ethics or NLP applications.
Evaluation Method: Project-based evaluation including expert system rules, Bayesian reasoning correctness, and quality of presentation on emerging AI topics.

SEMESTER-V

COURSE 13 A: ARTIFICIAL INTELLIGENCE

Practical **Credits: 1** **2 hrs/week**

SWI-Prolog environment for practice without installation.

1. Write Prolog facts for a family tree.
    a. Define rules for ancestor/2, sibling/2, cousin/2.
    1. Query for ancestors, descendants, and siblings.
2. Implement member/2, append/3, reverse/2, length/2.
3. Write a predicate to find the maximum element of a list.
4. Flatten a nested list into a single-level list.
5. Write Prolog rules to calculate factorial and Fibonacci numbers.
6. Implement GCD of two numbers using recursion.
7. Demonstrate the use of cut (!) and fail predicates.
8. Represent a simple graph using edge/2 predicates.
9. Write a recursive DFS to find a path between two nodes.
10. Implement BFS to find a path between two nodes in a graph.
11. Compare DFS and BFS by finding path lengths.
12. Solve the 8-puzzle or grid problem using Greedy Best-First search and A*.
13. Represent a map with regions and adjacency constraints.
    1. Assign colors to regions using backtracking.
    2. Ensure no adjacent regions share the same color.
14. Place N queens on an N×N chessboard such that no two queens threaten each other.
    1. Generate all possible solutions using backtracking.
15. Encode facts and rules using propositional and first-order logic.
16. Implement forward and backward chaining for simple queries.
17. Build a rule-based expert system (e.g., medical diagnosis or plant disease).
    1. Include knowledge base, inference engine, and explanation facility.
    2. Test the system with sample queries.
18. Write a DCG grammar for simple English sentences.
    1. Parse sentences to generate a syntax tree.
19. Implement a simple deterministic Naïve Bayes calculation for categorical data.

SEMESTER-V

COURSE 13 B: CLOUD COMPUTING FOR DATA SCIENCE

Theory **Credits: 3** **3 hrs/week**

Course Objectives

1. Introduce the fundamentals of cloud computing and its role in data science.
2. Provide understanding of virtualization, service, and deployment models.
3. Familiarize students with cloud storage, data management, and databases.
4. Expose students to cloud-based big data and machine learning platforms.
5. Train students in building, deploying, and monitoring ML pipelines on the cloud.

Course Outcomes

At the end of this course, students will be able to:

1. Explain cloud computing concepts including service models, deployment models, and virtualization.
2. Demonstrate cloud storage and database services for managing large-scale data.
3. Apply cloud-based platforms to run data science and machine learning workflows.
4. Build and deploy ML models on cloud services using AutoML and managed ML platforms.
5. Evaluate and monitor cloud-deployed solutions with respect to scalability, performance, and cost.

Unit 1. Introduction to Cloud Computing

Definition & Evolution of Cloud Computing, Service-Oriented Architecture (SOA) & Web Services, Utility & Grid Computing concepts, Characteristics of Cloud Computing

Cloud Computing Architecture: Front-end, Back-end, Networking, Delivery Models

Cloud Service Models: SaaS, PaaS, IaaS, Continuous Delivery using PaaS

Unit 2. Virtualization & Deployment Models

Concept & importance of Virtualization, Types of Virtualizations: Application, Network, Desktop, Storage, Server, Data Virtualization

Cloud Deployment Models: Public, Private, Community, Hybrid

Role of Cloud Computing in Data Science, Advantages of Cloud in Machine Learning

Unit 3. Cloud Storage & Data Management

Cloud Storage: Introduction, Benefits, Use Cases (Backup, Archiving, DR, Content Delivery)

Cloud Storage Systems: Block-based, File-based, Object-based storages

Key-Value Databases: Features & limitations

Batch vs. Streaming data for ML pipelines

Cloud Data Warehouses: AWS Redshift, Google BigQuery

Unit 4. Cloud Platforms for Data Science & ML

Machine Learning in the Cloud: Benefits & Limitations, Cloud-based ML Services: AIaaS, GPUaaS

