SCHEME & SYLLABI

for

PG DIPLOMA

In

Data Science and Analytics

(w.e.f. session 2026-2027)

(As per NEP 2020 & UGC Curriculum and Credit Framework for Undergraduate Programmes – CCFUP)

(Choice Based Credit System with Multiple Entry & Exit and NSQF/NCrF Alignment)

COMMUNITY COLLEGE OF SKILL DEVELOPMENT

J. C. BOSE UNIVERSITY OF SCIENCE AND TECHNOLOGY, YMCA, FARIDABAD
SECTOR-6, FARIDABAD, HARYANA-121006

J. C. Bose University of Science and Technology, YMCA, Faridabad

(formerly YMCA University of Science and Technology)

A State Govt. University established vide State Legislative Act No. 21 of 2009

SECTOR-6, FARIDABAD, HARYANA-121006

VISION

"J.C. Bose University of Science and Technology, YMCA, Faridabad aspires to be a nationally and internationally acclaimed leader in technical and higher education in all spheres which transforms the life of students through integration of teaching, research and character building."

MISSION

- To contribute to the development of science and technology by synthesizing teaching, research and creative activities.
- To provide an enviable research environment and state-of-the-art technological exposure to its scholars.
- To develop human potential to its fullest extent and make them emerge as world class leaders in their professions and enthuse them towards their social responsibilities.

Community College of Skill Development (CCSD)

VISION

"To be a centre of excellence in vocational and skill-based higher education that produces competent, self-reliant and industry-ready professionals with sound technical skills, ethical values and a spirit of innovation and lifelong learning."

MISSION

- To impart quality vocational education integrating theory, hands-on practice and on-the-job training aligned with NSQF/NCrF and industry standards.
- To strengthen employability and entrepreneurship through competency-based, outcome-oriented and skill-intensive curricula.
- To foster multiple entry and exit pathways that empower learners to progress seamlessly across academic and vocational qualifications.

ABOUT THE COMMUNITY COLLEGE OF SKILL DEVELOPMENT (CCSD)

The Community College of Skill Development (CCSD) at J.C. Bose University of Science and Technology, YMCA, Faridabad, was established to promote industry-oriented, competency-based education and lifelong learning in alignment with the National Education Policy (NEP) 2020, National Skills Qualifications Framework (NSQF), and National Credit Framework (NCrF). CCSD offers a range of skill-integrated academic programmes at undergraduate, diploma, and post-graduate levels that combine academic knowledge with practical training, laboratory work, projects, internships, and industry engagement to enhance employability and professional competence.

ABOUT THE PROGRAMME

The Post Graduate Diploma in Data Science and Analytics is designed to address the growing demand for skilled professionals in the fields of Data Science, Artificial Intelligence, Machine Learning, Business Analytics, Big Data, Cloud Analytics, and Generative AI. The programme provides a strong foundation in statistics, programming, database systems, data analytics, machine learning, deep learning, computer vision, cloud computing, and modern AI technologies. It emphasizes hands-on learning through laboratory courses, case studies, industry-oriented assignments, and a capstone project that enables learners to apply analytical techniques to real-world datasets and business problems. Developed in accordance with contemporary industry requirements and emerging technological trends, the programme follows an outcome-based and multidisciplinary approach to learning. The curriculum integrates theoretical concepts with practical applications using industry-standard tools, programming languages, databases, cloud platforms, and analytics frameworks. The programme leverages the University's modern computing infrastructure, advanced laboratories, digital learning resources, and industry collaborations to equip learners with the knowledge, technical skills, analytical thinking, and problem-solving abilities required for successful careers in Data Science, Analytics, Artificial Intelligence, Business Intelligence, and related technology-driven domains.

TABLE 1 PROGRAMME EDUCATIONAL OBJECTIVES (PEOs)

PEO1	To equip students with advanced knowledge and practical skills in Data Science, Analytics, Machine Learning, Artificial Intelligence, Database Systems, Big Data Technologies, and Data Visualization for solving real-world business and societal problems.
PEO2	To enable students to collect, process, analyze, and interpret structured and unstructured data using statistical, computational, and analytical techniques for informed decision-making.
PEO3	To develop professional competence, ethical responsibility, teamwork, leadership qualities, and effective communication skills required in multidisciplinary and data-driven environments.
PEO4	To promote lifelong learning, innovation, research aptitude, and adaptation to emerging technologies in Data Science, Artificial Intelligence, and Analytics.

TABLE 2 PROGRAMME OUTCOMES (POs)

PO1	Domain Knowledge – Apply concepts of Data Science, Statistics, Machine Learning and Artificial Intelligence.
PO2	Problem Analysis – Identify, formulate and analyze data-driven problems.
PO3	Design & Development of Solutions – Develop analytical models and data-driven solutions.
PO4	Investigation & Data Analysis – Collect, process, analyze and interpret data using appropriate methodologies.
PO5	Modern Tool Usage – Use programming languages, databases, analytical software and visualization tools

	effectively.
PO6	Ethics & Professional Responsibility – Apply ethical principles, privacy standards and professional practices in data handling.
PO7	Communication & Teamwork – Communicate analytical findings effectively and work collaboratively in multidisciplinary environments.
PO8	Lifelong Learning – Engage in independent learning and adapt to emerging technologies in Data Science and Analytics.

TABLE 3 PROGRAMME SPECIFIC OUTCOMES (PSOs)

PSO1	Apply data science, statistical analysis, machine learning and AI techniques to solve real-world problems.
PSO2	Develop and deploy data-driven solutions using programming, databases, cloud and visualization technologies.

TABLE 4 SUSTAINABLE DEVELOPMENT GOALS (SDGs) ADDRESSED

SDG 4	Quality Education - Promotes inclusive and equitable quality education through advanced learning in Data Science, Artificial Intelligence, Analytics, and emerging digital technologies while fostering digital literacy, critical thinking, analytical skills, and lifelong learning.
SDG 8	Decent Work and Economic Growth - Develops industry-relevant competencies in Data Science, Machine Learning, Artificial Intelligence, Business Analytics, and Cloud Technologies, enhancing employability, innovation, entrepreneurship, and workforce readiness for the digital economy.
SDG 9	Industry, Innovation and Infrastructure - Supports innovation, intelligent systems, digital transformation, data-driven decision-making, advanced analytics, and the development of scalable technological solutions for sustainable industrial and economic growth.
SDG 11	Sustainable Cities and Communities - Facilitates the use of data analytics, artificial intelligence, and predictive modeling for smart city applications, urban planning, public services optimization, transportation management, and sustainable community development.
SDG 12	Responsible Consumption and Production - Encourages responsible use of data resources, efficient computing practices, data governance, ethical AI, and sustainable utilization of digital technologies for informed decision-making and resource optimization.
SDG 16	Peace, Justice and Strong Institutions - Promotes ethical data practices, responsible AI, transparency, privacy protection, cybersecurity awareness, and data-driven governance for accountable and trustworthy institutions.
SDG 17	Partnerships for the Goals - Strengthens collaboration among academia, industry, government, research organizations, and technology partners through project-based learning, industry engagement, internships, and innovation-driven data science initiatives.

