ಬೆಂಗಳೂರು ನಗರ ವಿಶ್ವವಿದ್ಯಾನಿಲಯ

BENGALURU CITY UNIVERSITY

BENGALURU CITY UNIVERSITY

Syllabus of Third and Fourth Semesters for

Bachelor of Computer Applications
in
Data Science (BCA-DS)
(CBCS Scheme)

Under
State Education Policy

Effective from the Academic Year
2025 – 2026

Board of Studies in Computer Science for UG

No : BCU/BoS/Comp.Sci. & Appln.(PG & UG)/389/2024-25 dated 19-2-2025

1	Prof. Ramesh B Kudenatti Department of Mathematics Bengaluru City University,Bengaluru-560056	Chairperson
2	Prof. Guru D S Department of Studies in Computer Science University of Mysore, Mysore-570006	Member
3	Prof. Aziz Makandar Department of Computer Science Karnataka State Akkamahadevi Women University, Jnanashakti Campus, Vijayapura-586109	Member
4	Prof. Suneetha Department of Computer Science, Karnataka State Open University, Muktha Gangothri, Mysuru-570006	Member
5	Prof. Veena R Department of MCA, Seshadripuram College, Seshadripuram, Bengaluru-560020	Member
6	Prof. Kiran Kumar M N Department of Computer Applications, BMS College of Commerce and Management, Bengaluru-560004	Member
7	Prof. Latha B Department of Computer Science Vijaya College, R V Road, Basavanagudi, Bengaluru-560004	Member
8	Prof. R Shanthi Krishna Department of Computer Applications, SSMRV College, Jayanagar, Bengaluru-560041	Member
9	Prof. Roopa H R Department of Computer Applications, Seshadripuram Institute of Commerce and Management, Seshadripuram, Bengaluru-560020	Member
10	Sri Seby Kallarakkal CEO-Nabler Web Solutions, Bengaluru-560052	Member

Name of the Degree Program	**:**	**Bachelor of Computer Applications- BCA (DS)**
Discipline Course	**:**	**Computer Science**
Starting Year of Implementation	**:**	**2024-25 (I & II Semesters)** **2025-26 (III & IV Semesters)** **2026-27 (V & VI Semesters)**

Programme Outcomes (POs)

PO 1	Apply knowledge of computer science fundamentals, data science principles, and mathematical foundations to design and implement computational solutions.
PO 2	Identify, analyse and solve structured and unstructured real-world problems using algorithmic thinking, data structures, and efficient programming strategies.
PO 3	Collect, clean, visualize and analyze data using tools and techniques to derive meaningful insights for informed decision-making.
PO 4	Use modern technologies and tools such as AWS Cloud, Machine Learning libraries, visualization platforms, and statistical software to develop data-driven applications.
PO 5	Demonstrate project management skills through capstone projects and practical labs and apply research methodologies to explore contemporary issues in data science and analytics.
PO 6	Build software solutions using multiple paradigms including Object-Oriented Programming, Unix Shell Scripting, and Scientific Programming, promoting versatility in software development.
PO 7	Understand professional, ethical, legal and societal responsibilities in data usage, privacy, AI implementation and technological impact.
PO 8	Communicate effectively in both verbal and written formats through project reports, lab records, presentations and document code.
PO 9	Function effectively as an individual and as a member or leader in diverse teams in multidisciplinary environments such as group projects and internships.
PO 10	Recognize the need for continuous learning in a rapidly evolving tech world and demonstrate readiness to adopt innovations in AI, ML, IoT, and cloud computing for solving real-world challenges.

ASSESSMENT

Weightage for the Assessments (in percentage)

Type of Course	Formative Assessment/ I.A.	Summative Assessment (S.A.)
Theory	20%	80 %
Practical	20%	80 %

Detailed Structure for BCA- DS Course

Semester	Course Code	Paper Title	Teaching Hours / Week	Marks		Duration of Exam in Hours	Credits
				Exam	IA		
I	24BCA11	Discrete Structure	03	80	20	03	03
	24BCA12	Problem Solving Technique	03	80	20	03	03
	24BCA13	Computer Architecture	03	80	20	03	03
	24BCA12P	PST Lab	04	40	10	03	02
	24BCA13P	Computer Architecture Lab	04	40	10	03	02
	24BCA1P	Office Automation Tools Lab	04	40	10	03	02
	24BCAL11	Language-I	04	80	20	03	03
	24BCAL12	Language-II	04	80	20	03	03
	24BCACC1	Constitutional Values-I	02	40	10	1.5	02
						Total Credits	**23**
II	24BCA21	Data Structure	03	80	20	03	03
	24BCA22	Object Oriented Programming Using Java	03	80	20	03	03
	24BCA23	Operating System	03	80	20	03	03
	24BCA21P	Data Structure Lab	04	40	10	03	02
	24BCA22P	Java Programming Lab	04	40	10	03	02
	24BCA23P	Unix & Shell Programming Lab	04	40	10	03	02
	24BCAL21	Language-I	04	80	20	03	03
	24BCAL22	Language-II	04	80	20	03	03
	24BCACC2	Constitutional Values-II	02	40	10	1.5	02
	24BCACC3	Environmental Studies	02	40	10	1.5	02
						Total Credits	**25**