Managed ML Platforms: Overview & advantages, Cloud ML Platforms: AWS SageMaker, Azure ML Studio, Google Cloud AutoML

Unit 5. Training & Deployment of ML on Cloud

Factors for selecting Cloud ML Platforms: ETL/ELT pipeline support, Scale-up/out training, ML frameworks, Pre-tuned services

Steps for Training ML Models in Cloud: Data source identification, Feature engineering, Training, Validation, Deployment, Monitoring, Monitoring & improving cloud-deployed ML models

Case studies & industry applications

Text / Reference Books

1. Handbook of Cloud Computing, Dr. Anand Nayyar, BPB Publications (2019)
2. Cloud Computing: A Practical Approach, Toby Velte, Anthony Velte, Robert C., McGraw Hill
3. Cloud Computing for Data Analysis, Noah Gift, Alfredo Deza, Pragmatic AI Labs
4. Data Science on the Google Cloud Platform, Valliappa Lakshmanan, O'Reilly
5. Machine Learning in the AWS Cloud: Amazon SageMaker, Abhishek Mishra, Wiley

Activities:

Outcome: Explain cloud computing concepts including service models, deployment models, and virtualization.

Activity: Prepare a comparative chart/report on cloud service and deployment models with real-world examples (AWS, Azure, GCP).

Evaluation Method: Report submission & viva (assess clarity, accuracy, and examples used).

Outcome: Demonstrate cloud storage and database services for managing large-scale data.
Activity: Perform lab experiments on block/file/object storage and execute queries in BigQuery / RDS.
Evaluation Method: Lab performance & practical exam (students demonstrate CRUD operations and explain storage use cases).

Outcome: Apply cloud-based platforms to run data science and machine learning workflows.
Activity: Implement a cloud-based ETL pipeline (e.g., using AWS Glue / Dataflow) for preparing a dataset.
Evaluation Method: Lab report + demo evaluation (workflow completeness, correctness of execution).

Outcome: Build and deploy ML models on cloud services using AutoML and managed ML platforms.
Activity: Train and deploy a classification/regression model using AWS SageMaker / Azure ML / Google AutoML.
Evaluation Method: Practical demo + oral viva (assess deployment success, prediction results, and understanding of pipeline).

Outcome: Evaluate and monitor cloud-deployed solutions with respect to scalability, performance, and cost.
Activity: Configure monitoring (e.g., CloudWatch, Stackdriver) for a deployed ML service and analyze resource usage.
Evaluation Method: Mini-project report & presentation (grading based on monitoring setup, analysis quality, cost optimization suggestions).

SEMESTER-V

COURSE 13 B: CLOUD COMPUTING FOR DATA SCIENCE

Practical **Credits: 1** **2 hrs/week**

1. Create Virtual Machine using VMware workstation for Windows/Linux
2. Install & configure WAMP/XAMPP/Apache on the VM and host a sample page
3. Install and configure a cloud account (AWS/Azure/GCP free tier).
4. Create and manage storage buckets; upload and access datasets.
5. Launch an instance and configure Block-based storage (EBS)
6. Create & configure File-based storage on cloud VM (EFS/Network FS).
7. Set up Jupyter Notebook/Colab on cloud VM.
8. Connect to cloud-hosted database services (AWS RDS, BigQuery, Cosmos DB).
9. Implement a batch ETL pipeline in the cloud.
10. Launch a SageMaker notebook, attach IAM role and S3 bucket, run sample notebook
11. Build a classification/regression model using AWS SageMaker / Azure ML Studio / GCP AI Platform.
12. Implement a simple ETL job: extract (RDS / CSV), transform, load into cloud DW (e.g., Redshift / BigQuery).
13. Use CloudWatch / Stackdriver to monitor endpoints, set alarms and auto-scale rules.
14. Use cloud AutoML services for dataset prediction tasks.
15. Deploy a trained ML model as a REST API endpoint in the cloud.

