

P.G. Diploma in Data Science Semester I (2025-26)

Discipline Centric Core Course (DCC)

FOCCC27001T: Fundamentals of Computers & Operating System

(20 CIA + 80 EoSE. = Max. Marks: 100)

Course Credits	No. of Teaching Hours Per Week	Total No. of Teaching Hours
4 Credits	4 Hours	60 Hours
Course Outcome: On successful completion of the course, the students will be able to:		
• Understand the basic components, functions, and types of computers and their applications. • Describe the architecture, working, and characteristics of input, output, memory, and storage devices. • Explain the concept of software, types of software, and role of system and application software. • Analyze the role and functions of operating systems in managing hardware and software resources. • Compare different types of operating systems and demonstrate understanding of their features (e.g., batch, multi-user, real-time).		
SYLLABUS		
Unit-I: Evolution of computers, Classification of computers: Micro, Mini, Mainframe, Supercomputers, Characteristics and applications of computers, Block diagram of a computer, Input, Output and Storage Devices		
Unit-II: Data and Information Management: Concept of Data, Information, and Knowledge, Types and sources of data, Introduction to databases, File systems vs. DBMS, Basic concepts of data storage and retrieval, Data security and privacy basics		
Unit-III: Data Communication and Networks: Introduction to data communication, Network types: LAN, MAN, WAN, PAN, Topologies: Bus, Ring, Star, Mesh, Hybrid, Internet, Intranet, and Extranet, Basics of IP address, domain names, and DNS, Network devices: Router, Switch, Hub, Modem		
Unit-IV: Applications and Emerging Trends: Role of IT in business, education, healthcare, and governance, E-Governance, E-Commerce, M-Commerce, Social, legal, and ethical aspects of IT, Cloud Computing and IoT – introduction, Cyber security		
Unit-V: Computer Software and Operating Systems: System Software and Application Software, Operating Systems: Functions and types (DOS, Windows, Linux, Android), Introduction to programming languages: Machine, Assembly, High-level languages, Open-source software and licensing, Utility programs and their types		
SUGGESTED BOOKS		
1. "Fundamentals of Computers" – V. Rajaraman 2. "Computer Fundamentals" – P.K. Sinha & Priti Sinha 3. "Introduction to Computers" – Peter Norton 4. "Operating System Concepts" – Abraham Silberschatz, Peter B. Galvin, Greg Gagne 5. "Operating Systems" – Achyut S. Godbole & Atul Kahate		

P.G. Diploma in Data Science Semester I (2025-26)

Discipline Centric Core Course (DCC)

IDSCC27001T: Introduction to Data Science & Data Lifecycle

(20 CIA + 80 EoSE. = Max. Marks: 100)

Course Credits	No. of Teaching Hours Per Week	Total No. of Teaching Hours
3 Credits	3 Hours	45 Hours
Course Outcome: On successful completion of the course, the students will be able to:		
• Define and explain key concepts and components of data science. • Identify types and sources of data and their applications. • Describe and apply the data science lifecycle stages. • Analyze case studies to understand data-driven decision-making. • Recognize ethical issues and trends in data science practice.		
SYLLABUS		
Unit-I: Fundamentals of Data Science: Definition, scope and evolution of data science, Importance of data science in business, governance, and research, Components of data science: Data, Algorithms, Models, Tools. Roles in data science: Data Scientist, Data Analyst, Data Engineer. Interdisciplinary nature (Math, Stats, CS, Domain Knowledge)		
Unit-II: Types and Sources of Data: Structured, semi-structured, and unstructured data, Qualitative vs. Quantitative data, Sources: Web data, IoT, Surveys, Databases, APIs. Big Data vs. Traditional Data, Open data, proprietary data, and real-time streaming data		
Unit-III: Data Science Lifecycle Phases: Problem Definition, Data Collection, Data Cleaning & Preprocessing, Exploratory Data Analysis (EDA), Modeling & Algorithms, Evaluation & Interpretation, Deployment, Feedback & Maintenance, Iterative nature of the lifecycle, Tools used in each stage (brief)		
Unit-IV: Data-Driven Decision Making: Decision-making with data vs. intuition-based decisions, Case studies from domains (Healthcare, Finance, Marketing, etc.), Importance of domain knowledge, Key Performance Indicators (KPIs) and Metrics. Basics of A/B Testing		
Unit-V: Ethics, Privacy, and Future Trends: Data privacy and protection (GDPR, Indian Data Protection Law basics), Ethical concerns in AI/ML (bias, fairness, transparency), Data democratization and open data challenges, Emerging trends: AutoML, Explainable AI, Edge Data Science, Career opportunities and learning pathways		
SUGGESTED BOOKS		
1. Doing Data Science – Cathy O’Neil & Rachel Schutt 2. Data Science for Business – Foster Provost & Tom Fawcett 3. The Data Science Handbook – Field Cady 4. Python for Data Analysis – Wes McKinney		

Discipline Centric Core Course (DCC)

PDSCC27001T: Programming in Python for Data Science

(20 CIA + 80 EoSE. = Max. Marks: 100)