Semester	Course Code	Paper Title	Teaching Hours / Week	Marks		Duration of Exam in Hours	Credits
				Exam	IA		
III	24BCA31	Database Management System	04	80	20	03	03
	24BCA32	Foundations of Data Science	04	80	20	03	03
	24BCA33	Python Programming	04	80	20	03	03
	24BCA34	Data Visualization	02	40	10	1.5	02
	24BCA31P	DBMS Lab	03	40	10	03	02
	24BCA32P	Data Science Lab	03	40	10	03	02
	24BCA33P	Python Programming Lab	02	40	10	03	02
	24BCAL31	Language-I	04	80	20	03	03
	24BCAL32	Language-II	04	80	20	03	03
Total Credits							**23**
IV	24BCA41	Artificial Intelligence	04	80	20	03	03
	24BCA42	Data Analytics	04	80	20	03	03
	24BCA43	Scientific Programming using R	04	80	20	03	03
	24BCA44	Data Mining	02	40	10	1.5	02
	24BCA41P	Artificial Intelligence Lab	03	40	10	03	02
	24BCA42P	Data Analytics Lab	03	40	10	03	02
	24BCA43P	R Programming Lab	03	40	10	03	02
	24BCASE1	Basics of Natural Language Processing	02	40	10	1.5	02
	24BCAL41	Language-I	04	80	20	03	03
	24BCAL42	Language-II	04	80	20	03	03
Total Credits							**25**

Semester	Course Code	Paper Title	Teaching Hours / Week	Marks		Duration of Exam in Hours	Credits
				Exam	IA		
V	24BCA51	Data Science Project Management	04	80	20	03	03
	24BCA52	Business Data Analytics	04	80	20	03	03
	24BCA53	Time Series Analysis	04	80	20	03	03
	24BCA54	Machine Learning	04	80	20	03	03
	24BCA52P	BDA Lab	03	40	10	03	02
	24BCA53P	Time Series Analysis Lab	03	40	10	03	02
	24BCASE2	Quantitative Techniques	03	40	10	1.5	02
	24BCAPJ	Project	08	120	30	03	06
Total Credits							**24**
VI	24BCA61	Cloud Computing & AWS for Data Science	04	80	20	03	03
	24BCA62	Inferential Statistics	04	80	20	03	03
	24BCA63	Deep Learning	04	80	20	03	03
	24BCA64	Internet of Things	04	80	20	03	03
	24BCA62P	Inferential Statistics Lab	03	40	10	03	02
	24BCAIS	Internship	--	80	20	03	06
Total Credits							**20**
Overall Total Credits							**140**

Detailed Syllabus for BCA – Data Science

SEMESTER – III

Theory	**24BCA31: Database Management System**
Teaching Hours : 04 Hours/Week	**Credits : 03**
Duration of Exam : 03 Hours	**Maximum Marks : 100 (Exam 80 + IA 20)**

Course Outcomes

COs	Description
CO1	Understand the fundamental concepts of data, database systems, DBMS architecture, data models and the various types of DBMS along with their classifications.
CO2	Design high-level conceptual data models using the Entity-Relationship approach, apply design principles and understand physical storage structures, file organization, and indexing mechanisms.
CO3	Apply the relational model concepts, perform normalization using functional dependencies and write effective SQL queries to manage and manipulate data constraints and views.
CO4	Utilize relational algebra to query databases, analyze query processing and optimization techniques and explain transaction management, concurrency control and recovery mechanisms in DBMS. Basics of PL/SQL.

UNIT – I Fundamentals of Database Systems and Architecture 14 Hours

Introduction to Data and Database, History of Database. Characteristics of the Database Approach. Significance and Advantages of Database Management Systems. Actors on the Scene, Workers behind the Scene. System Structure: Instance and Schema, Data Models, Data Independence. Three Schema Architecture. Database Languages and Interfaces. The Database System Environment. Centralized and Client-Server Architecture. Classification of Database Management System.

UNIT – II Database Design and Storage Structures 14 Hours

High-Level Conceptual Data Models for Database Design. Entity Types, Entity Sets, Attributes and Keys. Relationship Types, Relationship Sets, Roles and Structural Constraints. Weak Entity Types. Extended ER Features. Refining the ER Design. Naming Conventions and Design Issues. ER to Relational Mapping. File Organization and Storage. Secondary Storage Devices. File Organization Techniques. Single-Level Ordered Index. Multi-Level Indexes. Indexes on Multiple Keys.

UNIT – III Relational Model, Normalization and SQL 14 Hours

Relational Model Concepts. Relational Model Constraints and Relational Database Schema. Update Operations and Dealing with Constraint Violations. Anomalies in a Database. Functional Dependency. Armstrong's Axioms. Closure of a Relation and Attributes. Lossless Join and Dependency Preservation. Normalization: 1NF, 2NF, 3NF, BCNF. Structure of Relational Databases. SQL: Data Definition and Data Types. Specifying Constraints in SQL. Schema Change Statements. Insert, Delete and Update Statements. Views (Virtual Tables). Assertions and Triggers.

UNIT – IV Query Processing, Transactions and PL/SQL 14 Hours

Unary Relational Operations: SELECT and PROJECT. Relational Algebra Operations. Binary Relational Operations: JOIN and DIVISION. Additional Relational Operations. Query Processing and Optimization: Evaluation of Relational Algebra Expressions. Query Equivalence. Introduction to Transaction Processing. Transaction and System Concepts. States of a Transaction. Desirable Properties of Transactions (ACID). Concurrency Control Techniques: Two-Phase Locking Techniques. Backup and Recovery from Failures, Basics of PL/SQL Programming.

TEXT BOOKS:

1. Elmasri R. and Navathe S.B., *Fundamentals of Database Systems*, 7th Edition, Addison-Wesley, 2016.
2. Silberschatz A., Korth H.F., and Sudarshan S., *Database System Concepts*, 7th Edition, Tata McGraw Hill, 2019.
3. Ivan Bayross, *SQL, PL/SQL – The Programming Language of Oracle*, 4th Edition, BPB Publications.

REFERENCE BOOKS:

1. C.J. Date, A. Kannan, S. Swamynathan, *An Introduction to Database Systems*, 8th Edition, Pearson Education, 2009.
2. Raghu Ramakrishnan and Johannes Gehrke, *Database Management Systems*, 3rd Edition, McGraw Hill, 2003.
3. Kevin Loney and George Koch, *Oracle 9i – The Complete Reference*, McGraw-Hill International Edition.