SEMESTER-VI

COURSE 14 A: NEURAL NETWORKS AND DEEP LEARNING

Theory **Credits: 3** **3 hrs/week**

Course Objectives:

1. Introduce the fundamental concepts of Artificial Neural Networks and Deep Learning, along with their historical and biological inspirations.
2. Provide an in-depth understanding of different neural network architectures including Perceptron, DNN, CNN, RNN, and advanced models.
3. Develop hands-on skills to design, train, and evaluate deep learning models using popular frameworks such as TensorFlow and Keras.
4. Expose students to applications of deep learning in computer vision, natural language processing, and generative modeling.
5. Enable students to critically analyze challenges in deep learning such as overfitting, bias, and ethical concerns.

Course Outcomes:

After successful completion of this course, students will be able to:

1. Explain the principles of neural networks, perceptrons, activation functions, and the evolution of deep learning.
2. Apply concepts of forward/backward propagation, weight initialization, and optimization techniques to train deep neural networks.
3. Design and implement convolutional neural networks (LeNet, AlexNet, VGG) for image classification tasks.
4. Build and analyze recurrent neural networks (RNN, LSTM, GRU) for sequential data and natural language processing applications.
5. Experiment with advanced deep learning concepts such as transfer learning, generative models, and transformers using pre-trained models.

Unit 1. Foundations of Deep Learning:

What is Artificial Intelligence, Machine Learning, and Deep Learning? History and applications of deep learning, Biological vs. Artificial Neurons

Introduction to Neural Networks, Perceptron and activation functions (Linear, ReLU, Sigmoid, Tanh, Softmax), Types of Neural Networks (shallow vs. deep, feedforward vs. recurrent), Gradient descent and backpropagation (conceptual only), Concept of loss functions (MSE, cross-entropy) at intuitive level

Unit 2. Deep Neural Networks:

Forward and backward propagation, Weight initialization, learning rate, and optimization algorithms (SGD, Adam, RMSProp), Overfitting & underfitting: Regularization, Dropout, Batch normalization, Activation functions in deep networks, Loss functions in detail (MSE, cross-entropy, hinge loss)
Introduction to Keras/TensorFlow framework

Unit 3. Convolutional Neural Networks (CNNs):

Introduction to images and pixels, Filters/kernels, padding, and pooling, CNN architecture and layers (Conv, Pooling, Fully Connected, Softmax), Classical CNN architectures: LeNet-5 (digit recognition - first CNN model), AlexNet (ImageNet breakthrough - deeper CNN), VGG (concept of depth, simplicity)
Applications in image classification, object detection, facial recognition

Unit 4. Recurrent Neural Networks (RNNs) and NLP

Sequences and time series data, Introduction to RNNs: vanishing/exploding gradient issue LSTM and GRU (intuitive and architectural view), Word embeddings: Word2Vec, GloVe, introduction to contextual embeddings (BERT at high level)
Applications: Sentiment analysis, text generation, simple time-series forecasting

Unit 5. Advanced & Emerging Topics:

Generative models: GANs (Generator & Discriminator intuition), VAEs (introduction only), Transformers: attention mechanism (intuitive), BERT, GPT family (overview), Transfer learning & fine-tuning pre-trained models (vision & NLP), AI ethics: Bias, fairness, privacy, safety, explainability

Textbooks:

1. Chollet, F. (2018). *Deep Learning with Python* (1st ed.). Manning Publications.
2. Nielsen, M. A. (2015). *Neural Networks and Deep Learning*. Determination Press. *(Available free online: http://neuralnetworksanddeeplearning.com)*

Reference Books:

1. Géron, A. (2019). *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow* (2nd ed.). O'Reilly Media.

2. Howard, J., & Gugger, S. (2020). *Practical Deep Learning for Coders*. O'Reilly Media. *(Based on the fast.ai course)*
3. Shane, J. (2019). *You Look Like a Thing and I Love You: How AI Works and Why It's Making the World a Weirder Place*. Voracious / Hachette Book Group.