Course Credits	No. of Teaching Hours Per Week	Total No. of Teaching Hours
3 Credits	3 Hours	45 Hours
Course Outcome: On successful completion of the course, the students will be able to:		
• Understand the basics of Python programming syntax and semantics. • Implement control structures, functions, and modules in Python. • Understand functional properties of built-in data structures in Python • Apply object-oriented programming features in Python. • Utilize Python libraries such as NumPy and Pandas for Data Science applications.		
SYLLABUS		
Unit-I: Introduction to Python Programming: Features of Python, Installation and working with Python IDEs, Basic Syntax, Variables, and Data Types, Input / Output operations, Type conversions and operators		
Unit-II: Control Structures and Functions: Conditional statements (if, if-else, nested if), Loops: for, while, nested loops, break, continue, pass, Defining and calling functions, Arguments and return values, Lambda functions, Recursion, Modules and packages		
Unit-III: Python Data Structures: Lists, Tuples, Sets, and Dictionaries, List comprehensions, String manipulation and operations, Iterators and Generators, Working with built-in functions on data structures		
Unit-IV: Object-Oriented Programming in Python: Introduction to OOP: Classes and Objects, Constructor, Destructor, Inheritance, Polymorphism, Method Overriding, Encapsulation and data hiding		
Unit-V: Python for Data Science: NumPy: Arrays, Operations, Indexing, Broadcasting, Pandas: Series, DataFrames, Reading CSV/Excel files, Data cleaning, filtering, grouping, merging, Simple plotting with Matplotlib, Data visualization basics		
SUGGESTED BOOKS		
1. Core Python Programming, R. Nageswara Rao, Dreamtech Press, Second Edition, 2018 2. Python Programming, Dr. M. Suresh Anand, Dr. R. Jothikumar, Dr. N. Vadivelan, Notion Press, First Edition, 2020 3. The Complete Reference Python, Martin C. Brown, McGraw Hill Education, Fourth Edition, 2018 4. Think Python, Allen B. Downey, O'Reilly Media, 2016 5. Programming and Problem Solving with Python, Amit Ashok Kamthane, Ashok Namdev Kamthane, McGraw Hill HED, First Edition, 2017 6. Advanced Python Programming, Sakis Kasampalis, Quan Nguyen, Dr. Gabriele Lanaro, Ingram short title, 2019		

Discipline Centric Core Course (DCC)

DSPCC27001T: Descriptive Statistics & Probability

(20 CIA + 80 EoSE. = Max. Marks: 100)

Course Credits	No. of Teaching Hours Per Week	Total No. of Teaching Hours
3 Credits	3 Hours	45 Hours
Course Outcome: On successful completion of the course, the students will be able to:		
- Understand the key concepts of descriptive statistics including data types, graphical representation, and central tendency. - Understand the concept of central tendency, its uses and interpretation - Apply techniques of data dispersion, skewness, and kurtosis in data analysis. - Understand fundamental concepts of probability and its theorem. - Understand probability distribution and can apply statistical thinking to real-world data interpretation and decision-making in Data Science.		
SYLLABUS		
Unit-I: Introduction to Statistics and Data: Definition and scope of statistics, Types of data: qualitative, quantitative, discrete, continuous, Scales of measurement: nominal, ordinal, interval, ratio, Methods of data collection, Frequency distribution and tabulation, Graphical representation: Histogram, Bar chart, Pie chart, Ogive		
Unit-II: Measures of Central Tendency: Arithmetic mean, median, mode, Geometric mean, harmonic mean, Partition values: quartiles, deciles, percentiles, Uses and interpretation in Data Science		
Unit-III: Measures of Dispersion and Shape: Range, interquartile range, variance, standard deviation, mean deviation, Coefficient of variation, Moments: raw and central moments, Skewness: Karl Pearson & Bowley's method, Kurtosis and interpretation		
Unit-IV: Basics of Probability : Experiment, sample space, events, Classical, empirical and axiomatic definitions of probability, Conditional probability and Bayes' Theorem Independence of events, Applications in Data Science scenarios		
Unit-V: Probability Distributions: Random variables: discrete and continuous, Expectation, variance of random variables, Discrete distributions: Binomial Distribution, Poisson Distribution, Continuous distributions:, Normal Distribution, Standard normal curve and Z-scores, Real-life applications of distributions in Data Science		
SUGGESTED BOOKS		
- Goon, Gupta & Dasgupta – Fundamentals of Statistics – World Press - S.C. Gupta – Fundamentals of Mathematical Statistics – Sultan Chand - Navidi, W. – Statistics for Engineers and Scientists – McGraw-Hill		

Ability Enhancement Compulsory Course (AECC)

ECTAC27001T: English Communication & Technical Writing

(20 CIA + 80 EoSE. = Max. Marks: 100)