Theory	24BCA32: Foundations of Data Science
Teaching Hours : 04 Hours/Week	Credits: 03
Duration of Exam: 03 Hours	Maximum Marks: 100 (Exam 80 + IA 20)

Course Outcomes

COs	Description
CO1	Understand the evolution, roles and applications of Data Science, along with data security concerns.
CO2	Use different strategies to collect, clean, integrate, and transform data for analysis.
CO3	Compute key statistical measures like mean, standard deviation, skewness, and correlation, and visualize data using plots.
CO4	Develop simple and multiple regression models, evaluate them using visual tools, and apply them for predictions.

UNIT - I Foundations of Data Science and Its Applications 14 Hours

Introduction to Data Science, Evolution of Data Science , Data Science Roles, Stages in a Data Science Project , Applications of Data Science in various fields , Data Security Issues.

UNIT - II Data Acquisition and Preprocessing Techniques 14 Hours

Data Collection Strategies , Data Pre-Processing Overview , Data Cleaning , Data Integration and Transformation , Data Reduction , Data Discretization.

UNIT - III Exploratory Data Analysis and Statistical Techniques 14 Hours

Descriptive Statistics , Mean, Standard Deviation, Skewness and Kurtosis , Box Plots , Pivot Table, Heat Map ,Correlation Statistics , ANOVA.

UNIT - IV Regression Analysis and Predictive Modeling 14 Hours

Simple and Multiple Regression , Model Evaluation using Visualization , Residual Plot , Distribution Plot – Polynomial Regression and Pipelines , Measures for In- sample Evaluation – Prediction and Decision Making.

TEXT BOOKS:

1. Jojo Moolayil, “Smarter Decisions : The Intersection of IoT and Data Science”, 1st Edition, Packt Publishing Ltd., UK, 2016.
2. Cathy O’Neil and Rachel Schutt, “Doing Data Science”, 1st Edition,O'Reilly, USA, 2013.
3. David Dietrich, Barry Heller, Beibei Yang, “Data Science and Big data Analytics”, 1st Edition, EMC Education Services, 2015.

REFERENCE BOOKS:

1 Joel Grus, “Data Science from Scratch”, 1st Edition, O’REILLY, USA, 2015.

2. Rafael A. Irizarry, “Introduction to Data Science”, 1st Edition,Chapman & Hall,USA, 2022.

3 Gupta. S.C. & Kapoor,V.K. , “Fundamentals of Mathematical Statistics”, 12th Edition Sultan Chand & Sons Pvt. Ltd., New Delhi,2002.

Theory	24BCA33: Python Programming
Teaching Hours : 04 Hours/Week	Credits: 03
Duration of Exam: 03 Hours	Maximum Marks: 100 (Exam 80 + IA 20)

Course Outcomes

COs	Description
CO1	Understand the basic concepts and apply them to develop programs.
CO2	Demonstrate proficiency in handling Python's complex data structures and apply file handling for various formats.
CO3	Apply object-oriented programming concepts in Python and use libraries for efficient data manipulation and analysis.
CO4	Utilize Python packages and APIs for effective data visualization and apply them in real-time data analysis projects.

UNIT – I Foundations of Python Programming 14 Hours

Introduction to Python: Python Interpreter/Shell, Identifiers, Keywords, Statements and Expressions, Variables, Operators, Precedence and Associativity, Data Types, Indentation, Comments, Reading Input, Print Output, Type Conversions, The type() function and is operator, Dynamic and Strongly Typed Language. Control Flow: Conditional blocks: if, else, elif, Nested if. Looping: while, for, range, loop manipulation using break, continue, else, pass. Functions: Function Definition and Calling, Built-In Functions, Return Statement, Default Parameters, Scope and Lifetime of Variables, Command Line Arguments. Strings: Creating and Storing Strings, String Operations, Slicing, Joining, String Methods.

UNIT – II Data Structures and File Handling 14 Hours

Lists: Creating Lists, Basic List Operations, Indexing and Slicing, List Methods, The del Statement. Dictionaries: Creating Dictionaries, Accessing and Modifying Key-Value Pairs, Dictionary Methods. Tuples: Creating Tuples, Tuple Operations, Indexing, Slicing, Tuple Methods, Relationships between Tuples, Lists, and Dictionaries. Sets and Frozensets: Creating Sets, Set Operations, Set Methods. Iterators and Iterables. File Handling: Types of Files, Reading and Writing Text and Binary Files, CSV File Handling, Pickle Module.

UNIT – III Object-Oriented Programming and Data Handling Libraries 14 Hours

Object-Oriented Programming: Classes and Objects in Python, Constructor Method, Multiple Objects, Class vs Data Attributes, Encapsulation, Inheritance, Polymorphism. Introduction to Python Libraries for Data Handling: NumPy – Arrays and Operations, Pandas – Series and DataFrames, Indexing and Querying, Handling Missing Values, Data Aggregation, Grouping, and Summarization.

UNIT – IV Data Analysis and Visualization 14 Hours

Importing and Exporting Data (CSV, JSON), Understanding and Formatting Data. Using Matplotlib and Plotly for Visualization, Generating and Plotting Data (Line Graphs, Bar Charts), Random Walks, Dice Simulation, Working with APIs, Downloading Data and Visualizing Repositories using Plotly, Mapping Global Datasets with JSON.

TEXT BOOKS:

1. Wesley J. Chun, *Core Python Applications Programming*, 3rd Edition, Pearson Education, 2016.
2. Yashavant Kanetkar and Aditya Kanetkar, *Let Us Python*, 3rd Edition, BPB Publications.
3. Jeeva Jose & P. Sojan Lal, *Introduction to Computing and Problem Solving with Python*, Khanna Publishers, New Delhi, 2016.
4. Eric Matthes, *Python Crash Course – A Hands-On, Project-Based Introduction to Programming*, 2nd Edition, No Starch Press, 2019.
5. Gowrishankar S, Veena A, *Introduction to Python Programming*, 1st Edition, CRC Press/Taylor & Francis, 2018. ISBN: 978-0815394372.

