# Information Booklet cum Syllabus

## Of

## Data Science

## Revision-I

## January 2026

**National Institute of Electronics and Information Technology**
An Autonomous Scientific Society under
Ministry of Electronics and Information Technology, Government of India

NIELIT Gorakhpur
M.M.M.U.T. Campus, Deoria Road
Gorakhpur (U.P.) -273010

## CONTENTS



## 1. About Course

This course is designed to provide a strong foundation in Python programming, data analysis, data visualization, and machine learning. Learners will gain hands-on experience with essential data science tools and techniques used in real-world applications. The course emphasizes practical learning through examples, exercises, and projects.

By the end of the course, students will be able to analyze data, build machine learning models, visualize insights effectively, and apply Python-based solutions to data-driven problems. This program equips learners with the core skills required to begin a career in Data Science using Python.

## 2. NIELIT

National Institute of Electronics and Information Technology, NIELIT, (Erstwhile DOEACC Society) is an autonomous scientific society of the Ministry of Electronics& Information Technology, Government of India. The Society is registered under the Societies Registration Act, 1860. NIELIT was set up to carry out Human Resource Development and related activities in the area of Information, Electronics & Communications Technology (IECT). NIELIT is engaged both in Formal & Non-Formal Education in the areas of IECT besides development of industry-oriented quality education and training programmes in the state-of-the-art areas. NIELIT has endeavoured to establish standards to be the country's premier institution for Examination and Certification in the field of IECT. It is also one of the National Examination Body, which accredits institutes/organizations for conducting courses in IT and Electronics in the non-formal sector.

## 3. Objective of Course

The objective of this course is to equip learners with strong foundations in Python programming, data analysis, machine learning, and advanced data science techniques, along with hands-on experience in real-world applications.

After completing the course, learners will be able to:

- Develop a strong understanding of Python programming fundamentals for data science applications
- Perform data handling, cleaning, preprocessing, and visualization using industry-standard Python libraries.
- Apply basic statistical concepts to interpret and analyze data effectively.
- Use Git and GitHub for version control and collaborative data science workflows.
- Build and evaluate machine learning models using supervised and unsupervised learning techniques.
- Perform predictive and exploratory data analysis using appropriate model evaluation methods.
- Understand the fundamentals, applications, and ethical considerations of Generative AI.
- Work with databases and real-world datasets for data-driven decision making.
- Build, evaluate, and deploy data science models using modern tools.
- Apply Generative AI techniques in practical data science workflows.
- Complete an end-to-end capstone project, demonstrating problem-solving, model development, deployment, and project management using GitHub

## 4. Job Roles of Course

This Course is designed to equip a learner with necessary skills as per following job roles:

- Data Analyst
- Junior Data Scientist
- Machine Learning Engineer (Entry Level)
- Python Developer
- Business Intelligence Analyst

## 5. Eligibility

12th Pass, Basic Knowledge of Computer

## 6. Total duration of the course

180 Hours (Theory: 80 Hrs, Practical/Tutorial: 100 Hrs)

## 7. Course Details

### 7.1. Course Outline and Objective of Each Unit

| S. No. | Module Name | Duration (Theory) in Hours | Duration (Practical) in Hours | Total Learning Hrs. | Learning Objectives |
|---|---|---|---|---|---|
| **1** | **Python & Data Foundations** | 25 | 30 | 55 | • Understand Python programming fundamentals for data analysis.<br>• Perform data handling, cleaning, and preprocessing using Python libraries.<br>• Create meaningful data visualizations for analytical insights.<br>• Apply basic statistical concepts for data interpretation.<br>• Use Git and GitHub for version control and collaborative data science workflows |
| **2** | **Data Science & Machine Learning** | 25 | 35 | 60 | • Work with relational databases using SQL for data analysis.<br>• Build, evaluate, and deploy data science models.<br>• Apply Generative AI techniques in data science workflows.<br>• Implement end-to-end data science project development.<br>• Manage and document projects using GitHub and industry best practices |
| **3** | **Advanced Data Science & Capstone Projects** | 30 | 35 | 65 | • Understand the concepts and types of machine learning.<br>• Apply supervised and unsupervised learning algorithms to datasets.<br>• Perform model evaluation and validation techniques.<br>• Analyze predictive and exploratory machine learning results.<br>• Understand the fundamentals, applications, and ethical considerations of Generative AI |