TABLE 5 COURSE CODES AND DEFINITIONS (UGC CCFUP CATEGORIES)

L	Lecture
T	Tutorial
P	Practical
DSC	Discipline Specific Course (core vocational/technical)
Project	Project / Field Work

NSQF JOB-ROLE & NATIONAL OCCUPATIONAL STANDARDS (NOS) ALIGNMENT

The Post Graduate Diploma in Data Science and Analytics is designed to develop industry-relevant competencies in Data Science, Machine Learning, Artificial Intelligence, Business Analytics, Big Data, Cloud Analytics, and Generative AI. The programme integrates statistics, programming, databases, analytics, and AI technologies through practical and project-based learning. It prepares learners for emerging job roles such as Data Analyst, Data Scientist, Machine Learning Engineer, Business Intelligence Analyst, and AI Associate. The curriculum emphasizes hands-on skills, analytical thinking, problem-solving, and data-driven decision-making required in modern digital enterprises.

Job Role: AI – Data Scientist
QP Code: SSC/Q8104
NQR Code: QG-05-IT-01526-2023-V1.1-NASSCOM
Occupation: Artificial Intelligence and Big Data Analytics
NSQF Level: 6
Sector: IT-ITeS (NASSCOM Sector Skill Council)
Version: 3.0

NOS Code	NOS Title	Aligned Subject(s)	Justification (Brief)
SSC/N8101 V2.0	Import Data as per Specifications	Foundations of Data Science and Python Programming; Data Science Programming Lab	Covers data acquisition, data sources, data collection methods, file handling, data import techniques, and preparation of datasets using Python and analytical tools.
SSC/N8102 V2.0	Preprocess Data as per Specifications	Foundations of Data Science and Python Programming; Data Analytics and Visualization; Data Science Programming Lab	Develops competencies in data cleaning, transformation, handling missing values, feature preparation, data profiling, validation, and preprocessing required for analytics and machine learning.
SSC/N8103 V2.0	Perform Exploratory Data Analysis as per Specifications	Applied Statistics; Data Analytics and Visualization	Introduces descriptive and inferential statistics, data exploration, visualization techniques, trend analysis, hypothesis testing, and interpretation of analytical results.
SSC/N8104 V3.0	Perform Research and Design of Algorithmic Models	Machine Learning and Predictive Analytics; Deep Learning and Computer Vision	Provides knowledge of machine learning algorithms, model selection, neural networks, deep learning architectures, research methodologies, and algorithm design for solving complex analytical problems.
SSC/N8105 V2.0	Apply Pre-designed Algorithmic Models to Specified Use Cases	Machine Learning and Data Analytics Lab; Artificial Intelligence and Generative AI; Capstone Project in Data Science and Analytics	Develops practical skills in implementing, testing, evaluating, deploying, and optimizing machine learning and AI models for real-world applications and business scenarios.
SSC/N9014 V1.0	Maintain an Inclusive, Environmentally Sustainable Workplace	Artificial Intelligence and Generative AI; Capstone Project in Data Science and Analytics	Promotes ethical AI practices, sustainability awareness, responsible use of technology, professional conduct, inclusivity, and environmental considerations in technology-driven environments.
DGT/VSQ/N0102 V1.0	Employability Skills	Capstone Project in Data Science and Analytics	Develops communication, teamwork, leadership, project management, problem-solving, documentation, presentation, and professional skills required in data science and analytics careers.

J. C. BOSE UNIVERSITY OF SCIENCE AND TECHNOLOGY, YMCA, FARIDABAD
Community College of Skill Development (CCSD)
[PG Diploma Data Science and Analytics | NSQF Level 6]

SCHEME FOR SEMESTER – I

Course Code	Course Title	Teaching Schedule				Teaching Hours			Marks for Sessional	Marks for End Term Examination		Total Marks	Credits	Course Type
		L	T	P	Total	L	P	O/A		Theory	Practical			
DSP-107-V1	Foundations of Data Science and Python Programming	4	0	0	4	60	0	60	40	60		100	4	DSC
DSP-101-V1	Applied Statistics	4	0	0	4	60	0	60	40	60		100	4	DSC
DSP-105-V1	Database Systems and SQL	4	0	0	4	60	0	60	40	60		100	4	DSC
DSP-103-V1	Data Analytics and Visualization	4	0	0	4	60	0	60	40	60		100	4	DSC
DSP-109-V1	Data Science Programming Lab	0	0	4	4	0	60	0	25		25	50	2	DSC LAB
DSP-111-V1	Database and Analytics Lab	0	0	4	4	0	60	0	25		25	50	2	DSC LAB
Total		**16**	**0**	**8**	**24**	**240**	**120**	**240**	**210**	**240**	**50**	**500**	**20**	

J. C. BOSE UNIVERSITY OF SCIENCE AND TECHNOLOGY, YMCA, FARIDABAD
Community College of Skill Development (CCSD)
[PG Diploma Data Science and Analytics | NSQF Level 6]

SCHEME FOR SEMESTER – II

Course Code	Course Title	Teaching Schedule				Teaching Hours			Marks for Sessional	Marks for End Term Examination		Total Marks	Credits	Course Type
		L	T	P	Total	L	P	O/A		Theory	Practical			
DSP-108-V1	Machine Learning and Predictive Analytics	4	0	0	4	60	0	60	40	60		100	4	DSC
DSP-106-V1	Deep Learning and Computer Vision	4	0	0	4	60	0	60	40	60		100	4	DSC
DSP-104-V1	Big Data Technologies and Cloud Analytics	4	0	0	4	60	0	60	40	60		100	4	DSC
DSP-102-V1	Artificial Intelligence and Generative AI	4	0	0	4	60	0	60	40	60		100	4	DSC
DSP-112-V1	Machine Learning and Data Analytics Lab	0	0	4	4	0	60	0	25		25	50	2	DSC LAB
DSP-110-V1	Capstone Project in Data Science and Analytics	0	0	4	4	0	60	0	25		25	50	2	DSC LAB
Total		**16**	**0**	**8**	**24**	**240**	**120**	**240**	**210**	**240**	**50**	**500**	**20**	

DETAILED SYLLABI

FOUNDATIONS OF DATA SCIENCE AND PYTHON PROGRAMMING
DSP-107-V1

Total credits: 4

L	T	P
4	0	0

Theory: 60

Sessional: 40

Course Objectives:
The course aims to provide students with a strong foundation in Data Science and Python Programming. It introduces the data science lifecycle, data-driven decision-making, and fundamental programming concepts using Python. The course enables students to develop problem-solving skills, perform basic data analysis, and prepare for advanced topics in machine learning and analytics.

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

- **CO1:** Explain the fundamentals of data science, data analytics, and their applications in various domains.
- **CO2:** Apply Python programming constructs including variables, operators, control structures, functions, and modules to solve computational problems.
- **CO3:** Utilize Python data structures and file handling techniques for data processing and management.
- **CO4:** Develop Python programs for basic data analysis, visualization, and problem-solving applications.

Course Contents:

Unit-I: Introduction to Data Science

Overview of Data Science, Data Science Lifecycle, Types of Data, Sources of Data, Data Collection and Preprocessing, Data-driven Decision Making, Applications of Data Science in Business, Healthcare, Finance, Education and Industry, Introduction to Python and its Features, Python Installation and Development Environments.