REFERENCE BOOKS:

1. Allen B. Downey et al., *How to Think Like a Computer Scientist: Learning with Python*, John Wiley, 2015.
2. John Zelle, *Python Programming: An Introduction to Computer Science*, 2nd Edition, Course Technology, Cengage Learning, 2013.
3. A.N. Kamthane & A.A. Kamthane, *Programming and Problem Solving with Python*, McGraw Hill Education, 2017.
4. Mark Lutz, *Learning Python*, 5th Edition, O'Reilly Publications, 2013. ISBN: 978-1449355739.
5. Ljubomir Perkovic, *Introduction to Computing Using Python – An Application Development Focus*, Wiley, 2012.

Theory	24BCA34: Data Visualization
Teaching Hours : 02 Hours/Week	Credits : 02
Duration of Exam : 1.5 Hours	Maximum Marks : 50 (Exam40 + IA10)

Course outcomes

COs	Description
CO1	Understand the significance of data visualization, types of data and the visualization process to support effective decision making.
CO2	Compare features of popular visualization tools and apply data storytelling principles to design interactive and meaningful dashboards.
CO3	Apply visualization design principles, use of color, and data modeling techniques to create clear and insightful visual representations.

UNIT – I: Introduction to Data Visualization and Tools 14 Hours

Definition and importance of data visualization – Role of data visualization in decision making – Types of data: numerical, categorical, temporal, geographical – Data visualization process: data collection, exploration, analysis, visualization, interpretation – Challenges and limitations of data visualization. Overview of Visualization Tools (e.g., Excel, Tableau, Power BI, Python) – Comparing and contrasting features and use cases among these tools.

UNIT – II: Data Storytelling and Visualization Design 14 Hours

Principles of Data Storytelling: Narrative and Context – Best Practices for Dashboard Layout and Interactivity – Principles of Good Visualization Design – Understanding and Using Color in Visualizations – Importance of Data Modelling in Visualization

TEXT BOOKS:

1. Cole Nussbaumer Knaflic,"Storytelling with Data: A Data Visualization Guide for Business Professionals", 1st edition,Wiley, USA,2015.
2. Edward Tufte,“The Visual Display of Quantitative Information”,2nd Edition , Graphics Press, USA,2001.

REFERENCE BOOKS:

1. Kieran Healy,"Data Visualization: A Practical Introduction", 1st Edition, Princeton University Press,USA, 2018.
2. Alberto Ferrari and Marco Russo, "Analyzing Data with Power BI and Power Pivot for Excel",1st edition, Microsoft Press,USA, 2017.
3. Devin Knight, Brian Knight, Mitchell Pearson, and Manuel Quintana, "Microsoft Power BI Complete Reference",1st edition, Packt Publishing,UK,2018.

Lab	24BCA31P: DBMS Lab
Teaching Hours : 03 Hours/Week	Credits : 02
Duration of Exam : 3 Hours	Maximum Marks : 50 (Exam 40 + IA 10)

Part A

1. Create a table STUDENT with the following fields: RollNo, Name, DOB, Department, Marks.

Insert at least 5 records
Display all records
Update marks for a specific student
Delete a student record

2. Create a table COURSE with CourseID, CourseName, Credits.

Alter the table to add a field Department
Drop the field Credits

3. Create two tables EMPLOYEE1 and EMPLOYEE2 with the following attributes: (FNAME, MNAME, LNAME, SSN, BDATE, ADDRESS, SEX, SALARY, SUPERSSN, DNO). Perform UNION, INTERSECT, and MINUS operations on them.

4. Use the STUDENT table to perform Aggregate functions and Scalar functions

5. Using the STUDENT and COURSE tables,

Find students who scored above average marks using a subquery.
List students enrolled in 'Computer Science' using a subquery.

6. Create DEPARTMENT and STUDENT tables.

Write queries to display student names along with their department names using JOIN operations.

7. Create a view to show student names and marks from the STUDENT table where Marks >75.

Query the view and update the view.

8. Demonstrate the use of GRANT and REVOKE on the STUDENT table.

Use COMMIT and ROLLBACK after INSERT and DELETE commands.

Part B

9. Write a PL/SQL block to accept a number and check if it is even or odd.

10. Write a PL/SQL program to divide two numbers and handle the exception if the denominator is zero.

11. Use a parameterized cursor to display students based on department input.

12. Create a stored function to calculate grade based on marks:

Above 80: Distinction
60–79: First Class
40–59: Second Class
Below 40: Fail

13. Write a stored procedure to update marks of a student given their RollNo and new marks.

14. Create a BEFORE INSERT trigger on STUDENT to ensure marks are not entered as negative.

15. Create a trigger to log changes into a table STUDENT_LOG whenever the marks are updated.

16. Use a cursor with a loop to count and display the number of students in each department.

Lab	24BCA32P: Data Science Lab
Teaching Hours : 03 Hours/Week	Credits : 02
Duration of Exam : 3 Hours	Maximum Marks : 50 (Exam40 + IA10)

Part - A

1. Load a dataset from Kaggle and implement the basic operations – head(), info(),describe(), detect missing values, drop or fill null values, remove duplicate values.
2. Create 2 datasets – customers.csv(customerid, Name, City) and purchases.csv (customerid, purchaseamount). Implement the following operations:
   a)Merge 2 datasets b) Create new column: Discount = purchaseamount * 0.1
   c) Normalize purchaseamount using Min-Max scaling
3. From iris dataset, implement dimensionality reduction technique to reduce 4d to 2d. Visualize original vs reduced dataset using matplotlib.
4. Create a dataset containing customer id and age and implement Data Discretization.
5. Create a dataset containing students marks in 5 different subjects and display the mean and standard deviation for each subject.
6. Create a dataset containing Student Id, Height(in CM) and compute the Skewness and Kurtosis.
7. Create a dataset containing the 5 consecutive monthly sales of 3 different products and draw box plot to identify the outliers and distribution spread.
8. Create a dataset containing Customer Age, Income, Spending Score and Savings. Calculate Correlation matrix and plot heat map.
9. Create a dataset containing Car's Id, Weight, Horsepower and MPG. Calculate Pearson Correlation Coefficients.
10. Create a dataset containing test scores of the student with 3 different teaching methods. Perform one-way ANOVA to test if means differ significantly.