Course Credits	No. of Teaching Hours Per Week	Total No. of Teaching Hours
4 Credits	4 Hours	60 Hours
Course Outcome: On successful completion of the course, the students will be able to: • Students will gain a comprehensive understanding of basic sounds of English and identify key literary forms. • Students will Interpret and appreciate selected poetic and literary texts by Shakespeare, Tagore, and R.K. Narayan, demonstrating a grasp of theme, tone, and literary techniques. • Students will apply knowledge of English grammar structures such as form classes, articles, prepositions, modal auxiliaries, and the use of prefixes, suffixes, and connectives in context.. • Students will explore the concept of Using appropriate tenses, voice (active/passive), and speech (direct/indirect) so as to learn English effectively. • Students will compose formal and informal letters, reports, and job applications with clarity, correct structure, and suitable vocabulary.		
SYLLABUS		
Unit-I: The Sounds of English: Consonants, Mono-thongs, Diphthongs. An Acquaintance with Literary Forms:- Elegy, Ballad, and Sonnet An Acquaintance with Figures of Speech:- Simile, Metaphor, Personification, and Irony		
Unit-II: Poetry: William Shakespeare – All the World is a stage. Rabindranath Tagore – Where the Mind is without Fear. Act Play/Novel: R. K. Narayan – Vendor of Sweets.		
Unit-III: Introduction to Form Classes, Tenses and its uses. Articles, Preposition and Modal auxiliaries.		
Unit-IV: Active and Passive Voice, Direct and Indirect Speech, Prefixes and Suffixes, Connectives.		
Unit-V: English Writing Skills: Formal Letters & Informal Letters, Report Writing, Application for Job / Job Letter.		
SUGGESTED BOOKS		
1. "An Introduction to Literary Forms" by W.H. Hudson. 2. "The Vendor of Sweets" by R.K. Narayan 3. "English Verse: An Introduction" by David Kennedy 4. "High School English Grammar and Composition" by P.C. Wren & H. Martin. 5. "Effective English Communication" by M.A. Yadugiri & Geetha Rajeevan.		

Discipline Centric Core Course (DCC)

PPLCC27001P: Python Programming Lab

(20 CIA + 80 EoSE. = Max. Marks: 100)

Course Credits	No. of Teaching Hours Per Week	Total No. of Teaching Hours
1 Credits	2 Hours	30 Hours
SUGGESTED LABORATORY EXERCISES		
1. Write a program in Python to count repeated characters in a string. 2. Write a program in Python to implement various string operators. 3. Write a program in Python to implement different types of built in string oriented functions. 4. Write a program in Python to create, append, and remove list elements from a given list. (List elements must be from user side) 5. Write a program in Python to display the sum and square of given integer elements from a given list. 6. Write a program in Python to find the largest and smallest number in a user defined list. 7. Write a program in Python to find unique and duplicate items of a user defined list. 8. Write a program in Python to count the numbers of characters in the string and store them in a dictionary data structure. 9. Write a program in Python to implement linear search. 10. Write a program in Python to implement various built in functions for Tuple. 11. Write a program in Python to implement Regular Expression (RegEx). 12. Write a program in Python to define a module to find Fibonacci Numbers and import the module to another program. 13. Write a program in Python to define a module and import a specific function in that module to another program. 14. Write a Python script named demo.py. This script should prompt the user for the names of two text files. The contents of the first file should be input and written to the second file. 15. Write a program in Python to accept student details and store in a database.		

Discipline Centric Core Course (DCC)

DVLCC27001P: Data Visualization Lab (Matplotlib, Seaborn)

(20 CIA + 80 EoSE. = Max. Marks: 100)

Course Credits	No. of Teaching Hours Per Week	Total No. of Teaching Hours
1 Credits	2 Hours	30 Hours
SUGGESTED LABORATORY EXERCISES		
Part A: Using Matplotlib Basic Plots: • Create a line plot showing monthly sales of a company. • Create a bar chart to compare marks of students in three subjects. • Plot a pie chart showing the percentage of expenses (Rent, Food, Transport, Miscellaneous). Intermediate Visuals: • Use subplot() to draw multiple plots in a single figure. • Customize plot: Add title, axis labels, legend, and grid. Advanced Features: • Plot a histogram of random numbers using NumPy. • Use scatter() to visualize the relationship between age and income. • Part B: Using Seaborn Basic Seaborn Plots: • Load a built-in dataset (e.g., tips, iris, or penguins) using sns.load_dataset(). • Plot a boxplot to show the distribution of total bill across different days. • Create a violin plot for petal length grouped by species (use iris dataset). Categorical Plots: • Use countplot() to visualize gender distribution in the dataset. • Use barplot() to display average tip per day. Relationship Plots: • Plot a pairplot() for the iris dataset. • Use lmplo() to show linear relationship between total_bill and tip. • Create a dashboard-style visualization using matplotlib.pyplot.subplots() to display: ✓ A line chart ✓ A pie chart ✓ A histogram ✓ A scatter plot (All in one figure using subplots.)		

Discipline Centric Core Course (DCC)

SLPCC27001P: Statistics Lab using Python (NumPy, SciPy)

(20 CIA + 80 EoSE. = Max. Marks: 100)

Course Credits	No. of Teaching Hours Per Week	Total No. of Teaching Hours
1 Credits	2 Hours	30 Hours
SUGGESTED LABORATORY EXERCISES		
Part A: Descriptive Statistics using NumPy		
1. Mean, Median, Mode Write a Python program using NumPy to: Find the mean, median, and mode of a given dataset.		
2. Standard Deviation and Variance Calculate standard deviation and variance for the same dataset.		
3. Range and Percentiles Find the range, 25th, 50th, and 75th percentiles using NumPy functions.		
Part B: Probability Distributions using SciPy		
4. Normal Distribution Plot Generate a normal distribution with mean=0, std=1, and plot it using <code>matplotlib</code>.		
5. Binomial Distribution Simulate a binomial distribution ($n = 10$, $p = 0.5$), and calculate the probability of exactly 5 successes.		
6. T-Distribution and Confidence Interval Given a sample, compute the 95% confidence interval using <code>scipy.stats.t.interval</code>.		
Part C: Hypothesis Testing		
7. One-Sample t-test Perform a one-sample t-test to check if the sample mean is significantly different from a given population mean.		
8. Two-Sample t-test Compare the means of two samples using a two-sample t-test.		
Part D: Real-world Dataset Analysis		
9. Read CSV Data & Analyze - Load a dataset (e.g., <code>student_scores.csv</code>) using <code>pandas</code>, then: - Compute summary statistics. - Plot histogram and boxplot. - Perform hypothesis testing on student marks.		