Unit-II: Python Programming Fundamentals

Python Syntax, Variables and Data Types, Input and Output Operations, Operators, Conditional Statements, Loops, Control Flow Statements, Functions, Modules and Packages, Scope of Variables, Built-in Functions.

Unit-III: Data Structures and File Handling

Strings, Lists, Tuples, Dictionaries, Sets, List Comprehensions, String Manipulation, File Handling, Reading and Writing Files, Exception Handling, Introduction to NumPy Arrays and Basic Operations.

Unit-IV: Data Analysis and Visualization Basics

Introduction to Pandas, DataFrames and Series, Data Import and Export, Data Cleaning Concepts, Descriptive Statistics, Data Visualization using Matplotlib, Bar Charts, Line Charts, Pie Charts, Histograms, and Basic Data Analysis Applications.

Practical Exercises:

The learners are required to:

1. Develop Python programs using variables, operators, conditional statements, loops, and functions.
2. Implement programs using lists, tuples, dictionaries, sets, and file handling operations.
3. Perform basic data manipulation and analysis using NumPy and Pandas libraries.
4. Create graphical visualizations and analytical reports using Matplotlib based on real-world datasets.

Suggested Readings:

1. Jake VanderPlas, Python Data Science Handbook.
2. Wes McKinney, Python for Data Analysis.
3. Joel Grus, Data Science from Scratch.
4. Allen B. Downey, Think Python.
5. Reema Thareja, Python Programming: Using Problem Solving Approach.

CO–PO Mapping

Course Outcomes (Cos)	PO1 Domain Knowledge	PO2 Problem Analysis	PO3 Design & Development	PO4 Investigation & Data Analysis	PO5 Modern Tool Usage	PO6 Ethics & Professional Responsibility	PO7 Communication & Teamwork	PO8 Lifelong Learning
CO1	3	2	1	2	1	1	1	2
CO2	3	3	2	1	3	–	–	2
CO3	3	3	2	2	3	–	–	2
CO4	3	3	3	3	3	1	1	3

Correlation: 1 = Low, 2 = Medium, 3 = High; "–" = No Correlation

CO–PSO Mapping

Course Outcomes (Cos)	PSO1: Data Science, Statistics, ML & AI	PSO2: Programming, Databases, Cloud & Visualization
CO1	2	1
CO2	2	2
CO3	2	2
CO4	3	3

APPLIED STATISTICS
DSP-101-V1

Total Credits: 4

L	T	P
4	0	0

Theory: 60

Sessional: 40

Course Objectives:
The course aims to provide students with the fundamental concepts and techniques of statistics required for data analysis and decision-making. It focuses on data collection, descriptive statistics, probability distributions, statistical inference, correlation, regression, and hypothesis testing. The course enables students to apply statistical methods in data science, business analytics, and research applications.

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

- **CO1:** Explain statistical concepts and summarize data using descriptive statistical measures.
- **CO2:** Apply probability theory and probability distributions to analyze uncertain events and data.
- **CO3:** Perform statistical inference, hypothesis testing, and estimation for data-driven decision-making.
- **CO4:** Analyze relationships among variables using correlation and regression techniques.

Course Contents:

Unit-I: Descriptive Statistics

Introduction to Statistics, Types of Data, Data Collection Methods, Classification and Tabulation of Data, Frequency Distribution, Measures of Central Tendency (Mean, Median, Mode), Measures of Dispersion (Range, Variance, Standard Deviation), Skewness and Kurtosis, Data Visualization Techniques.

Unit-II: Probability and Probability Distributions

Basic Concepts of Probability, Conditional Probability, Bayes' Theorem, Random Variables, Probability Distribution Functions, Binomial Distribution, Poisson Distribution, Normal Distribution, Applications of Probability in Data Science and Analytics.

Unit-III

Statistical Inference and Hypothesis Testing: Sampling Techniques, Sampling Distributions, Estimation Theory, Confidence Intervals, Hypothesis Testing, Z-Test, t-Test, Chi-Square Test, Analysis of Variance (ANOVA), Statistical Significance and Decision Making.

Unit-IV

Correlation Analysis, Pearson and Spearman Correlation Coefficients, Simple Linear Regression, Multiple Regression Concepts, Model Evaluation Measures, Prediction and Forecasting, Statistical Quality Control, Control Charts, Sampling Inspection Techniques, Error Detection and Data Validation Methods, Applications of Statistics in Data Quality Management.

Practical Exercises:

1. Compute measures of central tendency and dispersion using real-world datasets.
2. Apply probability distributions to solve analytical and decision-making problems.
3. Perform hypothesis testing and confidence interval estimation using statistical software or spreadsheets.
4. Conduct correlation and regression analysis and interpret the results for predictive decision-making.

DATABASE SYSTEMS AND SQL
DSP-105-V1

Total Credits: 4

L	T	P
4	0	0

Theory: 60
Sessional: 40

Suggested Readings:

1. S. C. Gupta and V. K. Kapoor, Fundamentals of Applied Statistics.
2. Douglas C. Montgomery and George C. Runger, Applied Statistics and Probability for Engineers.
3. Ken Black, Business Statistics for Contemporary Decision Making.
4. David Freedman, Robert Pisani and Roger Purves, Statistics.
5. James T. McClave and Terry Sincich, Statistics.

Course Objectives:

The course aims to provide students with a comprehensive understanding of database systems and Structured Query Language (SQL). It focuses on database design, relational database concepts,

CO–PO Mapping

Course Outcomes (COs)	PO1 Domain Knowledge	PO2 Problem Analysis	PO3 Design & Development	PO4 Investigation & Data Analysis	PO5 Modern Tool Usage	PO6 Ethics & Professional Responsibility	PO7 Communication & Teamwork	PO8 Lifelong Learning
CO1	3	2	1	3	1	–	–	2
CO2	3	3	2	3	1	–	–	2
CO3	3	3	2	3	2	1	–	2
CO4	3	3	3	3	2	1	1	3

Correlation: 1 = Low, 2 = Medium, 3 = High; “–” = No Correlation

CO–PSO Mapping

Course Outcomes (COs)	PSO1: Data Science, Statistics, ML & AI	PSO2: Programming, Databases, Cloud & Visualization
CO1	3	1
CO2	3	1
CO3	3	2
CO4	3	2

normalization, SQL queries, database management, and data retrieval techniques. The course enables students to efficiently store, manage, retrieve, and analyze data for data science and analytics applications.

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

- **CO1:** Explain database concepts, relational database models, and database management system architectures.
- **CO2:** Design relational databases using entity-relationship modeling and normalization techniques.
- **CO3:** Develop SQL queries for data definition, manipulation, retrieval, and database management operations.
- **CO4:** Apply database systems and SQL techniques for data analysis, reporting, and decision-making applications.

Course Contents:

Unit-I: Introduction to Database Systems

Introduction to Database Systems, Database Management Systems (DBMS), File System vs DBMS, Database Architecture, Types of Databases, Data Models, Relational Database Model, Database Users, Database Applications, Introduction to Relational Database Management Systems (RDBMS).