Part – B

1. From Boston Housing Dataset, predict the housing price based on average number of rooms. Evaluate Plot regression line and residuals.
2. From AutoMPG Dataset, predict miles per gallon using multiple car features. Check residual plots and model accuracy.
3. Implement Polynomial Regression on Salary Data (Employee Experience Vs Salary). Fit Polynomial regression (degree 2 or 3) for better fit. Visualize polynomial curve fit and residuals.
4. Use Diabetes dataset from sklearn to create a pipeline that scales features and apply linear regression. Evaluate R^2 and RMSE.
5. Use Advertising Dataset to build regression model and analyze residual plots. Plot residual histogram and scatter plots.
6. Use California Housing dataset and train the regression model and plot the distribution of prediction errors. Use Seaborn distplot/kdeplot for residuals.

Lab	24BCA33P: Python Programming Lab
Teaching Hours : 03 Hours/Week	Credits : 02
Duration of Exam : 3 Hours	Maximum Marks : 50 (Exam 40 + IA 10)

Part A

1. Write a Python program to declare variables, perform arithmetic operations, and display results.
2. Create a program to check if a number is even or odd using if-else.
3. Write a Python program to print the first n Fibonacci numbers using a for loop.
4. Implement a program that accepts a string and counts the number of vowels and consonants.
5. Create a program to store student details in a dictionary and retrieve details based on user input.
6. Demonstrate the use of break, continue, and pass in loops.
7. Write a program to create NumPy arrays, perform element-wise operations, and reshape arrays.
8. Create a Pandas Series and perform indexing, slicing, and querying operations.
9. Load a dataset into a Pandas Data Frame and perform sorting and filtering operations.
10. Write a program to handle missing values by filling them with mean/median values.

Part B

1. String Operations: Write a program to count the occurrences of each word in a given string.
2. Implement a program to insert, delete, and update elements in a list.
3. Create a dictionary with employee details and perform CRUD operations.
4. Implement a program to generate prime numbers up to n using a generator function.
5. Write a Python program to calculate the factorial of a number using recursion.
6. Write a program to read and write student marks into a text file and display the contents.
7. Plot a line graph and a bar chart using Matplotlib.
8. Write a program to load a CSV file into Pandas and perform basic data analysis.
9. Implement a program to group a dataset by category and calculate summary statistics.
10. Load a dataset from Scikit-learn and display its basic properties.

SEMESTER – IV

Theory	24BCA41: Artificial Intelligence
Teaching Hours : 04 Hours/Week	Credits : 03
Duration of Exam : 03 Hours	Maximum Marks : 100 (Exam 80 + IA 20)

Course Outcomes

COs	Description
CO1	Understand the fundamental concepts of Artificial Intelligence and problem-solving strategies.
CO2	Apply logical reasoning techniques, perform inferences, solve constraint satisfaction problems and basic learning methods.
CO3	Design AI systems using planning techniques and reasoning, interpret data using perception models.
CO4	Explore the application domains and evaluate ethical and societal implications of AI technologies.

UNIT – I Fundamentals of Artificial Intelligence and Search Techniques 14 Hours

Definitions, Applications, and Scope. Intelligent Agents: Agents and Environments, Concept of Rationality, Nature of the Environment, Structure of Agents. Knowledge-Based Agents: Introduction, The Wumpus World as an Example. Problem Solving: Problem-Solving Agents, Formulating Problems. Search Techniques: Uninformed Search Strategies – Depth First Search (DFS), Breadth First Search (BFS), Iterative Deepening Search. Informed Search Strategies – Best First Search, A* Search, AO* Search, Means-End Analysis. Adversarial Search and Games: Two-Player Zero-Sum Games, Minmax Algorithm, Alpha-Beta Pruning.

UNIT – II Knowledge Representation, Reasoning, and Learning Paradigms 14 Hours

Propositional Logic, First-Order Predicate Logic, Differences between Propositional and First-Order Inference. Inference Techniques: Unification and Lifting, Forward Chaining, Backward Chaining, Resolution, Truth Maintenance Systems. Constraint Satisfaction Problems (CSPs): Definition, Examples, Backtracking Search. Learning Concepts: Rote Learning, Learning by Taking Advice, Learning in Problem Solving, Learning from Examples. Decision Trees and Winston's Learning Program.

UNIT – III Planning, Reasoning and Perception 14 Hours

Introduction to Planning: Planning Problem, State-Space Search, The Blocks World Problem, STRIPS Representation. Handling Uncertainty: Non-Monotonic Reasoning, Probabilistic Reasoning, Introduction to Fuzzy Logic and Fuzzy Set Theory. Introduction to Perception: Computer Vision – Image Classification, Object Detection. Natural Language Processing (NLP): Introduction, Syntactic Processing, Semantic Analysis, Discourse and Pragmatic Processing.

UNIT – IV Machine Learning, Neural Networks, and AI Ethics 14 Hours

Types of Learning – Supervised, Unsupervised, and Reinforcement Learning. Neural Networks: Basics of Artificial Neural Networks (ANN), Deep Learning Concepts , Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), Long Short-Term Memory (LSTM) Networks and their Applications. Expert Systems: Architecture, Components, and Role of Expert Systems with Two Case Studies. Legal and Ethical Issues in AI: Societal Impact, Bias, Privacy, and Accountability in AI Systems.

