Information Booklet cum Syllabus

Of

Data Science

Revision-I

April 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

Sl. No.	Title	Page No.
1.	About the course	2
2.	NIELIT	2
3.	Objective of Course	3
4.	Job Role of Course	3
5.	Eligibility	3
6.	Duration of Course	3
7.	Course Details	4-5
8.	Reference Books/ Study Materials	6
9.	Practical Assignments	6

1. About Course

This course provides a strong foundation in Data Science by integrating key concepts in statistics, programming, and machine learning. Learners gain hands-on experience with tools like Python, Jupyter Notebook, and libraries such as Pandas and Scikit-learn through practical implementation and real-world projects.

By the end of the program, students will be able to clean and analyze data, build and evaluate machine learning models, and communicate insights effectively using data visualization. This course equips learners with the essential skills required to pursue careers in Data Science, Machine Learning, and Data Analytics.

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 endeavored 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 a strong foundation in **Data Science**, combining **statistical analysis**, **programming**, and **machine learning** techniques, along with hands-on experience in solving real-world data problems and building predictive models.

After completing the course, learners will be able to:

- Develop a solid understanding of the data science lifecycle, including **data collection**, **cleaning**, **exploratory data analysis (EDA), modeling**, and **interpretation**.
- Gain proficiency in programming using Python for **data manipulation**, **analysis**, and automation.
- Perform data cleaning and preprocessing using libraries like Pandas and NumPy to handle missing values, outliers, and inconsistencies.
- Conduct **exploratory data analysis** and create insightful visualizations using tools such as Matplotlib and Seaborn.
- Build and evaluate machine learning models using **Scikit-learn** for tasks like **classification**, **regression**, and **clustering**.
- Apply feature engineering and **model optimization** techniques to improve predictive performance.
- Communicate insights effectively through data storytelling and **visualization** to support data-driven **decision-making**.

4. Job Roles of Course

This course is designed to equip learners with the necessary skills for the following job roles:

- Junior Data Scientist
- Machine Learning Engineer (Entry-Level)
- Data Science Trainee
- Predictive Analytics Associate
- Applied Data Science Intern

5. Eligibility

Pursing Diploma /Graduation

6. Total duration of the course

90 Hours (Theory: 35 Hrs, Practical/Tutorial: 55 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 for Data Science**	**6**	**10**	**16**	• Understand Python fundamentals (variables, data types, loops, functions) • Work with libraries like Pandas & NumPy for data manipulation • Perform data cleaning and pre-processing • Visualize data using Matplotlib & Seaborn
2	**Machine Learning**	**8**	**12**	**20**	• Understand supervised & unsupervised learning concepts • Implement algorithms like Linear Regression, Decision Trees, and KNN • Perform model evaluation and validation techniques • Apply feature engineering techniques
3	**Deep Learning**	**6**	**10**	**16**	• Understand neural networks and deep learning concepts • Build ANN models using TensorFlow/Keras • Work with CNNs for image data and basic NLP models • Optimize models using tuning techniques
4	**Power BI**	**10**	**15**	**25**	• Connect and transform data from multiple sources. • Learn to Create interactive dashboards and reports • Use DAX functions and calculated measures • Apply data storytelling techniques for business insights

5	**Career Practicum & Capstone Project**	**5**	**8**	**13**	• Develop an end-to-end data science project from raw dataset • Apply data cleaning, analysis, modeling, and visualization techniques • Understand model deployment basics • Build presentation and storytelling skills for business impact

7.2. Detailed Course

Module Name	Unit	Contents	Hrs.
Python for Data Science	**Python Programming Fundamentals**	• Variables, Data Types, Operators • Control Structures (Loops, Conditionals) • Functions and Modules	**16**
	Data Analysis Libraries	• Data manipulation using **Pandas** • Numerical computing with **NumPy** • Data visualization using **Matplotlib** & **Seaborn**	
Machine Learning	**Supervised Learning**	• **Linear** & **Logistic** Regression • Decision Trees & KNN, SVM • Model evaluation (Accuracy, Precision, Recall)	**20**
	Unsupervised Learning	• Clustering (**K-Means**, Hierarchical) • Dimensionality Reduction (**PCA**) • Feature engineering basics	
Deep Learning	**Neural Networks**	• Perceptron and Multi-layer Networks • Activation Functions & Backpropagation • ANN Model building with TensorFlow/Keras	**16**
	Advanced Deep Learning	• Convolutional Neural Networks (CNN) • Introduction to NLP • Model tuning and optimization	
Power BI	**Data Visualization & BI**	• Data import and transformation • Creating dashboards and reports • DAX functions and calculated fields	**25**
	Business Intelligence & Storytelling	• Interactive visuals and filters • Dashboard design best practices • Presenting insights for decision-making	
Career Practicum & Capstone Project	**Applied Project & Portfolio**	• Develop an end-to-end project using real-world dataset • Perform data cleaning, EDA, and model building • Create dashboards or visual reports	**13**

		• Present insights with business storytelling • Introduction to deployment concepts and best practices	

8. Reference Books/ Study Materials

1. Python for Data Science Handbook – Jake VanderPlas
2. Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow – Aurélien Géron
3. Deep Learning – Ian Goodfellow, Yoshua Bengio & Aaron Courville
4. Practical Statistics for Data Scientists – Peter Bruce & Andrew Bruce
5. Data Science for Business – Foster Provost & Tom Fawcett

9. Practical Assignments

1. Write a Python program to load a dataset and perform basic data exploration, including checking data types, missing values, and summary statistics using Pandas.

2. Perform data cleaning on a dataset by handling missing values, removing duplicates, and correcting inconsistent formats using NumPy and Pandas.

3. Conduct Exploratory Data Analysis (EDA) on a dataset and create visualizations such as histograms, boxplots, and scatter plots using Matplotlib and Seaborn.

4. Build a Linear Regression model using Scikit-learn to predict a continuous variable (e.g., house prices or sales) and evaluate model performance.

5. Implement a Classification model (e.g., Logistic Regression or Decision Tree) to predict categories such as customer churn or loan approval, and measure accuracy, precision, and recall.

6. Perform clustering using K-Means to segment customers based on purchasing behavior and interpret the clusters for business insights.

7. Build a basic Neural Network model using TensorFlow or Keras to classify simple datasets (e.g., handwritten digits or image categories).

8. Create a data visualization dashboard using Power BI or Tableau to present KPIs such as total sales, trends, and category performance.

9. Perform feature engineering on a dataset by creating new meaningful variables and improving model performance through scaling and encoding techniques.

10. Develop an end-to-end mini project: load a real-world dataset, clean and analyze it, build a predictive model, and present insights using visualizations and a summary report.