Unit-II: Database Design and Normalization

Entity Relationship (ER) Model, Entities, Attributes, Relationships, ER Diagrams, Mapping ER Models to Relational Schemas, Keys and Constraints, Functional Dependencies, Introduction to Normalization, First Normal Form (1NF), Second Normal Form (2NF), Third Normal Form (3NF), Database Design Principles.

Unit-III: Structured Query Language (SQL)

Introduction to SQL, Data Definition Language (DDL), Data Manipulation Language (DML), Data Query Language (DQL), Data Control Language (DCL), Data Types, Creating and Managing Tables, INSERT, UPDATE, DELETE, SELECT Statements, Filtering and Sorting Data, Aggregate Functions, GROUP BY and HAVING Clauses.

Unit-IV: Advanced SQL and Database Applications

Joins (Inner, Left, Right and Full Joins), Subqueries, Views, Indexes, Constraints, Stored Procedures and Triggers (Concepts), Transactions and ACID Properties, Database Security, Backup and Recovery Concepts, Data Validation Techniques, Data Integrity Constraints, Data Quality Management using SQL, Applications of SQL in Data Analytics and Business Intelligence.

Practical Exercises:

1. Design ER diagrams and convert them into relational database schemas.
2. Create and manage databases and tables using SQL DDL commands.
3. Perform data insertion, updating, deletion, and retrieval using SQL queries and joins.
4. Develop analytical reports using aggregate functions, subqueries, views, and database transactions.

Suggested Readings:

1. Abraham Silberschatz, Henry F. Korth and S. Sudarshan, Database System Concepts.
2. Ramez Elmasri and Shamkant B. Navathe, Fundamentals of Database Systems.
3. Thomas Connolly and Carolyn Begg, Database Systems: A Practical Approach to Design, Implementation and Management.
4. C. J. Date, An Introduction to Database Systems.
5. Coronel and Morris, Database Systems: Design, Implementation and Management.

CO–PO Mapping

Course Outcomes (COs)	PO1 Domain Knowledge	PO2 Problem Analysis	PO3 Design & Development	PO4 Investigation & Data Analysis	PO5 Modern Tool Usage	PO6 Ethics & Professional Responsibility	PO7 Communication & Teamwork	PO8 Lifelong Learning
CO1	3	2	1	1	2	–	–	2
CO2	3	3	3	2	2	–	–	2
CO3	3	3	3	2	3	–	–	2
CO4	3	3	2	3	3	1	1	3

Correlation: 1 = Low, 2 = Medium, 3 = High; "–" = No Correlation

CO–PSO Mapping

Course Outcomes (COs)	PSO1: Data Science, Statistics, ML & AI	PSO2: Programming, Databases, Cloud & Visualization
CO1	2	2
CO2	2	3
CO3	2	3
CO4	3	3

DATA ANALYTICS AND VISUALIZATION
DSP-103-V1

Total Credits: 4

L T P

Theory: 60

4 0 0

Sessional: 40

Course Objectives:
The course aims to provide students with the knowledge and skills required for data analytics and visualization. It focuses on data preprocessing, exploratory data analysis, business analytics, data visualization techniques, dashboard development, and interpretation of analytical results. The course enables students to transform raw data into meaningful insights for informed decision-making.

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

- **CO1:** Apply data preprocessing and exploratory data analysis techniques to prepare datasets for analytics.
- **CO2:** Analyze structured and unstructured data using descriptive and diagnostic analytics methods.
- **CO3:** Create effective visualizations and dashboards to communicate analytical insights.
- **CO4:** Interpret analytical results and support data-driven decision-making in business and organizational environments.

Course Contents:

Unit-I: Introduction to Data Analytics

Introduction to Data Analytics, Types of Analytics (Descriptive, Diagnostic, Predictive and Prescriptive Analytics), Data Analytics Lifecycle, Data Collection and Integration, Data Cleaning and Preprocessing, Handling Missing Values, Data Transformation, Data Quality Dimensions (Accuracy, Completeness, Consistency, Timeliness, Validity and Uniqueness), Data Profiling Techniques, Data Quality Assessment, Data Quality Metrics and KPIs, Introduction to Data Governance, Introduction to Business Analytics.

Unit-II: Exploratory Data Analysis

Exploratory Data Analysis (EDA), Data Summarization Techniques, Descriptive Statistics for Analytics, Outlier Detection, Correlation Analysis, Data Aggregation and Grouping, Feature Understanding, Pattern Identification, Introduction to Analytical Reporting.

Unit-III: Data Visualization Techniques

Principles of Data Visualization, Visual Perception and Design, Charts and Graphs, Bar Charts, Line Charts, Pie Charts, Histograms, Scatter Plots, Heat Maps, Interactive Visualizations, Data Storytelling, Visualization Best Practices.

Unit-IV: Business Intelligence and Dashboard Development

Business Intelligence and Dashboard Development: Introduction to Business Intelligence (BI), Key Performance Indicators (KPIs), Dashboard Design Principles, Dashboard Development using Modern Visualization Tools, Reporting and Analytics, Business Insights Generation, Decision Support Systems, Case Studies in Data Analytics and Visualization.

Practical Exercises:

1. Perform data cleaning, preprocessing, and exploratory data analysis on real-world datasets.
2. Generate statistical summaries and analytical reports using data analytics tools.
3. Create visualizations such as bar charts, line charts, scatter plots, histograms, and dashboards.
4. Develop analytical dashboards and interpret business insights for decision-making applications.

Suggested Readings:

1. Cole Nussbaumer Knaflic, Storytelling with Data.
2. Ben Fry, Visualizing Data.
3. Wilfried Grossmann and Stefanie Rinderle-Ma, Fundamentals of Business Analytics.
4. Wes McKinney, Python for Data Analysis.
5. Cathy O'Neil and Rachel Schutt, Doing Data Science.

CO–PO Mapping

Course Outcomes (Cos)	PO1 Domain Knowledge	PO2 Problem Analysis	PO3 Design & Development	PO4 Investigation & Data Analysis	PO5 Modern Tool Usage	PO6 Ethics & Professional Responsibility	PO7 Communication & Teamwork	PO8 Lifelong Learning
CO1	3	3	2	3	2	–	–	2
CO2	3	3	2	3	2	–	–	2
CO3	3	2	3	2	3	–	2	2
CO4	3	3	2	3	2	1	2	3

Correlation: 1 = Low, 2 = Medium, 3 = High; "–" = No Correlation

CO–PSO Mapping

Course Outcomes (COs)	PSO1: Data Science, Statistics, ML & AI	PSO2: Programming, Databases, Cloud & Visualization
CO1	3	2
CO2	3	2
CO3	2	3
CO4	3	2

DATA SCIENCE PROGRAMMING LAB
DSP-109-V1

Total Credits: 2

L T P — Internal: 25

0 0 4 — External: 25

Course Objectives:
The course aims to develop practical programming skills required for Data Science using Python. It focuses on Python programming fundamentals, data structures, file handling, numerical computing, data manipulation, visualization, and exploratory data analysis. The course enables students to implement data science solutions using industry-standard Python libraries.