TEXT BOOKS:

1. M.C. Trivedi, *A Classical Approach to Artificial Intelligence*, Khanna Book Publishing Company, 2024 (AICTE Recommended).
2. Nilsson Nils J., *Artificial Intelligence: A New Synthesis*, Morgan Kaufmann Publishers Inc., ISBN: 978-1-55860-467-4.
3. Dan W. Patterson, *Introduction to Artificial Intelligence & Expert Systems*, PHI Learning, 2010.
4. Rajiv Chopra, *Data Science with Artificial Intelligence, Machine Learning and Deep Learning*, Khanna Book Publishing Company, 2024.
5. Russell, S. and Norvig, P., *Artificial Intelligence – A Modern Approach*, 3rd Edition, Prentice Hall.

REFERENCE BOOKS:

1. M.C. Trivedi, *Introduction to AI and Machine Learning*, Khanna Book Publishing Company, 2024.
2. Van Hirtum, A. & Kolski, C., *Constraint Satisfaction Problems: Algorithms and Applications*, Springer, 2020.
3. Elaine Rich, Kevin Knight, Shivasankar B. Nair, *Artificial Intelligence*, 3rd Edition, McGraw Hill, 2019.
4. Rajiv Chopra, *Machine Learning and Machine Intelligence*, Khanna Book Publishing Company, 2024.

Theory	24BCA42: Data Analytics
Teaching Hours : 04 Hours/Week	Credits : 03
Duration of Exam : 03 Hours	Maximum Marks : 100 (Exam 80 + IA 20)

Course Outcome

COs	Description
CO1	Understand the evolution, types, applications, and benefits of data analytics along with the essential skills required for business, web, and text analytics.
CO2	Analyze relationships between variables using correlation and regression techniques and solve problems using Pearson's and Spearman's methods.
CO3	Apply probability concepts, probability distributions, and hypothesis testing techniques such as t-test, ANOVA, and Chi-square in data analysis.
CO4	Develop interactive dashboards by importing, transforming, and visualizing data using Power BI tools for effective business intelligence reporting.

UNIT-I Introduction to Data Analytics 14 Hours

Evolution of Data Analytics, Data Analytics Overview, Types of Data Analytics -Descriptive Analytics -Diagnostic Analytics -Predictive Analytics -Prescriptive Analytics, Importance and Benefits of Data Analytics. Different Applications of Analytics in Business, Text Analytics and Web Analytics, Skills for Business Analytics.

UNIT-II Correlation & Regression 14 Hours

Correlation: Introduction, Meaning of Correlation, Types of correlation ,probable error, Karl pearson's coefficient of correlation for individual series only, Spearman's Rank correlation for individual series only. Regression: Introduction, definition, difference between correlation and regression, Simple linear regression, properties of regression coefficients, Regression equation x on y, Regression equation y on x , Simple Problems.

UNIT- III Probability & Statistical Methods 14 Hours

Sample Space, Types of Events, Measures of probability, conditional probability, Bayes' theorem, Random variable. Probability Distributions: Binomial, Poisson and Normal Distributions – Definitions, means, variances and applications of these distributions. Simple problems. Estimation and Hypothesis Testing- t-test, Analysis of variance (ANOVA) and Chi-square test.

UNIT-IV Power BI 14 Hours

Introduction to Power BI -What is Business Intelligence (BI)?, Overview of Power BI and its components: Power BI Desktop, Service, and Mobile , Installing Power BI Desktop, Power BI Interface & Navigation, Data sources supported by Power BI. Data Loading & Transformation using Power Query **:** Understanding Power Query Editor, Loading data from various sources (Excel, Web, SQL),Cleaning and transforming data:Remove columns/rows, Rename columns, Merge and Append queries, Data types and formatting.Visualizations and Dashboards-Types of charts: Bar, Column, Pie, Line, Map, Matrix, Card, Gauge, etc., Creating interactive dashboards, Formatting visuals and tooltips, Using slicers and filters, Drill through and bookmarks.

TEXT BOOKS:

1. Kumar, U.D. :Business Analytics – The Science of Data – Driven Decision Making, 1st Edition, Wiley, USA, 2014.
2. Dr Anil Maheshwari, Data Analytics Made Accessible, 1st Edition, Amazon.com Services LLC., USA, 2016.
3. Johnson, R.A., Miller, I. and Freund, Probability and Statistics for Engineers, 8th Edition,Pearson, USA, 2013.

REFERENCE BOOKS:

1. Gert, H.N., Thorlund, L. and Thorlund, J :Business Analytics for Managers – Taking Business Intelligence Beyond Reporting, 1st Edition, Wiley, USA,2018.
2. Dr. Gaurav Aroraa , Chitra Lele , Dr. Munish Jindal ; " Data Analytics: Principles, Tools, and Practices: A Complete Guide for Advanced Data Analytics Using the Latest Trends, Tools, and Technologies, 1st Edition, Amazon, USA, 2021.

Theory	24BCA43: Scientific Programming using R
Teaching Hours : 04 Hours/Week	Credits : 03
Duration of Exam : 03 Hours	Maximum Marks : 100 (Exam 80 + IA 20)

Course Outcomes

COs	Description
CO1	Demonstrate the ability to install and utilize R packages for performing data analysis tasks.
CO2	Apply R programming skills for efficient data analysis and statistical computing.
CO3	Use key statistical packages in R for analytics and data interpretation.
CO4	Integrate foundational concepts from statistics, probability, and linear algebra within the R programming environment to develop real time applications

UNIT – I Introduction to R and Fundamental Programming Constructs 14 Hours
Introduction to scientific programming, R basics, code editors for R, finding help, control structures, conditional executions, loops. Functions in R.

UNIT – II Utilities and System-Level Programming in R 14 Hours
Useful utilities, debugging utility, regular expressions, interpreting character string as expression, time-date-sleep, calling external software with system commands, running R commands.