P.G. Diploma in Data Science Semester II (2025-26)

Discipline Centric Core Course (DCC)

FDACC27002T: Fundamentals of Data Analytics

(20 CIA + 80 EoSE. = Max. Marks: 100)

Course Credits	No. of Teaching Hours Per Week	Total No. of Teaching Hours
4 Credits	4 Hours	60 Hours
Course Outcome: On successful completion of the course, the students will be able to: • Understand key concepts of data analytics and its lifecycle. • Perform basic data cleaning and preprocessing tasks. • Apply statistical tools to analyze datasets. • Visualize data effectively using modern tools. • Use Python/Excel/Power BI for real-world analytics problems.		
SYLLABUS		
Unit-I: Introduction to Data Analytics What is Data Analytics? Types of Data Analytics: Descriptive, Diagnostic, Predictive, Prescriptive Data Science vs. Data Analytics Applications of Data Analytics in various fields (Business, Health, Education, etc.) Analytics Process Life Cycle		
Unit-II: Data Collection and Preparation Types and Sources of Data: Structured, Semi-structured, Unstructured Data Collection Methods Data Cleaning, Data Transformation, Data Integration Handling Missing Data, Outliers		
Unit-III: Basic Statistics for Analytics Measures of Central Tendency (Mean, Median, Mode) Measures of Dispersion (Range, Variance, Standard Deviation) Correlation and Regression Probability Basics and Distributions (Normal, Binomial)		
Unit-IV: Data Visualization Importance and Principles of Data Visualization Types of Charts: Bar, Pie, Line, Histogram, Heatmap Introduction to tools: MS Excel, Tableau, Power BI, Python (Matplotlib, Seaborn)		
Unit-V: Tools and Technologies Excel for Data Analysis (Functions, Pivot Tables, Charts) Python (Intro to Pandas, Numpy, Matplotlib) Power BI / Tableau: Data Import, Visual Analytics Introduction to R (optional)		
SUGGESTED BOOKS		
1. Data Analytics Made Accessible – A. Maheshwari 2. Python for Data Analysis – Wes McKinney 3. Data Science for Business – F. Provost and T. Fawcett 4. Practical Statistics for Data Scientists – Peter Bruce		

Discipline Centric Core Course (DCC)

DBSCC27002T: Database Systems (SQL + NoSQL Basics)

(20 CIA + 80 EoSE. = Max. Marks: 100)

Course Credits	No. of Teaching Hours Per Week	Total No. of Teaching Hours
3 Credits	3 Hours	45 Hours
Course Outcome: On successful completion of the course, the students will be able to: • Understand the concepts, architecture, and models of database systems. • Design and implement relational databases using ER models and normalization. • Write and execute SQL queries for data retrieval and manipulation. • Understand the need for NoSQL databases and their types. • Explore basic operations in NoSQL databases like MongoDB for semi-structured data.		
SYLLABUS		
Unit-I: Introduction to Database Systems: Definition and purpose of a database, Database system vs. file system, Components of DBMS: DDL, DML, DCL, Architecture: 3-tier view of DBMS, Database users and administrators, Data models: hierarchical, network, relational, object-oriented		
Unit-II: Data Modeling & Relational Model: Entity-Relationship (ER) model: entities, attributes, relationships, ER diagrams: keys, cardinality, participation, Conversion of ER to relational schema, Relational model: schema, instances, keys, integrity constraints, Normalization: 1NF, 2NF, 3NF with examples		
Unit-III: SQL for Data Science: SQL basics: DDL, DML, DCL commands, Querying: SELECT, WHERE, ORDER BY, GROUP BY, HAVING, JOINS: INNER, LEFT, RIGHT, FULL, Subqueries, views, indexes, Aggregate functions: COUNT, AVG, SUM, MIN, MAX, Creating and managing tables, constraints (PK, FK, UNIQUE, NOT NULL)		
Unit-IV: Introduction to NoSQL Databases: Limitations of RDBMS in big data era, What is NoSQL?, Types of NoSQL databases: Key-value, Column-family, Document, Graph, CAP theorem, Comparison: SQL vs NoSQL		
Unit-V: NoSQL with MongoDB: Introduction to MongoDB, Collections and documents, CRUD operations in MongoDB, Querying documents using filter conditions, Indexing and simple aggregations, Use cases of MongoDB in data science projects		
SUGGESTED BOOKS		
1. Elmasri & Navathe – Fundamentals of Database Systems – Pearson 2. Kristina Chodorow – MongoDB: The Definitive Guide – O'Reilly 3. Silberschatz, Korth & Sudarshan – Database System Concepts – McGraw Hill 4. Pramod J. Sadalage & Martin Fowler – NoSQL Distilled – Pearson 5. Ramez Elmasri – Modern Database Management – Pearson		

Discipline Centric Core Course (DCC)