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

- **CO1:** Develop Python programs using variables, control structures, functions, and modules.
- **CO2:** Apply Python data structures and file handling techniques for data processing tasks.
- **CO3:** Utilize NumPy and Pandas libraries for data manipulation and analysis.
- **CO4:** Create visualizations and analytical reports using Python-based data science tools.

List of Programs

1. Write Python programs for arithmetic, logical, and relational operations.
2. Implement conditional statements and looping constructs.
3. Develop programs using functions and recursion.
4. Perform string manipulation and pattern generation.
5. Implement list, tuple, dictionary, and set operations.
6. Develop programs using list comprehensions.
7. Read from and write to text and CSV files.
8. Implement exception handling in Python programs.
9. Create and manipulate NumPy arrays.
10. Perform mathematical and statistical operations using NumPy.
11. Import datasets using Pandas DataFrames.
12. Perform data cleaning and preprocessing operations.
13. Handle missing values and duplicate records.
14. Apply filtering, sorting, grouping, and aggregation techniques.
15. Generate descriptive statistical summaries.
16. Create line charts using Matplotlib.
17. Create bar charts and pie charts using Matplotlib.
18. Develop histograms and scatter plots for data exploration.
19. Perform exploratory data analysis on real-world datasets.
20. Prepare analytical reports using Python and data visualization techniques.

Practical Exercises:

The learners are required to

1. Develop Python applications using programming constructs, functions, and data structures.
2. Perform file handling and exception management for data processing applications.
3. Analyze and manipulate datasets using NumPy and Pandas libraries.
4. Create visualizations and exploratory data analysis reports using Python-based tools.

Suggested Readings:

1. Wes McKinney, Python for Data Analysis.
2. Jake VanderPlas, Python Data Science Handbook.
3. Allen B. Downey, Think Python.
4. Reema Thareja, Python Programming: Using Problem Solving Approach.
5. Joel Grus, Data Science from Scratch.

CO–PO Mapping

Course Outcomes (COs)	PO1 Domain Knowledge	PO2 Problem Analysis	PO3 Design & Development	PO4 Investigation & Data Analysis	PO5 Modern Tool Usage	PO6 Ethics & Professional Responsibility	PO7 Communication & Teamwork	PO8 Lifelong Learning
CO1	3	3	3	1	3	–	–	2
CO2	3	3	2	2	3	–	–	2
CO3	3	3	2	3	3	–	–	2
CO4	3	2	2	3	3	–	2	3

Correlation: 1 = Low, 2 = Medium, 3 = High; "–" = No Correlation

CO–PSO Mapping

Course Outcomes (COs)	PSO1 Data Science, Statistics, ML & AI	PSO2 Programming, Databases, Cloud & Visualization
CO1	2	3
CO2	2	3
CO3	3	3
CO4	3	3

DATABASE AND ANALYTICS LAB
DSP-111-V1

Total Credits: 2

L T P — Internal: 25

0 0 4 — External: 25

Course Objectives:
The course aims to provide hands-on experience in relational database management, SQL programming, data extraction, data transformation, and analytical reporting. It focuses on database design, query development, data management, and analytics using SQL and modern database tools. The course enables students to manage and analyze data effectively for business intelligence and data science applications.

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

- **CO1:** Design and implement relational databases using database design principles.
- **CO2:** Develop SQL queries for data definition, manipulation, retrieval, and management operations.
- **CO3:** Apply advanced SQL techniques for data analysis and reporting.
- **CO4:** Perform database-driven analytics and generate meaningful insights for decision-making.

List of Programs

1. Create and manage databases using SQL commands.
2. Create tables with appropriate constraints and data types.
3. Implement primary keys, foreign keys, unique keys, and check constraints.
4. Insert, update, and delete records in database tables.
5. Retrieve data using SELECT statements and filtering conditions.
6. Perform sorting and aggregation operations.
7. Apply SQL aggregate functions (COUNT, SUM, AVG, MIN, MAX).
8. Use GROUP BY and HAVING clauses for analytical reporting.
9. Perform Inner Join, Left Join, Right Join, and Full Join operations.
10. Develop nested queries and correlated subqueries.
11. Create and manage database views.
12. Implement indexes for query optimization.
13. Perform transaction management using COMMIT and ROLLBACK.
14. Create stored procedures and triggers.
15. Design and implement normalized database schemas.
16. Import and export data between databases and CSV files.
17. Perform data cleaning and transformation using SQL.
18. Perform data profiling to identify missing, inconsistent, and duplicate records in datasets.
19. Implement data validation rules and quality checks using SQL queries.
20. Generate data quality reports and dashboards for monitoring data accuracy and completeness.
21. Generate business reports using analytical SQL queries.
22. Analyze datasets using SQL-based statistical functions.
23. Develop mini-projects involving database design and analytics.

Practical Exercises:

The learners are required to

1. Design and implement a relational database system for a real-world application.
2. Develop SQL queries for data retrieval, aggregation, and analytical reporting.
3. Apply joins, subqueries, views, procedures, and transactions for database management.
4. Generate analytical reports and business insights using SQL and database tools.

Suggested Readings:

1. Ramez Elmasri and Shamkant B. Navathe, Fundamentals of Database Systems.

2. Abraham Silberschatz, Henry F. Korth and S. Sudarshan, Database System Concepts.
3. Thomas Connolly and Carolyn Begg, Database Systems.
4. C. J. Date, An Introduction to Database Systems.
5. Coronel and Morris, Database Systems: Design, Implementation and Management.

CO–PO Mapping

Course Outcomes (COs)	PO1 Domain Knowledge	PO2 Problem Analysis	PO3 Design & Development	PO4 Investigation & Data Analysis	PO5 Modern Tool Usage	PO6 Ethics & Professional Responsibility	PO7 Communication & Teamwork	PO8 Lifelong Learning
CO1	3	2	3	1	3	–	–	2
CO2	3	3	3	2	3	–	–	2
CO3	3	3	2	3	3	–	1	2
CO4	3	3	2	3	3	1	2	3

Correlation: 1 = Low, 2 = Medium, 3 = High; "–" = No Correlation

CO–PSO Mapping

Course Outcomes (Cos)	PSO1 Data Science, Statistics, ML & AI	PSO2 Programming, Databases, Cloud & Visualization
CO1	2	3
CO2	2	3
CO3	3	3
CO4	3	3

MACHINE LEARNING AND PREDICTIVE ANALYTICS
DSP-108-V1

Total Credits: 4

L T P — Theory: 60

4 0 0 — Sessional: 40

Course Objectives:
The course aims to introduce students to the concepts, techniques, and applications of Machine Learning and Predictive Analytics. It focuses on data preprocessing, supervised and unsupervised learning algorithms, model evaluation, predictive modeling, and real-world applications. The course enables students to develop intelligent systems and data-driven solutions for prediction and decision-making.

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

- **CO1:** Explain the fundamentals of machine learning and predictive analytics and their applications in various domains.
- **CO2:** Apply data preprocessing, feature engineering, and machine learning algorithms to solve analytical problems.
- **CO3:** Develop and evaluate predictive models using supervised and unsupervised learning techniques.
- **CO4:** Interpret machine learning results and deploy predictive analytics solutions for data-driven decision-making.