Activities:

Outcome: Explain the principles of neural networks, perceptrons, activation functions, and the evolution of deep learning.

Activity: *Concept Mapping Exercise* - Students work in small groups to draw a timeline-based concept map linking biological neurons → perceptron → multilayer perceptron → activation functions → deep learning.

Evaluation Method: Rubric-based evaluation of concept maps (clarity, correctness, and depth of connections).

Outcome: Apply concepts of forward/backward propagation, weight initialization, and optimization techniques to train deep neural networks.

Activity: *Hands-on Coding Task* - Students implement a simple 3-layer neural network from scratch in Python/NumPy (without Keras/TensorFlow) to observe propagation and optimization effects.

Evaluation Method: Code demonstration + oral viva where students explain how changing initialization/learning rate affects training.

Outcome: Design and implement convolutional neural networks (LeNet, AlexNet, VGG) for image classification tasks.

Activity: *Mini Project (Image Classification)* - Students form teams to train CNNs (LeNet, AlexNet, VGG) on a dataset like MNIST/CIFAR-10 and compare results.

Evaluation Method: Project report + presentation with accuracy comparison, confusion matrix, and discussion of design choices.

Outcome: Build and analyze recurrent neural networks (RNN, LSTM, GRU) for sequential data and natural language processing applications.

Activity: *Case Study on Sentiment Analysis* - Students use IMDB reviews dataset to build RNN/LSTM/GRU models for sentiment classification and analyze performance.

Evaluation Method: Submission of case study report with experimental setup, evaluation metrics (accuracy/F1-score), and reflection on differences among models.

Outcome: Experiment with advanced deep learning concepts such as transfer learning, generative models, and transformers using pre-trained models.

Activity: *Model Exploration & Demonstration* - Students choose one advanced technique (Transfer Learning on ResNet, GAN for image generation, or Transformer for text classification) and prepare a live demo in class.

Evaluation Method: Evaluation of demo + short reflective note (1–2 pages) on challenges, benefits, and application potential of the chosen technique.

SEMESTER-VI

COURSE 14 A: NEURAL NETWORKS AND DEEP LEARNING

Practical **Credits: 1** **2 hrs/week**

1. Build a perceptron from scratch in Python
2. Use Google Teachable Machine or Tensor Flow Playground
3. Visualize various Activation Functions and their Gradients
4. Build and train a deep neural network for classification (e.g., MNIST digits)
5. Experiment with dropout, batch normalization, and different activations
6. Train a CNN to classify fashion images (Fashion-MNIST)
7. Visualize filters and feature maps
8. Fine-tune a pre-trained CNN (Mobile Net, VGG) for a small dataset
9. Build an LSTM model for movie review sentiment analysis (IMDb dataset)
10. Generate text using a simple character-level RNN
11. Use a pre-trained model (like MobileNet or BERT) for a simple task
12. Use Huggingface to deploy a Sentiment Analysis App for Swiggy Reviews

SEMESTER-VI

COURSE 14 B: TIME SERIES ANALYSIS AND FORECASTING

Theory **Credits: 3** **3 hrs/week**

Course Objectives

The course aims to:

1. Provide fundamental understanding of time series data, components, and characteristics.
2. Train students in identifying, modeling, and forecasting using ARMA/ARIMA/SARIMA models.
3. Introduce state-space and multivariate approaches for complex data.
4. Familiarize students with modern forecasting methods, including spectral and evaluation techniques.
5. Enable hands-on practice with real-world datasets using R/Python statistical libraries.