BMLCC27002T: Basics of Machine Learning

(20 CIA + 80 EoSE. = Max. Marks: 100)

Course Credits	No. of Teaching Hours Per Week	Total No. of Teaching Hours
3 Credits	3 Hours	45 Hours
Course Outcome: On successful completion of the course, the students will be able to: • Understand the fundamental concepts and types of machine learning. • Preprocess and visualize data for effective model building. • Apply basic supervised learning algorithms for prediction and classification. • Understand unsupervised learning techniques for pattern discovery. • Evaluate model performance and understand overfitting, underfitting, and cross-validation.		
SYLLABUS		
Unit-I: Introduction to Machine Learning: Definition, goals, and applications of ML, Types of machine learning: supervised, unsupervised, reinforcement learning, ML process and workflow, Difference between AI, ML, and DL, ML tools and libraries: scikit-learn, NumPy, Pandas, Matplotlib		
Unit-II: Data Preprocessing & Visualization: Data collection, cleaning, and wrangling, Handling missing values, outliers, and categorical variables, Feature scaling: normalization and standardization, Data visualization using Matplotlib & Seaborn, Splitting data: training vs testing sets		
Unit-III: Supervised Learning Algorithms: Linear Regression: model, loss function, gradient descent, Logistic Regression, Decision Trees and Random Forests, k-Nearest Neighbors (k-NN), Use cases: sales prediction, email classification, etc.		
Unit-IV: Unsupervised Learning Algorithms: Clustering: k-Means, Hierarchical clustering, Dimensionality Reduction: PCA (basic idea), Use cases: customer segmentation, anomaly detection, Visualization of clusters and reduced dimensions		
Unit-V: Model Evaluation & Validation: Confusion matrix, accuracy, precision, recall, F1-score, ROC-AUC curve, Overfitting and underfitting, Cross-validation techniques, Bias-variance tradeoff		
SUGGESTED BOOKS		
1. Aurélien Géron – Hands-On Machine Learning with Scikit-Learn, Keras & TensorFlow – O'Reilly 2. Tom M. Mitchell – Machine Learning – McGraw Hill 3. Andriy Burkov – The Hundred-Page Machine Learning Book 4. Ethem Alpaydin – Introduction to Machine Learning 5. Jason Brownlee – Machine Learning Mastery with Python (eBook series)		

Discipline Elective Course (DSE)

EDASE27002T: Exploratory Data Analysis (EDA)

(20 CIA + 80 EoSE. = Max. Marks: 100)

Course Credits	No. of Teaching Hours Per Week	Total No. of Teaching Hours
4 Credits	4 Hours	60 Hours
Course Outcome: On successful completion of the course, the students will be able to: • Understand the role and workflow of EDA in the data science lifecycle. • Perform data cleaning, transformation, and preparation. • Use univariate, bivariate, and multivariate analysis techniques. • Create effective data visualizations for pattern discovery. • Draw actionable insights from data using Python libraries and real-world datasets.		
SYLLABUS		
Unit-I: Introduction to Exploratory Data Analysis: What is EDA and its importance in data science, EDA vs Data Preprocessing vs Feature Engineering, Types of EDA: Univariate, Bivariate, Multivariate, Overview of EDA tools (Python: Pandas, NumPy, Matplotlib, Seaborn)		
Unit-II: Data Cleaning and Preparation: Handling missing data: removal, imputation strategies, Identifying and treating outliers, Dealing with duplicates, Encoding categorical variables (label encoding, one-hot encoding), Data type conversions and binning		
Unit-III: Univariate and Bivariate Analysis: Descriptive statistics: mean, median, mode, range, IQR, std dev, Frequency distribution and histograms, Box plots, bar plots, violin plots, Scatter plots, correlation matrix, pair plots, Relationships and dependency between variables		
Unit-IV: Multivariate Analysis and Feature Exploration: Cross-tabulation and group-wise analysis, Heatmaps and joint plots, Multicollinearity and its impact, Dimensionality reduction overview (PCA – concept only), Exploratory feature selection and engineering basics		
Unit-V: Data Visualization and Reporting: Principles of effective data visualization, Visualization tools: Matplotlib, Seaborn, Plotly, Dashboards and storytelling with data, Creating EDA reports using Jupyter Notebook, Case study: EDA on real-world datasets (Iris, Titanic, COVID-19, etc.)		
SUGGESTED BOOKS		
1. Ben Jones – Communicating Data with Tableau (for visualization principles) 2. Wes McKinney – Python for Data Analysis – O'Reilly 3. Alberto Boschetti & Luca Massaron – Python Data Science Essentials – Packt		

Discipline Elective Course (DSE)

CLCSE27002T: Cloud Computing

(20 CIA + 80 EoSE. = Max. Marks: 100)