Course Contents

Unit-I: Introduction to Machine Learning

Introduction to Machine Learning, Types of Machine Learning, Applications of Machine Learning, Machine Learning Workflow, Data Collection and Preparation, Data Preprocessing, Feature Selection, Feature Engineering, Training and Testing Datasets, Introduction to Scikit-Learn.

Unit-II: Supervised Learning Techniques

Regression and Classification Concepts, Linear Regression, Multiple Linear Regression, Logistic Regression, Decision Trees, Random Forests, K-Nearest Neighbors (KNN), Support Vector Machines (SVM), Model Building and Prediction Techniques.

Unit-III: Unsupervised Learning Techniques

Clustering Concepts, K-Means Clustering, Hierarchical Clustering, Association Rule Mining, Dimensionality Reduction Concepts, Principal Component Analysis (PCA), Pattern Discovery and Customer Segmentation Applications.

Unit-IV: Predictive Analytics and Model Evaluation

Predictive Analytics Lifecycle, Model Evaluation Techniques, Confusion Matrix, Accuracy, Precision, Recall, F1-Score, ROC Curve and AUC, Cross Validation, Overfitting and Underfitting, Model Optimization, Predictive Analytics Applications in Business, Healthcare, Finance, Marketing, and Industry.

Practical Exercises:

1. Perform data preprocessing, feature selection, and exploratory analysis on real-world datasets.
2. Develop regression and classification models using machine learning algorithms.
3. Apply clustering and dimensionality reduction techniques for pattern discovery and segmentation.
4. Evaluate predictive models using performance metrics and generate analytical reports.

Suggested Readings:

1. Aurélien Géron, Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow.
2. Tom M. Mitchell, Machine Learning.
3. Ethem Alpaydin, Introduction to Machine Learning.
4. Andreas C. Müller and Sarah Guido, Introduction to Machine Learning with Python.
5. Trevor Hastie, Robert Tibshirani and Jerome Friedman, The Elements of Statistical Learning.

CO–PO Mapping

Course Outcomes (Cos)	PO1 Domain Knowledge	PO2 Problem Analysis	PO3 Design & Development	PO4 Investigation & Data Analysis	PO5 Modern Tool Usage	PO6 Ethics & Professional Responsibility	PO7 Communication & Teamwork	PO8 Lifelong Learning
CO1	3	2	1	2	2	–	–	2
CO2	3	3	3	3	3	–	–	2
CO3	3	3	3	3	3	–	1	2
CO4	3	3	2	3	3	1	2	3

Correlation: 1 = Low, 2 = Medium, 3 = High; "–" = No Correlation

CO–PSO Mapping

Course Outcomes (Cos)	PSO1 Data Science, Statistics, ML & AI	PSO2 Programming, Databases, Cloud & Visualization
CO1	3	1
CO2	3	2
CO3	3	2
CO4	3	3

DEEP LEARNING AND COMPUTER VISION
DSP-106-V1

Total Credits: 4

L T P — Theory: 60

4 0 0 — Sessional: 40

Course Objectives:
The course aims to introduce students to Deep Learning and Computer Vision techniques used in intelligent systems and modern AI applications. It focuses on neural networks, deep learning architectures, image processing, computer vision algorithms, and real-world applications. The course enables students to develop AI-based solutions for image classification, object detection, and visual analytics.

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

- **CO1:** Explain the fundamentals of deep learning, artificial neural networks, and computer vision systems.
- **CO2:** Apply deep learning techniques and neural network architectures for predictive and classification tasks.
- **CO3:** Develop computer vision models for image processing, image classification, and object recognition.
- **CO4:** Evaluate and deploy deep learning and computer vision solutions for real-world applications.

Course Contents:

Unit-I: Introduction to Deep Learning

Introduction to Deep Learning, Artificial Intelligence, Machine Learning and Deep Learning, Neural Networks, Perceptron Model, Activation Functions, Feed Forward Neural Networks, Backpropagation Algorithm, Gradient Descent, Loss Functions, Deep Learning Frameworks and Applications.

Unit-II: Deep Learning Architectures

Artificial Neural Networks (ANN), Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), Long Short-Term Memory (LSTM) Networks, Transfer Learning, Model Training and Optimization, Hyperparameter Tuning, Regularization Techniques, Introduction to TensorFlow and Keras.

Unit-III: Computer Vision Fundamentals

Introduction to Computer Vision, Digital Image Fundamentals, Image Representation, Image Processing Techniques, Image Enhancement, Image Filtering, Feature Extraction, Edge Detection, Image Segmentation, Computer Vision Workflow and Applications.

Unit-IV: Image Recognition and Advanced Applications

Image Classification, Object Detection, Face Recognition, Image Captioning Concepts, Video Analytics, Optical Character Recognition (OCR), Medical Image Analysis, Autonomous Systems, Computer Vision in Industry, Ethics and Challenges in AI-based Vision Systems.

Practical Exercises:

The learners are required to

1. Develop and train neural network models for classification and prediction tasks.
2. Implement deep learning models using TensorFlow and Keras frameworks.
3. Perform image processing, feature extraction, and image classification using computer vision

techniques.
4. Build and evaluate computer vision applications for object detection, image recognition, and visual analytics.

Suggested Readings:

1. Ian Goodfellow, Yoshua Bengio and Aaron Courville, Deep Learning.
2. François Chollet, Deep Learning with Python.
3. Richard Szeliski, Computer Vision: Algorithms and Applications.
4. Adrian Rosebrock, Deep Learning for Computer Vision with Python.
5. Aurélien Géron, Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow.

CO–PO Mapping

Course Outcomes	PO1 Domain Knowledge	PO2 Problem Analysis	PO3 Design & Development	PO4 Investigation & Data Analysis	PO5 Modern Tool Usage	PO6 Ethics & Professional Responsibility	PO7 Communication & Teamwork	PO8 Lifelong Learning
CO1	3	2	1	2	2	–	–	2
CO2	3	3	3	3	3	–	–	2
CO3	3	3	3	3	3	–	1	2
CO4	3	3	2	3	3	2	1	3

Correlation: 1 = Low, 2 = Medium, 3 = High; "–" = No Correlation

CO–PSO Mapping

Course Outcomes	PSO1 Data Science, Statistics, ML & AI	PSO2 Programming, Databases, Cloud & Visualization
CO1	3	1
CO2	3	2
CO3	3	2
CO4	3	3

BIG DATA TECHNOLOGIES AND CLOUD ANALYTICS
DSP-104-V1

Total Credits: 4

L T P — Theory: 60

4 0 0 — Sessional: 40

Course Objectives:
The course aims to provide students with knowledge of Big Data technologies and Cloud Analytics for managing, processing, and analyzing large-scale datasets. It focuses on Big Data concepts, distributed computing, Hadoop ecosystem, cloud computing platforms, cloud-based analytics, and modern data processing frameworks. The course enables students to develop scalable data solutions for business intelligence and data-driven decision-making.

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

- CO1: Explain the concepts, architecture, and applications of Big Data and Cloud Computing technologies.
- CO2: Apply distributed data processing frameworks and Big Data tools for managing large datasets.
- CO3: Utilize cloud platforms and cloud analytics services for data storage, processing, and analysis.
- CO4: Design scalable Big Data and Cloud-based analytics solutions for real-world applications.