Course Outcomes

By the end of the course, students will be able to:

1. Explain the concepts of time series, stationarity, and autocorrelation functions.
2. Apply ARMA/ARIMA/SARIMA models to real-world time series data.
3. Analyze multivariate and state-space time series using appropriate tools.
4. Implement forecasting workflows using R/Python for financial, business, and scientific datasets.
5. Evaluate forecast accuracy and select appropriate models using statistical criteria.

Unit 1. Fundamentals & Stationary Processes

Introduction to time series: types, components, forecasting process. Stationary processes: definitions, autocovariance, autocorrelation functions (ACF/PACF).

Model evaluation metrics. ACF/PACF example analyses.

Unit 2. ARMA & Forecasting with ARMA

ARMA(p,q) models: definition, estimation, forecasting approaches. Model identification: AIC, PACF/ACF, diagnostic checks. Practical examples of fitting ARMA and generating forecasts.

Unit 3. Non-Stationary & Seasonal Models

Non-stationary time series: differencing, unit roots. Seasonal models: SARIMA and multiplicative seasonal ARIMA. Identification, estimation, and diagnostic checks for seasonal models.

Unit 4. State-Space & Multivariate Time Series

Multivariate time series: Vector ARMA models (VARMA), estimation, forecasting. State-space representation: formulation, Kalman filter basics, forecasting in state-space models.

Unit 5. Advanced Topics & Forecast Evaluation

Spectral analysis: frequency-domain representation, spectral density. Forecast performance: measures, monitoring, choosing models.

Textbook:

1. Introduction to Time Series and Forecasting, Peter J. Brockwell & Richard A. Davis, 2nd Edition, Springer

Reference Books

1. Time Series Analysis: Forecasting and Control, George E. P. Box, Gwilym M. Jenkins & Gregory C. Reinsel
2. Introduction to Time Series Analysis and Forecasting, Douglas C. Montgomery, Cheryl L. Jennings, Murat Kulahci , (Wiley)
3. Time Series Analysis and Its Applications: With R Examples, R. H. Shumway & D. S. Stoffer

Activities:

Outcome: Explain the concepts of time series, stationarity, and autocorrelation functions

Activity: Students will prepare a seminar or short presentation explaining stationarity, ACF, PACF with a simple dataset example (like sales data).

Evaluation Method: Evaluated through presentation quality, understanding during viva, and a short concept quiz.

Outcome: Apply ARMA/ARIMA/SARIMA models to real-world time series data

Activity: Students will conduct a hands-on lab task to fit ARIMA and SARIMA models on stock price or rainfall data using Python/R.

Evaluation Method: Lab record submission, correctness of implementation, and a practical exam.

Outcome: Analyze multivariate and state-space time series using appropriate tools
Activity: Students will carry out a case study on macroeconomic datasets (like GDP, inflation, unemployment) using VAR or state-space modeling.
Evaluation Method: Case study report, results interpretation, and oral viva.

Outcome: Implement forecasting workflows using R/Python for financial, business, and scientific datasets
Activity: Students will design an end-to-end forecasting pipeline (data preprocessing → model building → forecasting → visualization). Example: forecasting COVID-19 daily cases or retail sales.
Evaluation Method: Project demo, code submission, and project report.

Outcome: Evaluate forecast accuracy and select appropriate models using statistical criteria
Activity: Students will compare multiple forecasting methods (e.g., ARIMA vs. Exponential Smoothing) on the same dataset and analyze performance using RMSE, MAE, and MAPE.
Evaluation Method: Written assignment, interpretation of metrics, and justification of chosen model.