Course Credits	No. of Teaching Hours Per Week	Total No. of Teaching Hours
4 Credits	4 Hours	60 Hours
Course Outcome: On successful completion of the course, the students will be able to:		
- Understand cloud computing fundamentals, models, and architecture. - Learn about virtualization, cloud infrastructure, and services. - Gain practical skills in using cloud platforms like AWS, Azure, or GCP. - Explore security, governance, and deployment strategies in the cloud.		
SYLLABUS		
Unit-I: Introduction to Cloud Computing Definition and Characteristics of Cloud Computing Cloud Deployment Models: Public, Private, Hybrid, Community Cloud Service Models: IaaS, PaaS, SaaS Benefits and Challenges of Cloud Computing Use Cases and Applications		
Unit-II: Cloud Architecture & Virtualization Cloud Reference Model Cloud Infrastructure: Data Centers, Hypervisors Virtualization: Concept, Types (Full, Para, OS-level) Virtual Machines, Containers (Docker) Resource Management and Scheduling		
Unit-III: Cloud Service Providers Overview of AWS, Microsoft Azure, Google Cloud Platform Cloud Storage: S3, Blob Storage Compute Services: EC2, Azure VMs, GCP Compute Engine Networking in Cloud: VPC, Subnets, Load Balancers		
Unit-IV: Cloud Security and Compliance Security Issues in Cloud Computing Identity & Access Management (IAM) Data Privacy and Protection Legal & Regulatory Compliance (GDPR, HIPAA) Risk Management and Governance		
Unit-V: Cloud Applications and Trends Serverless Computing and Function-as-a-Service (FaaS) Edge and Fog Computing Multi-cloud and Hybrid Cloud Strategies Cloud Automation and DevOps Tools Emerging Trends: AI in Cloud, Kubernetes, FinOps		
SUGGESTED BOOKS		
- Cloud Computing: Principles and Paradigms – Rajkumar Buyya - Cloud Computing: Concepts, Technology & Architecture – Thomas Erl - Mastering Cloud Computing – Rajkumar Buyya, Christian Vecchiola, Thamarai Selvi - AWS Certified Cloud Practitioner Guide – Ben Piper		

Discipline Centric Core Course (DCC)
SMDCC27002P: SQL and MongoDB Lab
(20 CIA + 80 EoSE. = Max. Marks: 100)

Course Credits	No. of Teaching Hours Per Week	Total No. of Teaching Hours
1 Credits	2 Hours	30 Hours
SUGGESTED LABORATORY EXERCISES		
Part A: SQL Assignment 1. Table Creation & Insertion Create the following tables: Table: STUDENTS <pre>sql CopyEdit ROLL_NO INT PRIMARY KEY, NAME VARCHAR(50), COURSE VARCHAR(50), DOB DATE, MARKS INT</pre> Table: COURSES <pre>sql CopyEdit COURSE_ID VARCHAR(10) PRIMARY KEY, COURSE_NAME VARCHAR(50), DURATION INT</pre> Tasks: Insert at least 5 records into each table. Display all records. 2. SQL Queries Write SQL queries for the following: - Display students who scored more than 70 marks. - Find the average marks of students. - List all students in the 'BCA' course. - Display the names of students sorted alphabetically. - Show students whose birthday falls in January. - Perform a join between STUDENTS and COURSES tables. - Count total number of students in each course. - Delete a student with ROLL_NO = 102. - Update marks of student ROLL_NO = 103 by adding 5 extra marks. Part B: MongoDB Assignment 1. Database and Collection Creation Create a database: collegeDB Create a collection: students <pre>json CopyEdit { "roll_no":101, "name":"Amit", "course":"BCA", "dob":"2003-05-10", "marks":75 }</pre> Insert at least 5 such documents. 2. MongoDB Queries		

Perform the following operations:

- a. Display all student documents.
- b. Find students who scored more than 70.
- c. Count total number of BCA students.
- d. Update the marks of student with roll_no 101.
- e. Delete a document where roll_no = 105.
- f. Sort students by name in ascending order.
- g. Find students born after the year 2002.
- h. Display only name and course fields.

Discipline Centric Core Course (DCC)

MLLCC27002P: Machine Learning Lab (Scikit-learn / Colab)

(20 CIA + 80 EoSE. = Max. Marks: 100)

Course Credits	No. of Teaching Hours Per Week	Total No. of Teaching Hours
1 Credits	2 Hours	30 Hours

SUGGESTED LABORATORY EXERCISES

Q1. Data Preprocessing

- Load the **Iris Dataset** using Scikit-learn.
- Display the shape, features, and target names.
- Perform basic **data visualization** using matplotlib or seaborn.

Q2. Train a Classification Model

- Use **K-Nearest Neighbors (KNN)** classifier on the Iris dataset.
- Split the data into train-test sets (70%-30%).
- Train the model and display accuracy on the test set.

Q3. Train a Regression Model

- Load the **Boston Housing Dataset** (or California Housing from `sklearn.datasets`).
- Use **Linear Regression** model.
- Display MSE and R^2 score for test data.

Q4. Train a Decision Tree Classifier

- Use the **Wine Dataset**.
- Train a **DecisionTreeClassifier**.
- Visualize the tree using `plot_tree()` from `sklearn.tree`.

Q5. Model Comparison

- Compare the accuracy of **KNN, Decision Tree, and Logistic Regression** classifiers on the Iris dataset.
- Tabulate the results.

Q6. Bonus Task

- Upload your own CSV dataset.
- Apply any classification or regression algorithm of your choice.
- Evaluate and visualize the model performance.