UNIT – III Object-Oriented Programming and Package Development in R 14 Hours
Object oriented programming in R, define class and objects in R, assign generics and methods. Packages in R, installation process of various packages in R, data science packages in R, Building R packages.

UNIT – IV Comparative Analysis and Real-World Applications of R 14 Hours
Comparison of R with other scientific programming software, implementation of various industry use cases of scientific programming using R.

TEXT BOOKS:

1. Mark Gardener, Beginning R: The Statistical Programming Language,1st Edition, Wiley, USA, 2013.
2. Roger Peng R Programming for Data Science, 1st Edition, Leanpub, Digital Publication, 2016.

REFERENCE BOOKS:

1. Grolemund, Garrett, Hands-On Programming with R,1st Edition, O'Reilly, USA, 2014.
2. Garrett Wickham, Garrett Grolemund, R for Data Science, 1st Edition, O'Reilly, USA, 2017.

Theory	24BCA44: Data Mining
Teaching Hours : 02 Hours/Week	Credits : 02
Duration of Exam : 1.5 Hours	Maximum Marks : 50 (Exam 40 + IA 10)

Course Outcomes

COs	Description
CO1	Explain the basic concepts, techniques, and issues involved in data mining and pattern discovery.
CO2	Apply frequent pattern mining algorithms like Apriori and FP-Growth to discover associations in datasets.
CO3	Implement classification techniques including regression, Bayesian, decision trees, and k-nearest neighbor for predictive modeling.

Unit I Introduction **14 Hours**

Introduction to Data Mining-Types of Data and Patterns Mined- Technologies- Applications. Major Issues in Data Mining. Mining Frequent Patterns: Basic Concept – Frequent Item Set Mining Methods -Apriori and FP Growth algorithms - Mining Association Rules.

Unit II Statistical-Based Algorithms **14 Hours**

Introduction - Statistical-Based Algorithms: Regression – Bayesian Classification. Distance Based Algorithms: Simple Approach - K Nearest Neighbors. Decision Tree-Based Algorithms: ID3 – C4.5 - CART - Scalable DT techniques.

TEXT BOOKS:

1. 1.Margaret H Dunham, "Data Mining Introductory and Advanced Topics", 1st Edition,Pearson Education,USA, 2012.
2. Jiawei Han and Micheline Kamber, "Data Mining - Concepts and Techniques", Third Edition, Elsevier, USA,2012

REFERENCE BOOKS:

1. GalitShmueli,et al, "Data Mining for Business Analytics: Concepts, Techniques and Applications in R", 1st Edition,Wiley, India, 2018.
2. Pang-Ning Tan, Michael Steinbach, Vipin Kumar, Introduction to Data Mining, Second edition,Addison Wesley,USA, 2018.

Lab	24BCA41P: Artificial Intelligence Lab
Teaching Hours : 03 Hours/Week	**Credits : 02**
Duration of Exam : 3 Hours	**Maximum Marks : 50 (Exam 40 + IA 10)**

1. Demonstrate basic problem-solving using Breadth-First Search on a simple grid.
2. Implement Depth-First Search (DFS) on a small graph.
3. Solve the Water Jug Problem using Breadth First Search (BFS).
4. Implement a Hill Climbing search to find the peak in a numeric dataset.
5. Apply the A* Search algorithm to find the shortest path in a 4x4 grid.
6. Implement the Minmax search algorithm for 2-player games. You may use a game tree with 3 plies. .
7. Use constraint propagation to solve a Magic Square puzzle.
8. Apply optimization techniques to find the maximum value in a list.
9. Represent and evaluate propositional logic expressions.
10. Implement a basic rule-based expert system for weather classification.
11. Implement a basic AI agent with simple decision-making rules.
12. Implement a basic Rule-Based Chatbot.
13. Using Python NLTK, perform the following Natural Language Processing (NLP) tasks for text content.
    a) Tokenizing b) Filtering Stop Words c) Stemming d) Part of Speech tagging
    e) Chunking f) Named Entity Recognition (NER)

Lab	24BCA42P: Data Analytics Lab
Teaching Hours : 03 Hours/Week	**Credits : 02**
Duration of Exam : 3 Hours	**Maximum Marks : 50 (Exam 40 + IA 10)**

Part - A

1. From the given dataset **players_info.csv** ,
   a) What is the probability distribution of genders among the players?
   b) What is the probability of each batting style?
   c) What is the probability of each bowling style?
   d) What is the probability distribution of player positions?
   e) What is the probability distribution of countries among the players?
2. 80% of all the visitors to Museum of Goa end up buying souvenirs from the souvenir shop at the museum. On the coming Sunday, if a random sample of 10 visitors is picked, Find the Probability that every visitor will end up buying from the souvenir shop. Find the Probability that a maximum of 7 visitors will buy souvenirs from the souvenir shop.
3. A testing agency wants to analyze the complexity of SAT exam 2022. They have collected the SAT scores of 1000 students in "**sat_score.csv**". Calculate the probability that a student will score less than 800 in SAT exam. Calculate the probability that a student will score more than 1300 in SAT exam.
4. A Marketing services company reported that the typical American spends a mean of 144 minutes (2.4 hours) per day accessing the Internet via a mobile device. Select a sample of 30 friends and family members whose mobile access time is stored in a CSV file "**InternetMobileTime.csv".** Is there evidence that the population mean time spent per day accessing the Internet via mobile device is different from 144 minutes? (Level of Significance $\alpha = 0.05$)
5. A hotel manager looks to enhance the initial expression that hotel guests have when they check in. Contributing to initial impressions is the time it takes to deliver a guest's luggage to the room after check_in. A random sample of 20 deliveries on a particular day is selected in Wing A of the hotel and a random sample of 20 deliveries is selected in Wing B. The results are stored in "**Luggage.csv**". Analyze the data and determine whether there is a difference between the mean delivery time in the 2 wings of the hotel.(Use $\alpha = 0.05$)
6. The file **'Concrete.csv'** contains the compressive strength in thousands of pounds/square inch, of 40 samples of concrete taken 2 and 7 days after pouring. At the 0.01 level of significance, is there evidence that the mean strength is lower at 2 days than at 7 days?
7. Two companies A and B were merged. After the first appraisal cycle post merger, employees originally belonging to company B have put an allegation that the management favours employees who were originally a part of company A. At 95% confidence perform a hypothesis test to validate if the claim holds good.