**7.2. Detailed Course**

| Module Name | Unit | Contents | Hrs. |
|---|---|---|---|
| **Python & Data Foundations** | **Introduction to Python Programming** | • Introduction to programming concepts<br>• Features and applications of Python<br>• Python installation and environment setup<br>• Python syntax, indentation, and comments | **55** |
| | **Python Basics & Control Structures** | • Data types and variables<br>• Operators and expressions<br>• Input and output functions<br>• Conditional statements<br>• Looping constructs (for, while) | |
| | **Data Structures & Libraries** | • Lists, tuples, sets, dictionaries<br>• NumPy arrays and operations<br>• Pandas Series and DataFrames<br>• Data import/export and cleaning | |
| | **Data Visualization, Statistics & Git** | • Data visualization using Matplotlib and Seaborn<br>• Descriptive statistics and probability basics<br>• Introduction to version control<br>• Git commands and GitHub repositories | |
| **Data Science & Machine Learning** | **Data Science Foundations (EDA & SQL)** | • Exploratory Data Analysis (EDA)<br>• Power BI<br>• Tableau<br>• Relational databases<br>• SQL queries, joins, subqueries & aggregate functions | **60** |
| | **Introduction to Machine Learning** | • Definition and scope of machine learning<br>• Types of machine learning<br>• Applications of machine learning<br>• ML workflow | |
| | **Supervised Learning** | • Linear Regression<br>• Logistic Regression<br>• K-Nearest Neighbors (KNN)<br>• Decision Trees and Random Forest<br>• Support vector Machine (SVM)<br>• Naive Bayes | |
| | **Unsupervised Learning & Evaluation** | • Clustering concepts<br>• K-Means and Hierarchical clustering<br>• Dimensionality reduction (PCA)<br>• Model evaluation metrics | |

| | | | |
|---|---|---|---|
| **Advanced Data Science & Capstone Projects** | **Deep Learning Basics** | • Introduction to neural networks<br>• Perceptron and multilayer networks<br>• Overview of deep learning architectures | **65** |
| | **Deep Learning Algorithms** | • Artificial Neutral Network (ANN)<br>• Activation functions and loss functions<br>• Convolutional Neural Networks (CNN)<br>• Recurrent Neural Networks (RNN), LSTM<br>• Autoencoder decoder architecture | |
| | **Introduction to Generative AI** | • Overview of Generative AI<br>• Predictive AI vs Generative AI<br>• Large Language Models (LLMs)<br>• Applications and ethical considerations | |
| | **Capstone Project** | • Problem identification and dataset selection<br>• Data preprocessing and modelling<br>• Model evaluation and deployment/presentation<br>• Project documentation using GitHub | |

## 8. Reference Books/ Study Materials


1) Python for Data Analysis – Wes McKinney

2) Hands-On Machine Learning with Scikit-Learn, Keras & TensorFlow – Aurélien Géron

3) Introduction to Machine Learning with Python – Andreas C. Müller & Sarah Guido

4) Practical Statistics for Data Scientists – Peter Bruce & Andrew Bruce

5) Generative Deep Learning – David Foster

6) Pro Git – Scott Chacon & Ben Straub


## 9. Practical Assignments

**1.** Write a Python program to demonstrate the use of variables, data types, operators, conditional statements, and looping constructs for solving basic computational problems.

**2.** Develop a Python program using sequence data types (List, Tuple, Set, and Dictionary) to perform insertion, deletion, searching, and updating operations.

**3.** Implement a program using NumPy to create arrays and perform array operations such as indexing, slicing, aggregation, and mathematical computations.

4. Using Pandas, perform data analysis on a given dataset by loading data, handling missing values, filtering records, grouping data, and generating summary statistics.

5. Create data visualizations using Matplotlib and Seaborn to represent trends, comparisons, and distributions from a given dataset.

6. Perform statistical analysis on a dataset by calculating mean, median, mode, variance, standard deviation, and interpreting the results.

7. Build a supervised machine learning model (Regression or Classification) using scikit-learn and evaluate its performance using appropriate metrics.

8. Apply an unsupervised learning algorithm (K-Means or Hierarchical Clustering) on a dataset and analyze the identified patterns or clusters.

9. Demonstrate the use of Generative AI tools by designing effective prompts and generating outputs for data summarization, insights, or analytical reporting.

10. Design and implement an end-to-end data science project, including data preprocessing, exploratory data analysis, model development, evaluation, and documentation using GitHub.