Discipline Centric Core Course (DCC)

PDACC27002P: Mini Project / Capstone Project in Data Analytics

(20 CIA + 80 EoSE. = Max. Marks: 100)

Course Credits	No. of Teaching Hours Per Week	Total No. of Teaching Hours
4 Credits	8 Hours	120 Hours
SUGGESTED LABORATORY EXERCISES		
Project Workflow / Phases		
1. Problem Identification		
- Choose a domain: Healthcare, Education, Retail, Social Media, Finance, etc. - Identify a problem that can be solved using data. - Define clear project objectives and expected outcomes.		
2. Data Collection		
- Choose reliable data sources: - Public datasets (Kaggle, UCI, data.gov.in) - Web scraping (BeautifulSoup, Scrapy) - APIs (Twitter, OpenWeatherMap) - Ensure data is relevant, recent, and sufficient in volume.		
Data Preprocessing		
- Handle missing data (impute or remove) - Normalize/standardize data - Encode categorical variables - Remove outliers - Data cleaning using tools like Pandas, NumPy - Exploratory Data Analysis (EDA) - Use statistics and visualizations to understand the data.		
Tools: Matplotlib, Seaborn, Power BI, Tableau		
Identify trends, correlations, and patterns.		
Model Building (Optional for Mini, Essential for Capstone)		
Apply suitable machine learning algorithms:		
Classification: Logistic Regression, Decision Trees		
Clustering: K-Means		
Prediction: Linear Regression		
Evaluate model performance using metrics: accuracy, precision, recall.		
Model Building (Optional for Mini, Essential for Capstone)		
Apply suitable machine learning algorithms:		
Classification: Logistic Regression, Decision Trees		
Clustering: K-Means		
Prediction: Linear Regression		
Evaluate model performance using metrics: accuracy, precision, recall.		
Project Documentation		
- Cover: - Introduction - Objective - Methodology - Tools & Technologies - Results		
Conclusion & Future Scope		

Skill Enhancement Course (SEC)

DHTSC27001P: Data Handling Tools: Excel & Google Sheets

(20 CIA + 80 EoSE. = Max. Marks: 100)

Course Credits	No. of Teaching Hours Per Week	Total No. of Teaching Hours
3 Credits	6 Hours	90 Hours
Course Outcome: On successful completion of the course, the students will be able to: • Understand the interface, features, and basic operations of Excel and Google Sheets. • Perform data entry, formatting, cleaning, and organization using spreadsheet tools. • Apply formulas, functions, and logical operators for data analysis. • Use charts, pivot tables, and dashboards for data visualization and reporting. • Automate tasks using basic scripting and collaborative features of Google Sheets.		
SYLLABUS		
Unit-I: Introduction to Spreadsheets: Overview of Excel and Google Sheets, Spreadsheet environment and interface, Creating, saving, and managing spreadsheets, Basic operations: insert, delete, format cells, rows, and columns, Data types and cell references (absolute, relative, mixed)		
Unit-II: Data Entry, Cleaning & Formatting: Data entry tips and keyboard shortcuts, Data validation and drop-down lists, Sorting, filtering, conditional formatting, Removing duplicates, trimming spaces, and handling missing data, Working with dates and text functions		
Unit-III: Formulas, Functions & Logical Operations: Arithmetic and logical operators, Common functions: SUM, AVERAGE, MIN, MAX, COUNT, IF, AND, OR, VLOOKUP, HLOOKUP, INDEX, MATCH, Nested functions and formula troubleshooting, Array formulas (in Google Sheets), Error types: #VALUE!, #DIV/0!, #N/A, #REF!		
Unit-IV: Data Analysis and Visualization: Creating charts: bar, pie, line, combo, Using sparklines and slicers (Excel), Creating Pivot Tables and Pivot Charts, Basic dashboard design principles, Use cases in business/data science reporting		
Unit-V: Collaboration, Scripting & Advanced Features: Sharing and collaborating in Google Sheets, Version history and commenting, Introduction to Google Apps Script and Macros (basic automation), Importing/exporting data (CSV, Excel, Google Drive, etc.), Connecting Sheets to external data (Google Forms, APIs)		
SUGGESTED BOOKS		
1. Lokesh Lalwani – Excel Data Analysis for Beginners – BPB Publications 2. Ben Collins – Google Sheets Data Analysis – Self-Published 3. Michael Alexander & Richard Kusleika – Excel Dashboards and Reports – Wiley		

Skill Enhancement Course (SEC)

DATSC27002P: Data Analytics Tools (Tableau / Power BI / Google Data Studio)

(20 CIA + 80 EoSE. = Max. Marks: 100)