Promotion Status

Company	P	NP	Total
A	15	9	24
B	16	15	31
Total	31	24	55

8. Traffic management inspector in a city wants to understand whether carbon emissions from different cars are different. For this reason, the inspector has taken random samples from all registered cars on the road in that city and would like to test if the amount of carbon emission release depends on fuel type at 5% significance level. **Dataset – AOVData.csv**
9. Find the relationship between the price of a laptop with other factors of the dataset **"laptops.csv"**.

Part - B

1. Install Power BI Desktop, explore the interface, and load data from an Excel file.
2. Use Power Query Editor to clean and transform a dataset.
3. Use Power Query Editor to combine data from multiple tables.
4. Visualize data using basic chart types - Bar Chart, Column Chart and Pie Chart by loading a dataset.
5. 5.Use cards, maps, and matrices for advanced visuals by loading a dataset.
6. 6.Enhance dashboard interactivity using slicers and filters.
7. Use drill-through pages and bookmarks for navigation and detail views.

Lab	24BCA43P: Scientific Programming using R Lab
Teaching Hours : 03 Hours/Week	Credits : 02
Duration of Exam : 3 Hours	Maximum Marks : 50 (Exam 40 + IA 10)

PART - A

1. Write a program to demonstrate variable assignments and data types in R.
2. Use if, else if, and else statements to categorize age groups.
3. Implement a for loop to calculate the factorial of a number.
4. Define a user-created function to compute the mean and standard deviation of a numeric vector.
5. Write a program using try Catch() to handle errors during division operations.
6. Demonstrate the use of debug(), browser(), and traceback() for debugging a faulty function.
7. Use regular expressions to extract all email addresses from a given text string.
8. Write a script that logs current time, pauses execution using Sys.sleep(), and resumes with another log.
9. Execute a system command from R to list all files in the current directory using system() or shell().

PART - B

1. Define an S3 class for "Student" and implement a method to calculate average marks.
2. Create an S4 class for "Employee" with slots like ID, Name, and Salary, and a method to increment salary.
3. Install and demonstrate the use of the ggplot2 package with a simple bar chart.
4. Explore dplyr functions to filter, arrange, and summarize a dataset.
5. Write and build a basic custom R package including documentation and metadata using devtools.
6. Implement a linear regression model on a sample dataset using lm() in R.
7. Perform a time-series plot for monthly sales data using ts() and plot().
8. Create a simple dashboard using shiny that displays summary statistics of a dataset.
9. Build a case study in R to analyze COVID-19 dataset: read data, clean, visualize, and interpret trends.

Theory: Skill Enhancement Course-I	24BCASE1: Basics of Natural Language Processing
Teaching Hours : 02 Hours/Week	Credits : 02
Duration of Exam : 1.5 Hours	Maximum Marks : 50 (Exam 40 + IA 10)

Course Outcomes

COs	Description
CO1	Understand and apply fundamental NLP concepts such as morphology, syntax, semantics, and various text preprocessing techniques like tokenization, stemming, and POS tagging.
CO2	Implement feature extraction techniques including BoW, TF-IDF, and word embeddings, and perform basic text similarity and clustering using NLP models and tools.

UNIT I: Introduction to NLP & Text Processing **14 Hours**

Overview of NLP - Definition and Scope of NLP, Challenges in NLP, Components: Morphology, Syntax, Semantics, Pragmatics. NLP and Linguistics- Role of linguistics in NLP, Language models: N-gram models, Bag of Words and TF-IDF. Text Preprocessing – Tokenization, Stop word removal, Stemming and Lemmatization, POS tagging, Regular expressions in text processing.

UNIT II: Syntax, Semantics, and Vectorization **14 Hours**

Syntax and Parsing - Parsing techniques: Shallow parsing, Dependency parsing, Context-Free Grammar (CFG). Semantics and Word Meaning - Lexical semantics, WordNet and its usage, Named Entity Recognition (NER). Feature Extraction Techniques - Bag of Words (BoW) revisited, TF-IDF vectorization Word Embeddings: Word2Vec, GloVe, One-hot encoding and limitations. Text Similarity & Clustering - Cosine Similarity, Text clustering using K-Means (intro).

TEXT BOOKS:

1. Daniel Jurafsky and James H. Martin, “Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition”, second edition,Prentice Hall, USA,2008 .
2. James A., Natural language Understanding 2e, Pearson Education, USA,1994
3. Bharati A., Sangal R., Chaitanya V.. Natural language processing: a Paninian perspective, 1st Edition,PHI, New Delhi,2000.

BLUEPRINT FOR QUESTION PAPER						
FOR 03 CREDITS						
Marks	**Unit I**	**Unit II**	**Unit III**	**Unit IV**	**Number of questions to be answered**	**Total**
2	**3**	**3**	**3**	**3**	**10**	**20**
6	**2**	**2**	**2**	**2**	**5**	**30**
10	**1**	**1**	**1**	**1**	**3**	**30**
					Total	**80**
FOR 02 CREDITS						
Marks	**Unit I**	**Unit II**	**Number of questions to be answered**			**Total**
2	**3**	**3**	**5**			**10**
5	**3**	**3**	**4**			**20**
10	**1**	**1**	**1**			**10**
					Total	**40**

Formative Assessment - 03 Credits

Category	Marks Allotted
Tests	10
Assignments	10
Total Marks	**20**

Formative Assessment - 02 Credits

Category	Marks Allotted
Tests	5
Assignments	5
Total Marks	**10**