SEMESTER-VI

COURSE 14 B: TIME SERIES ANALYSIS AND FORECASTING

Practical **Credits: 1** **2 hrs/week**

(Using R/Python statsmodels, pandas, forecast, or equivalent)

1. Import and visualize time series datasets (stock, weather, sales).
2. Perform decomposition of time series into trend, seasonal, residual components.
3. Compute and plot Autocorrelation Function (ACF) & Partial ACF (PACF).
4. Test stationarity using Augmented Dickey-Fuller (ADF) test.
5. Fit ARMA models and validate residuals.
6. Implement ARIMA and SARIMA models for seasonal data.
7. Perform model selection using AIC/BIC and cross-validation.
8. Forecast with ARIMA/SARIMA and plot prediction intervals.
9. Apply multivariate time series (VAR) to macroeconomic datasets.
10. Explore state-space models using Kalman filtering.
11. Conduct spectral analysis of a time series.
12. Compare forecasting methods: ARIMA vs. Exponential Smoothing vs. ML models.
13. Evaluate forecast performance with RMSE, MAPE, etc.

SEMESTER-VI

COURSE 15 A: NATURAL LANGUAGE PROCESSING

Theory **Credits: 3** **3 hrs/week**

Course Objectives:

1. Introduce the foundations of Natural Language Processing and its applications in real-world tasks.
2. Familiarize students with text preprocessing, linguistic analysis, and parsing techniques.
3. Equip learners with methods for information extraction, word representations, and sentiment classification.
4. Explore deep learning techniques for NLP, including RNNs, LSTMs, GRUs, and Transformers.
5. Provide hands-on experience with modern NLP tools (NLTK, spaCy, Hugging Face) for implementing applications such as chatbots, summarization, and document classification.

Course Outcomes:

At the end of this course, students will be able to:

1. Explain the principles, challenges, and applications of NLP and use basic text processing tools.
2. Apply preprocessing techniques (tokenization, stemming, lemmatization) and parsing methods to analyze language structures.
3. Implement information extraction and text representation methods (NER, embeddings, classification pipelines).
4. Build and evaluate deep learning models (RNN, LSTM, GRU, Transformer) for NLP tasks.
5. Utilize pre-trained transformer models (BERT, GPT) with Hugging Face for advanced NLP applications such as summarization, chatbots, and document classification.

Unit 1. Introduction to NLP and Language Fundamentals:

Definition, Goals, and Scope of NLP, Real-world Applications (Assistants, Chatbots, Translation, Summarization, QA, Spam Detection), Fundamentals of Language Processing, Ambiguities in NLP (Lexical, Structural, Contextual)

Installations: Python setup, NLTK, spaCy basics

Regular Expressions (Essential patterns, findall, split, sub, matching tokens)

Unit 2. Text Preprocessing and Linguistic Analysis:

Key NLP Terminologies: Morphology, Lexicon, Orthographic Rules

Finite State Transducers

Text Preprocessing Techniques: Tokenization, Stopword Removal, Stemming, Lemmatization

Grammar and Context-Free Grammar

Parsing Techniques: Top-down, Bottom-up, CYK Algorithm

Semantic Analysis: Elements, Meaning Representation

Unit 3. Information Extraction and Representation:

Named Entity Recognition (NER): Concepts, Examples, Using spaCy & NLTK

Word Embeddings: Word2Vec (Skip-Gram, CBOW), Comparison, Implementations, Bag of Words and N-grams, Text Classification Pipeline, Sentiment Analysis Applications

Ethical considerations in preprocessing & classification

Unit 4. Deep Learning for NLP:

Recurrent Neural Networks (RNN): Basics, RNN vs CNN/Feedforward NN, LSTM and GRU for Sequence Modeling, Transformer Models: Introduction, Pretrained Models (BERT, GPT), Hugging Face Ecosystem

Unit 5. Transformers and Modern NLP:

Transformer architecture basics (self-attention, encoder-decoder), BERT: Pretraining, Fine-tuning,

GPT and Generative NLP, Hugging Face Ecosystem (using pre-trained models)

Text Summarization: Extractive, Abstractive, Hybrid Approaches

Applications: Document Classification, Chatbots, Virtual Assistants

Textbook:

1. Natural Language Processing, Sini Raj Pulari, Umadevi Maramreddy, Shriram k. Vasudevan, Oxford University Press
2. Speech and Language Processing, An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition, Daniel Jurafsky, James H. Martin, Pearson Education, 2023.

