ABOUT THE COURSE

This is a comprehensive course designed to equip participants with the foundational knowledge and practical skills required to excel in the dynamic field of data science. Throughout the course, participants will delve into key concepts such as mathematical preliminaries, data analysis, visualization, and machine learning, all while harnessing the power of Python programming language. With a blend of theoretical instruction and hands-on exercises, participants will develop proficiency in handling and analyzing large datasets, extracting valuable insights, and making data-driven decisions. This course is designed for beginners seeking to enter the world of data science or a professional looking to enhance his/her skills in the field of data science. This course provides an excellent starting point for mastering the essential tools and techniques used in the industry.

CONTENTS OF THE COURSE

1. Introduction to Python for Data Science
   - Introduction: Tokens, Data Types, Variables, Standard Input Output, Control Flow. Data Structure: List, Tuple, Dictionary, Sets, Strings.
   - Functions, Types of Functions, Recursive Functions, Lambda Functions, Modules, Packages, File Handling, Exception Handling.
   - Object Oriented Programming: Classes, Objects, Methods, Constructors, Inheritance.
   - Numpy, Pandas & Matplotlib.
2. Mathematical Prerequisites for Data Science
   - Linear Algebra: Introduction to vectors, Vector operations, angle, projection, equation of a line, plane, hyperplane, circle, sphere, hypersphere, ellipse, ellipsoid, hyperellipsoid, square, rectangle, hyperrectangle. Vector space and subspace, linear system and subspace, matrix inversion, projection, basis transformations and determinants.
   - Probability and Statistics: Population & Sample, Gaussian Distribution and its PDF, CDF, Symmetric Distribution, Skewness & Kurtosis, Sampling Distribution and Central Limits Theorem, Q-Q Plot, Use of Distributions, Other Distributions, Box-Cox transform, Co-variance, Pearson Correlation Coefficient, Correlation and Causation, Use of Correlations, Confidence Interval, Hypothesis Testing.

3. Data Visualization, Summarization and Plotting
   - Exploratory data analysis: Summarizing Data and Plotting, Scatter plot, Pair plot, Histogram, Probability Density Function (PDF), Univariate analysis using PDF, CDF, Percentiles and Quantiles, Inter-Quartile Range (IQR), MAD, Box-plot with whiskers, Violin plots, Univariate, Bivariate, and Multivariate analysis, contour plot; Outlier Treatment, Measuring Asymmetry: Skewness and Pearson's Median Skewness Coefficient, Kernel Density Estimation.
   - Estimation: Sample and Estimated Mean, Variance and Standard Scores, Covariance, and Pearson's and Spearman's Rank Correlation.
4. Data and Data Science Process
   - Introduction and Motivation to Data Science, Data: Definition, Types and Facets of Data, Data Quality; Data Preprocessing: Aggregation, Sampling, Dimensionality reduction, Feature subset selection, Feature creation, Discretization and Binarization, Variable transformation; Measures of Similarity and Dissimilarity; Data science process, Class of Data Science Problems and Performance Measures.
   - Machine Learning Algorithms: Linear Regression, Logistic Regression and K-Nearest Neighbours (KNN), Decision Tree, Random Forest, Bagging, Boosting and Stacking of ML models.
5. Introduction to GitHub and Projects on Data Science Problems
   - Introduction to git and github. Solving following data science projects (not limited) and uploading into your github account.
   - Data Science Projects (Not limited)
     Classification of Iris Dataset, Custom-made Cancer Diagnosis, Mobile Price Prediction, Delhi Air Quality Index Forecasting, Predicting Diabetes, etc.

COURSE OBJECTIVES

1. To familiarize the participants with the basics of Python for data science.
2. To familiarize the participants with the basics mathematical preliminaries for data science.
3. To familiarize the participants with recent exploratory data analysis and data science process used in industry.
4. To familiarize the participants with Git, GitHub and solving some real-world data science projects.

Short Term Course
on
Data Science using Python
(DSP - 2026)
Hybrid Mode
(Online and Offline)
1st –5th JUNE 2026
Chairman
Prof. Bibhudutta Sahoo, HoD(CS)
Convener
Dr. Prasenjit Dey

Department of Computer Science and Engineering
National Institute of Technology
Rourkela-769008, Odisha
https://www.nitrkl.ac.in/

ABOUT NIT ROURKELA

National Institute of Technology (NIT) Rourkela is an institution of national importance funded by the Ministry of Education. NIT Rourkela was established as Regional Engineering College (REC) on August 15, 1961. In India, it was ranked 13 among engineering colleges by the National Institutional Ranking Framework (NIRF) in 2025. For details about the institute please visit us at https://www.nitrkl.ac.in/.

ABOUT DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

Department of Computer Science & Engineering, NIT, Rourkela was established in 1982. Since its inception, the Department is under dynamic progress and is able to establish the reputation for imparting quality education both at undergraduate and graduate programmes. The department also offers Ph. D. for regular as well as sponsored candidates. Please visit https://www.nitrkl.ac.in/CS/ to know more about the Department of CSE. The department has well equipped modern laboratories such as Software Engineering, Distributed Object Systems, Information Security & Data Communication, Image Processing & Cluster Computing and Advanced Database Engineering Labs for pursuing research keeping in view of the technological advancement.

TARGET PARTICIPANTS

The short-term course is of immense interest for UG/ PG students, research scholars/professionals, staff/ faculty members and industry professionals working in the area of Data Science. The participants from different Science and Engineering (Computer Science and Engineering, Electronics and Communication Engineering, Electrical Engineering, etc.) background will be benefitted with this course.

IMPORTANT DATES

Registration Starts	15th March 2026
Registration Ends	25th May 2026
Maximum Offline Participants (First Come First Serve Basis)	100
Registration Confirmation	27th May 2026
Course Schedule	1st - 5th JUNE 2026

PREREQUISITES

1. The offline participants should bring their laptop.
2. Basics of programming language and data structure will be a plus.

TOURIST PLACES NEARBY

Khandadhar Waterfall

Pitamahal Dam

Vedvyas Temple

Mandira Dam

REGISTRATION FEE PARTICULARS

Registration Fee	
Students	Rs. 1,180/- (online) Rs. 2,360/- (offline)
Faculty from Academic Institutions	Rs. 2,360/-
Employees from Industry and R&D Organizations	Rs. 3,540/-
Accommodation Charges	
Guest house (South / North block) Hostel	As Per Institute Norms

Registration fees include Registration Kit, Refreshment, Tea and Snacks and 18% GST. Lodging, boarding, lunch and dinner facility can be availed on separate payment basis and based on availability.

BANK ACCOUNT DETAILS FOR REGISTRATION

Account Name:	CONTINUING EDUCATION NIT ROURKELA
Account No.:	10138951784
Bank Name	State Bank of India(002109)
Branch:	NIT Rourkela Campus
IFSC Code	SBIN0002109

REGISTRATION FORM

To complete the online registration, the participants need to fill the following google form:
https://forms.gle/2gvcNTbYY2nopXRL9

Patron	Prof. K. Umamaheswar Rao, Director, NIT Rourkela
Chairperson	Prof. Bibhudatta Sahoo
Coordinator	Dr. Prasenjit Dey

Correspondence
Dr. Prasenjit Dey
Assistant Professor & Coordinator, **DSP -2026**
Department of Computer Science & Engineering
National Institute of Technology, Rourkela
Rourkela-769008, Odisha, India
Phone: 9674031771(Mob)
E-mail: **deyp@nitrkl.ac.in**