Course Contents:

Unit-I: Introduction to Big Data

Introduction to Big Data, Characteristics of Big Data (5Vs), Sources of Big Data, Big Data Analytics Lifecycle, Challenges of Big Data Management, Distributed Computing Concepts, Big Data Applications in Business, Healthcare, Finance, Social Media, and Industry.

Unit-II: Hadoop Ecosystem and Data Processing

Hadoop Architecture, Hadoop Distributed File System (HDFS), MapReduce Framework, YARN Architecture, Hadoop Ecosystem Components, Hive, Pig, HBase, Introduction to Apache Spark, Batch Processing and Real-Time Data Processing Concepts.

Unit-III: Cloud Computing

Introduction to Cloud Computing, Cloud Service Models (IaaS, PaaS, SaaS), Cloud Deployment Models (Public, Private, Hybrid, Community Cloud), Virtualization Concepts, Cloud Storage and Data Lakes, Cloud Security, Cloud Data Pipelines, Introduction to AWS (Amazon S3), Microsoft Azure (Azure Blob Storage), and Google Cloud Platform (Google Cloud Storage), Serverless Analytics Concepts.

Unit-IV: Cloud Analytics and Modern Data Platforms

Cloud-Based Data Analytics, Data Warehousing Concepts, Data Lakes, Cloud Databases, Analytics as a Service (AaaS), Real-Time Analytics, Stream Processing Concepts, Cloud-Based Machine Learning Services, Business Intelligence in the Cloud, Emerging Trends in Big Data and Cloud Analytics.

Practical Exercises:

The learners are required to

1. Explore Hadoop ecosystem components and perform distributed data storage operations.
2. Process and analyze large datasets using Hadoop or Apache Spark frameworks.
3. Deploy cloud-based storage and analytics services on leading cloud platforms.

4. Develop cloud analytics solutions and generate insights using real-world datasets.

Suggested Readings:

1. Tom White, Hadoop: The Definitive Guide.
2. Seema Acharya and Subhasini Chellappan, Big Data and Analytics.
3. Rajkumar Buyya, Christian Vecchiola and S. Thamarai Selvi, Mastering Cloud Computing.
4. Judith Hurwitz, Alan Nugent, Fern Halper and Marcia Kaufman, Big Data For Dummies.
5. Michael Minelli, Michele Chambers and Ambiga Dhiraj, Big Data, Big Analytics.

CO–PO Mapping

Course Outcomes (Cos)	PO1 Domain Knowledge	PO2 Problem Analysis	PO3 Design & Development	PO4 Investigation & Data Analysis	PO5 Modern Tool Usage	PO6 Ethics & Professional Responsibility	PO7 Communication & Teamwork	PO8 Lifelong Learning
CO1	3	2	1	2	2	–	–	2
CO2	3	3	3	3	3	–	–	2
CO3	3	3	2	3	3	1	–	2
CO4	3	3	3	3	3	1	1	3

Correlation: 1 = Low, 2 = Medium, 3 = High; "–" = No Correlation

CO–PSO Mapping

Course Outcomes (Cos)	PSO1: Data Science, Statistics, ML & AI	PSO2: Programming, Databases, Cloud & Visualization
CO1	2	2
CO2	3	3
CO3	2	3
CO4	3	3

ARTIFICIAL INTELLIGENCE AND GENERATIVE AI
DSP-102-V1

Total Credits: 4

L	T	P
4	0	0

Theory: 60

Sessional: 40

Course Objectives:
The course aims to introduce students to the concepts, techniques, and applications of Artificial Intelligence (AI) and Generative AI. It focuses on intelligent systems, machine learning foundations, large language models, generative models, prompt engineering, AI-powered applications, and ethical considerations. The course enables students to understand and utilize modern AI technologies for solving real-world problems.

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

- **CO1:** Explain the fundamental concepts, techniques, and applications of Artificial Intelligence and Generative AI.
- **CO2:** Apply AI models and Generative AI tools for content generation, problem-solving, and decision support.
- **CO3:** Analyze the working principles of Large Language Models (LLMs), generative models, and AI-powered systems.
- **CO4:** Evaluate ethical, legal, privacy, and societal implications associated with AI and Generative AI technologies.

Course Contents:

Unit-I: Introduction to Artificial Intelligence

Introduction to Artificial Intelligence, History and Evolution of AI, Types of AI, Intelligent Agents, Knowledge Representation, Reasoning and Problem Solving, Search Techniques, Applications of AI in Business, Healthcare, Finance, Education, and Industry.

Unit-II: Machine Learning Foundations and Generative AI

Overview of Machine Learning, Supervised and Unsupervised Learning, Deep Learning Concepts, Neural Networks, Introduction to Generative AI, Generative Models, Foundation Models, Large Language Models (LLMs), AI-powered Applications and Use Cases.

Unit-III: Prompt Engineering and AI Applications

Prompt Engineering Fundamentals, Text Generation, Image Generation, AI-assisted Coding, Chatbots and Virtual Assistants, Content Creation, Document Analysis, Multimodal AI Systems, Introduction to Retrieval-Augmented Generation (RAG), AI Agents, Vector Databases, Generative AI Applications in Data Science and Analytics.

Unit-IV: Responsible AI and Emerging Trends

Ethical AI, Responsible AI Principles, Bias and Fairness in AI Systems, Explainable AI (XAI), Privacy and Security Concerns, AI Governance, Intellectual Property Issues, Fine-Tuning vs Prompt Engineering, LLM Evaluation Concepts, Emerging AI Trends, Future of Generative AI and Intelligent Systems.

Practical Exercises:

The learners are required to

1. Explore AI and Generative AI tools for text, image, and content generation.
2. Design and evaluate prompts for various AI-assisted tasks and problem-solving scenarios.

3. Analyze real-world applications of Generative AI in business, analytics, and decision-making.
4. Prepare case studies on ethical, privacy, and governance issues associated with AI systems.

Suggested Readings:

1. Stuart Russell and Peter Norvig, Artificial Intelligence: A Modern Approach.
2. Melanie Mitchell, Artificial Intelligence: A Guide for Thinking Humans.
3. Bernard Marr, Generative AI in Practice.
4. Tom Taulli, Artificial Intelligence Basics.
5. OpenAI, Google, Microsoft, and Anthropic documentation on Generative AI technologies.

CO–PO Mapping

Course Outcomes (Cos)	PO1 Domain Knowledge	PO2 Problem Analysis	PO3 Design & Development	PO4 Investigation & Data Analysis	PO5 Modern Tool Usage	PO6 Ethics & Professional Responsibility	PO7 Communication & Teamwork	PO8 Lifelong Learning
CO1	3	2	1	2	2	1	–	2
CO2	3	3	3	2	3	1	1	2
CO3	3	3	2	3	3	1	1	2
CO4	2	2	1	2	2	3	2	3

Correlation: 1 = Low, 2 = Medium, 3 = High; "–" = No Correlation

CO–PSO Mapping

Course Outcomes (Cos)	PSO1: Data Science, Statistics, ML & AI	PSO2: Programming, Databases, Cloud & Visualization
CO1	3	1
CO2	3	2
CO3	3	2
CO4	2	1

MACHINE LEARNING AND DATA ANALYTICS LAB
DSP-112-V1

Total Credits: 2

L T P — Practical: 25

0 0 4 — Sessional: 25

Course Objectives:
The course aims to provide practical exposure to machine learning and data analytics techniques using Python-based tools and frameworks. It focuses on data preprocessing, exploratory data analysis, predictive modeling, clustering, model evaluation, and analytical reporting. The course enables students to develop machine learning solutions for real-world data science applications.