Reference Book:

1. Natural Language Processing and Information Retrieval, Tanveer Siddiqui, U.S. Tiwary , Oxford University Press.
2. 2. Natural Language Processing Recipes - Unlocking Text Data with Machine Learning and Deep Learning using Python, Akshay Kulkarni, Adarsha Shivananda, Apress, 2019.

Activities:

Outcome: Explain NLP fundamentals and basic text processing tools.

Activity: Quiz/ Assignment on Analyze ambiguities in Indian language sentences.

Evaluation Method: Quiz Score

Outcome: Apply preprocessing and parsing techniques to analyze language.

Activity: Case study: Building a simple grammar-based sentence parser.

Evaluation Method: Application of text preprocessing to raw text corpus, parsing

Outcome: Implement information extraction and text representation methods.

Activity: Hands-on NER using spaCy and NLTK.

Evaluation Method: Lab report submission on embeddings & NER.

Outcome: Build and evaluate deep learning models for NLP.

Activity: Group Discussion on Discussion: Compare RNN vs Transformers.

Evaluation Method: Depth of Understanding, Participation, Explanation

Outcome: Utilize pre-trained transformer models for advanced NLP applications.

Activity: Lab: Implement chatbot using GPT model.

Evaluation Method: Practical exam using Hugging Face models – Accuracy, Effectiveness

SEMESTER-VI

COURSE 15 A: NATURAL LANGUAGE PROCESSING

Practical **Credits: 1** **2 hrs/week**

1. Install Python, NLTK, and spaCy. Write a sample program to print available NLP corpora and models.
2. Write regex patterns for extracting emails, phone numbers, hashtags, and dates from a text file.
3. Demonstrate lexical and structural ambiguity with example sentences. Use NLTK parse trees to visualize.
4. Implement sentence and word tokenization using NLTK and spaCy. Compare outputs.
5. Write a program to remove stopwords and analyze text length reduction.
6. Apply Stemming and Lemmatization on a dataset and compare differences.
7. Use NLTK to demonstrate top-down and bottom-up parsing of a simple grammar.
8. Use spaCy to extract entities (e.g., names, locations, organizations) from news text.
9. Implement text representation and calculate similarity between documents.(Bag of Words and N-grams)
10. Build a sentiment classifier using Scikit-learn (Naive Bayes / Logistic Regression).
11. Implement a simple RNN to generate sentences character by character.
12. Hugging Face to load a pretrained BERT model and perform masked word prediction.
13. Implement extractive and abstractive summarization using Hugging Face pipelines.
14. Build a simple FAQ-based chatbot using Transformer-based embeddings.

SEMESTER-VI

COURSE 15 B: DATA ENGINEERING & MLOPS

Theory **Credits: 3** **3 hrs/week**

Course Objectives

1. To introduce the lifecycle and roles in Data Engineering.
2. To explore data architecture principles, distributed systems, and technology choices.
3. To analyze MLOps features, risks, and challenges in developing ML systems.
4. To design CI/CD pipelines and deployment strategies for ML models.
5. To understand monitoring, governance, and Responsible AI compliance in production ML.

Course Outcomes

At the end of the course, students will be able to:

1. Explain Data Engineering and its organizational roles.
2. Analyze major concepts in data architecture and distributed systems.
3. Apply MLOps features and evaluate challenges in ML model development.
4. Design and implement CI/CD pipelines for ML deployment.
5. Evaluate governance and Responsible AI practices in MLOps.

Unit 1. Foundations of Data Engineering

Data Engineering: definition, lifecycle, skills, activities. Evolution and roles of Data Engineers: technical vs business responsibilities, internal vs external roles. Relationship between Data Engineering and Data Science. Data lifecycle vs Data Engineering lifecycle.