Course Credits	No. of Teaching Hours Per Week	Total No. of Teaching Hours
3 Credits	6 Hours	90 Hours
Course Outcome: On successful completion of the course, the students will be able to: • Understand the need for data visualization and BI in data analytics. • Work with Tableau for interactive dashboards and visualizations. • Use Microsoft Power BI for business reporting and data modeling. • Utilize Google Data Studio for collaborative and real-time dashboards • Compare and apply appropriate BI tools for specific analytics tasks.		
SYLLABUS		
Unit-I: Introduction to BI and Visualization Tools: Role of visualization in data science, Business Intelligence (BI): concepts and applications, Overview of BI tools: Tableau, Power BI, Google Data Studio, Connecting to data sources (CSV, Excel, SQL, Google Sheets, APIs), Data types, hierarchies, filters, and relationships		
Unit-II: Data Analytics with Tableau : Tableau interface and workspace, Connecting and preparing data, Dimensions, measures, calculated fields, Creating bar charts, line charts, pie charts, maps, Filters, parameters, groups, and sets, Dashboards and storytelling features, Exporting and publishing Tableau reports		
Unit-III: Data Analytics with Power BI: Power BI Desktop overview and architecture, Data transformation using Power Query, Data modeling: relationships, star/snowflake schemas, DAX basics: calculated columns, measures, KPIs, Visual elements: slicers, cards, matrix, gauge, Creating dashboards and publishing reports to Power BI Service		
Unit-IV: Google Data Studio for Web-Based Reporting: Introduction to Google Data Studio (Looker Studio), Connecting to Google Sheets, BigQuery, and third-party sources, Building pages, charts, and controls, Filters, parameters, date range selectors, Sharing and embedding reports for collaboration		
Unit-V: Comparative Use Cases and Project Development: Tool comparison: Tableau vs Power BI vs Google Data Studio, Selecting the right tool for the right scenario (enterprise, academia, start-ups), Real-world project use cases (e.g., sales dashboard, health reports, survey analytics), Mini capstone project: Build a dashboard in tool of choice, Project presentation and peer review		
SUGGESTED BOOKS		
1. Ashutosh Nandeshwar – Tableau Data Visualization Cookbook – Packt 2. Brian Larson – Microsoft Power BI: Step by Step – McGraw Hill 3. Google – Looker Studio Help Docs (Online Free Resources)		

Value Added Course (VAC)

ICSVC27002T: Introduction to Cyber Security

(20 CIA + 80 EoSE. = Max. Marks: 100)

Course Credits	No. of Teaching Hours Per Week	Total No. of Teaching Hours
4 Credits	4 Hours	60 Hours
Course Outcome: On successful completion of the course, the students will be able to:		
• Explain the fundamentals of cyber security and its importance. • Recognize various cyber attacks and countermeasures. • Apply basic security principles to protect IT systems. • Analyze security threats in networks and software. • Demonstrate knowledge of legal and ethical issues in cyber security.		
SYLLABUS		
Unit-I: Introduction to Cyber Security Importance and scope of cyber security Threats, vulnerabilities, and attacks Types of attackers (script kiddies, insiders, hacktivists, etc.) Security goals: Confidentiality, Integrity, Availability (CIA)		
Unit-II: Cyber Crimes and Laws Types of cyber-crimes (hacking, phishing, identity theft, etc.) Indian IT Act 2000 and amendments Legal frameworks and cyber law policies Cyber forensics basics		
Unit-III: Network Security Fundamentals Introduction to network security Firewalls, IDS/IPS Cryptography basics: symmetric & asymmetric encryption Secure communication protocols (SSL/TLS, VPNs)		
Unit-IV: System and Application Security Operating system security principles Malware types and prevention Secure coding practices Authentication, authorization, and access control		
Unit-V: Cyber Security Tools and Practices Antivirus, anti-malware, and firewall tools Penetration testing and ethical hacking basics Incident response and disaster recovery Cyber hygiene and best practices for individuals		
SUGGESTED BOOKS		
1. Cybersecurity: The Beginner's Guide by Raef Meeuwisse 2. Principles of Information Security by Michael E. Whitman & Herbert J. Mattord 3. Cyber Security Essentials by James Graham, Richard Howard, and Ryan Olson 4. Information Security by Mark Stamp 5. IT Act 2000 (India) – Government Publications		

Value Added Course (VAC)

EDSVC27002T: Ethics in Data Science

(20 CIA + 80 EoSE. = Max. Marks: 100)

Course Credits	No. of Teaching Hours Per Week	Total No. of Teaching Hours
4 Credits	4 Hours	60 Hours
Course Outcome: On successful completion of the course, the students will be able to:		
• Identify ethical issues in data-driven decision-making. • Explain the importance of privacy, consent, and data protection laws. • Evaluate the fairness and bias in machine learning algorithms. • Demonstrate understanding of responsible AI practices. • Apply ethical principles to case studies and real-world projects.		
SYLLABUS		
Unit 1: Introduction to Ethics in Data Science: What is ethics? Importance in data science Data science life cycle and ethical touchpoints, Ethical frameworks (Utilitarianism, Deontology, Virtue Ethics) Case studies of ethical failures in tech		
Unit 2: Data Privacy and Protection: Personally Identifiable Information (PII) Informed consent, data anonymization, GDPR, HIPAA, Indian IT Act Surveillance and ethical issues in big data		
Unit 3: Fairness, Bias, and Discrimination in Algorithms: Bias in data collection and model design Discrimination in automated decision-making, Fairness metrics and mitigation strategies Case studies: COMPAS, Facial Recognition, Hiring tools		
Unit 4: Accountability, Transparency & Explainability: Black-box models and interpretability Algorithmic accountability Auditing data systems, Role of open data and transparency		
Unit 5: Responsible AI and Governance: Principles of responsible AI AI ethics guidelines (IEEE, OECD, UNESCO, NITI Aayog), Role of data scientists as ethical stewards		
SUGGESTED BOOKS		
1. Cathy O’Neil – <i>Weapons of Math Destruction</i> 2. Virginia Eubanks – <i>Automating Inequality</i> 3. Mike Loukides, Hilary Mason – <i>Ethics and Data Science</i> (O’Reilly) 4. NITI Aayog – <i>Responsible AI for All</i> (Govt. of India Report) 5. Floridi, Luciano – <i>The Ethics of Information</i>		