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

CO1: Perform data preprocessing and exploratory data analysis using machine learning tools.
CO2: Develop predictive models using supervised machine learning algorithms.
CO3: Apply clustering and dimensionality reduction techniques for data analytics applications.
CO4: Evaluate machine learning models and generate analytical insights from data.

List of Programs

1. Import and explore datasets using Pandas and NumPy.
2. Perform data cleaning, preprocessing, and feature engineering.
3. Handle missing values, outliers, and data transformations.
4. Conduct exploratory data analysis (EDA) using statistical summaries.
5. Visualize datasets using Matplotlib and Seaborn.
6. Implement Linear Regression for predictive analytics.
7. Implement Multiple Linear Regression for business forecasting.
8. Develop Logistic Regression models for classification problems.
9. Apply K-Nearest Neighbors (KNN) classification.
10. Build Decision Tree models for prediction and classification.
11. Implement Random Forest algorithms for predictive analytics.
12. Apply Support Vector Machine (SVM) models.
13. Perform K-Means Clustering for customer segmentation.
14. Implement Hierarchical Clustering techniques.
15. Apply Principal Component Analysis (PCA) for dimensionality reduction.
16. Evaluate models using Confusion Matrix and Classification Reports.
17. Calculate Accuracy, Precision, Recall, and F1-Score.
18. Perform Cross Validation and model comparison.
19. Develop predictive analytics solutions using real-world datasets.
20. Generate analytical reports and visual dashboards from machine learning results.

Practical Exercises:
The learners are required to

1. Perform data preprocessing, visualization, and exploratory data analysis on real-world datasets.
2. Develop regression, classification, and clustering models using machine learning algorithms.
3. Evaluate machine learning models using appropriate performance metrics and validation techniques.
4. Generate analytical reports and predictive insights using machine learning and data analytics tools.

Suggested Readings:

1. Aurélien Géron, Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow.
2. Andreas C. Müller and Sarah Guido, Introduction to Machine Learning with Python.
3. Wes McKinney, Python for Data Analysis.
4. Ethem Alpaydin, Introduction to Machine Learning.
5. Trevor Hastie, Robert Tibshirani and Jerome Friedman, The Elements of Statistical Learning.

CO–PO Mapping

CAPSTONE PROJECT IN DATA SCIENCE AND ANALYTICS

Course Outcomes (Cos)	PO1 Domain Knowledge	PO2 Problem Analysis	PO3 Design & Development	PO4 Investigation & Data Analysis	PO5 Modern Tool Usage	PO6 Ethics & Professional Responsibility	PO7 Communication & Teamwork	PO8 Lifelong Learning
CO1	3	3	2	3	3	–	–	2
CO2	3	3	3	3	3	–	–	2
CO3	3	3	2	3	3	–	1	2
CO4	3	3	2	3	3	1	2	3

Correlation: 1 = Low, 2 = Medium, 3 = High; "–" = No Correlation

CO–PSO Mapping

Course Outcomes (Cos)	PSO1: Data Science, Statistics, ML & AI	PSO2: Programming, Databases, Cloud & Visualization
CO1	3	2
CO2	3	2
CO3	3	2
CO4	3	3

DSP-110-V1

Total Credits: 2

L T P — Practical: 25

0 0 4 — Sessional: 25

Course Objectives:
The course aims to provide students with an opportunity to apply the knowledge and skills acquired throughout the programme to solve real-world data science and analytics problems. It focuses on problem identification, data collection, data preprocessing, model development, analysis, visualization, reporting, and presentation of findings. The course enables students to undertake independent and industry-oriented projects using modern data science tools and techniques.

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

- **CO1:** Identify, formulate, and analyze real-world problems suitable for data science and analytics solutions.
- **CO2:** Apply data collection, preprocessing, statistical analysis, machine learning, and visualization techniques to develop data-driven solutions.
- **CO3:** Design, implement, evaluate, and document a complete data science project using appropriate tools and technologies.
- **CO4:** Communicate project outcomes effectively through technical reports, presentations, and demonstrations.

Project Work
Students shall undertake an individual or group project related to Data Science, Machine Learning, Artificial Intelligence, Business Analytics, Predictive Analytics, Computer Vision, Big Data, Cloud Analytics, or related domains.
The project shall involve the following phases:

1. Problem Identification and Requirement Analysis
2. Literature Survey and Background Study
3. Dataset Collection and Understanding
4. Data Cleaning, Transformation, and Preprocessing
5. Exploratory Data Analysis and Visualization
6. Model Design and Development
7. Model Training, Testing, and Evaluation
8. Result Analysis and Interpretation
9. Version Control and Project Repository Management using GitHub
10. Report Writing and Documentation
11. Project Presentation, Demonstration and Viva-Voce Examination

Practical Exercises:
The learners are required to

1. Identify and formulate a real-world data science problem and prepare a project proposal.
2. Collect, preprocess, analyze, and visualize data using appropriate analytical tools and techniques.
3. Develop, evaluate, and document a complete data science or machine learning solution.
4. Present and demonstrate the project outcomes through technical reports, GitHub repositories, presentations, and viva-voce examinations.

Suggested Readings:

1. Joel Grus, Data Science from Scratch.
2. Jake VanderPlas, Python Data Science Handbook.
3. Aurélien Géron, Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow.
4. Wes McKinney, Python for Data Analysis.
5. Research papers, case studies, and technical documentation relevant to the selected project area.

Note:

1. Only the latest editions of the above books are recommended.

2. Projects should preferably address real-world problems and utilize modern data science, machine learning, AI, big data, or analytics tools.
3. The project report should include objectives, methodology, implementation details, results, conclusions, and future scope.
4. Students shall maintain proper project documentation, source code repositories, datasets, model artifacts, and version control records throughout the project lifecycle.

CO–PO Mapping

Course Outcomes (COs)	PO1 Domain Knowledge	PO2 Problem Analysis	PO3 Design & Development	PO4 Investigation & Data Analysis	PO5 Modern Tool Usage	PO6 Ethics & Professional Responsibility	PO7 Communication & Teamwork	PO8 Lifelong Learning
CO1	3	3	2	3	2	1	1	2
CO2	3	3	3	3	3	1	1	2
CO3	3	3	3	3	3	1	2	3
CO4	2	2	1	2	2	1	3	3

Correlation: 1 = Low, 2 = Medium, 3 = High; "–" = No Correlation

CO–PSO Mapping

Course Outcomes (Cos)	PSO1: Data Science, Statistics, ML & AI	PSO2: Programming, Databases, Cloud & Visualization
CO1	2	2
CO2	3	3
CO3	3	3
CO4	2	2