Unit 2. Data Architecture & Distributed Systems

Enterprise and Data Architecture definitions. Principles of good data architecture. Scalability, failure design, tiers, microservices, monolith vs modular. Event-driven architecture, hybrid cloud, multicloud, edge computing. Technology selection criteria: team size, interoperability, cost, TCO.

Unit 3. MLOps Fundamentals

MLOps challenges and risk mitigation. Responsible AI and scaling ML solutions. Key MLOps features: EDA, feature engineering, model training & evaluation, reproducibility. Deployment requirements, monitoring basics. Model versioning and experimentation tracking.

Unit 4. Model Deployment & CI/CD Pipelines

Preparing models for production. Runtime environments: dev to production adaptation.
CI/CD pipelines: building ML artifacts, testing pipelines. Deployment strategies: batch, online, A/B testing, canary releases. Containerization & scaling (Docker, Kubernetes).

Unit 5. Monitoring, Feedback Loops & Governance

Monitoring models in production: drift detection, ground truth evaluation.
Feedback loops: retraining workflows, online evaluation. Logging, monitoring frameworks.
Governance: regulations (GDPR, CCPA, GxP), Responsible AI principles.
Templates for governance, compliance, and model risk management.

Text/ Reference books

1. Fundamentals of Data Engineering, Joe Reis & Matt Housley, O'Reilly, 2022.
2. Introducing MLOps, Mark Treveil & Dataiku Team, O'Reilly, 2020

Web Resources:

https://www.ibm.com/think/topics/data-engineering
https://martinfowler.com/articles/microservices.html
https://towardsdatascience.com/a-gentle-introduction-to-mlops-7d64a3e890ff/

Activities

Outcome: Explain Data Engineering and roles
Activity: Prepare a concept map showing different Data Engineer roles in an organization.
Evaluation Method: Short presentation + written quiz.

Outcome: Analyze data architecture concepts
Activity: Case study on choosing between monolith, microservices, and event-driven architectures.
Evaluation Method: Case study report + viva.

Outcome: Apply MLOps features in ML development

Activity: Lab exercise on feature engineering & reproducibility using MLflow.

Evaluation Method: Lab record submission + demo.

Outcome: Design CI/CD pipelines for ML

Activity: Mini-project: build a simple ML CI/CD pipeline with GitHub Actions/Docker.

Evaluation Method: Project demo + evaluation rubric.

Outcome: Evaluate governance and Responsible AI practices

Activity: Group discussion & policy brief on GDPR/Responsible AI practices.

Evaluation Method: Written assignment + peer review.

SEMESTER-VI

COURSE 15 B: DATA ENGINEERING & MLOPS

Practical **Credits: 1** **2 hrs/week**

1. Install and configure a modern data engineering environment (Python, Jupyter, VSCode).
2. Explore and visualize the data lifecycle on a sample dataset (sales/weather).
3. ETL basics: extract data from CSV/JSON -> transform -> load into relational database.
4. Compare performance of batch vs event-driven ingestion using Apache Kafka or RabbitMQ.
5. Deploy a small dataset on Hadoop Distributed File System (HDFS) and perform simple operations.
6. Case study lab: design a microservices vs monolithic workflow for a mock business problem.
7. Perform Exploratory Data Analysis (EDA) and track experiments using **MLflow**.
8. Build a simple ML model (regression/classification) and enable **reproducibility** with version control.
9. Manage datasets and model versions using **DVC (Data Version Control)**.
10. Containerize an ML model with **Docker**.
11. Automate training and deployment with a **GitHub Actions CI/CD pipeline**.
12. Deploy an ML model as a REST API using **FastAPI / Flask**.
13. Implement model drift detection: monitor incoming data and compare with training data distribution.
14. Build a simple feedback loop: retrain a model automatically when drift exceeds a threshold.
15. Configure logging and monitoring using **Prometheus/Grafana** for a deployed ML model.
16. Case study: analyze GDPR/Responsible AI implications on a real dataset (e.g., facial recognition).