

TEQIP-III sponsored

Two Weeks

**Short Term Student Training
Programme**

On

“Data Science and Data Analytics”

13th to 23th May, 2020
(Excluding Sunday, 17th May)

Registration Fee & Other Details:

- **Course Registration is free for all participants**
- **Registration is through online mode only.**
- **Use following link for online registration:**

<http://tiny.cc/yo90kz>

Targeted Participants:

- **Final & Pre-final Year (UG students)**

Important Dates:

- Last date of Registration on or before **05-05-2020.**
- Confirmation of participation to selected participants only on or before **06-05- 2020 through E-mail only.**
- **STTP is scheduled from 13-05-2020 to 23-05-2020.**

Chief Patron

Prof.A. M. Rawani

Director, NIT Raipur

Patron

Prof. S. Bajpai

Head, CDC, NIT Raipur

Chairman

Prof. S. Sanyal

TEQIP Coordinator, NIT Raipur

Course Coordinators

Dr.S P Sahu

Assistant Professor, HOD(IT)

Dr. Govind P. Gupta

Assistant Professor

Department of Information Technology

Address for Correspondence

Dr. S. P. Sahu

HoD(IT), Assistant Professor

Department of Information Technology

NIT Raipur (CG) - 492 010

Email: spsahu.it@nitrr.ac.in

Contact No.: +91-9329173783

Dr. Govind P. Gupta

Assistant Professor

Department of Information Technology

NIT Raipur (CG) - 492 010

Email: gpgupta.it@nitrr.ac.in

Contact No.: +91-9891952480

TEQIP-III sponsored

Two Weeks

**Short Term Student Training
Programme**

on

“Data Science and Data Analytics”

13th to 23th May, 2020
(Excluding Sunday, 17th May)



Organized by

Department of Information Technology



**National Institute of Technology
Raipur-492 010 (Chhattisgarh)**

About the Institute:

National Institute of Technology Raipur situated in the capital of Chhattisgarh, has proven to be "avant-grade" in the field of science and technology over past few decades in this region. With sweet memory of foundation ceremony by our president Hon'ble Dr. Rajendra Prasad on 14th September 1956, the institute started with two departments namely Metallurgical and Mining Engineering. Later the inauguration of the Institute building was done by our Prime Minister Hon'ble Pt. Jawahar Lal Nehru on 14th March 1963. From 1st December 2005, the institute has become the National Institute of Technology Raipur. It is well connected with Mumbai, Delhi and all metro cities by regular flights and is on the main Howrah-Mumbai railway route. The institute is 5 km from the Raipur railway station and 18 km from airport on NH-6, the Great Eastern Road.

About Department: The Department of Information Technology came into existence in year 2000. IT department is the study, research and work place of approximately 300 students from the 28 states across India. The department provides an outstanding research environment and offers academic program leading to the award of B.Tech., M.Tech. and Ph.D. degree.

Speakers:

Expert lectures will be delivered by the faculties of the various branches of the institute and the experts from the industries

Objectives of the Course:

Aim of this STTP is to provide up-gradation of knowledge and skills for students/research scholars/faculties/scientists involved in active research in the area of **Data Science and Data analytics**. The course will provide in depth exposure to them. The lectures will be delivered by distinguished experts from IIT/NIT and Software Industries. This **two week short term course** will include lectures and **Special case studies** in the areas of **Data Science and Data analytics** and hands-on practical experiments using **Python data analytic tools** such as **NumPy, SciPy, Pandas, Matplotlib and Scikit-learn**.

Topics to be covered:

- **Introduction to Data Science**
- **Programming Tools for Data Science:** Basics of Python, Toolkits using Python: Matplotlib, NumPy, Scikit-learn, NLTK, Visualizing Data, Working with data.
- **Mathematical Foundations** needed for data science
- **Machine Learning and deep learning algorithms**

Note: A detail course contents are attached with this **Brochure**.

Other Important Information for participants:

- **Course Registration is free for all participants**
- Seats are limited to **50** and the participants will be selected on first come first serve basis.
- Certificates will be issued to the participants only after attending the complete course.
- **No TA/DA will be paid to the participants.**
- A Training / Workshop Kit will be provided to the participants.
- Food Coupon and tea during sessions will be provided to the participants.

Note: It is mandatory for all participants to appear in training assessment and certification through **SSC NASSCOM**.

CURRICULUM: DATA SCIENCE AND DATA ANALYTICS

DURATION

10 Days

60 Hours (40:Lectures [L], 20:Coding sessions [C])

PRE-REQUISITES

- Introduction to Programming
- Probability

OBJECTIVES

The objective of this course is to impart necessary knowledge of the mathematical foundations needed for data science and develop programming skills required to build data science applications.

LEARNING OUTCOMES

At the end of this course, the students will be able to:

- Demonstrate understanding of the mathematical foundations needed for data science.
- Collect, explore, clean, munge and manipulate data.
- Implement models such as k-nearest Neighbors, Naive Bayes, linear and logistic regression, decision trees, neural networks and clustering.
- Build data science applications using Python based toolkits.

DETAIL CONTENTS

1. Introduction to Data Science

L:1Hr

Concept of Data Science, Why/When/What, application in real scenarios.

2. Programming Tools for Data Science

C:6Hr

- a. Basics of Python (file handling, case-folding, spell check, split, strip, Regex, find, replace, etc.)
- b. Toolkits using Python: Matplotlib, NumPy, Scikit-learn, NLTK
- c. Visualizing Data: Bar Charts, Line Charts, Scatterplots
- d. Working with data: Reading Files, Scraping the Web, Using APIs (Example: Using the Twitter APIs), Cleaning and Munging, Manipulating Data, Rescaling, Dimensionality Reduction

3. Mathematical Foundations

L:12Hr

- a. Linear Algebra: Vectors, Matrices
- b. Multivariate calculus
- c. Statistics: Describing a Single Set of Data, Correlation, Simpson's Paradox, Correlation and Causation
- d. Probability: Dependence and Independence, Conditional Probability, Bayes's Theorem, Random Variables, Continuous Distributions, The Normal Distribution, The Central Limit Theorem
- e. Hypothesis and Inference: Statistical Hypothesis Testing, Confidence Intervals, P-hacking, Bayesian Inference
- f. Optimization: Unconstrained, Constrained, KKT conditions

4. Machine Learning

L:25Hr + C:10Hr

- a. Overview of Machine learning concepts – Bias/variance, overfitting and train/test splits.
- b. Types of Machine learning – Supervised, Unsupervised, Semi-supervised.
- c. Classification and Regression algorithms-Naïve Bayes, K-Nearest Neighbors, logistic regression, support vector machines (SVM), decision trees and induction rules, Hidden Markov Models
- d. Metrics
- e. Introduction to Bayes Theorem
- f. Linear Regression-model assumptions
- g. Regularization (lasso, ridge, elastic net) from both the statistical and Bayesian inference viewpoint
- h. Analysis of Time Series
- i. Bagging and Boosting (to balance bias and variance) and random forest
- j. Neural Networks-Learning And Generalization, Overview of Deep Learning
- k. Unsupervised learning: KMeans and Hierarchical clustering
- l. Reinforcement learning

5. Case Studies of Data Science Applications and Project

L:2Hr + C:4Hr

Weather forecasting, Stock market prediction, Object recognition, Real Time Sentiment Analysis, etc.

(SAMPLE) LIST OF PRACTICALS

1. Write a programme in Python to predict the class of the flower based on available attributes.
2. Write a programme in Python to predict if a loan will get approved or not.
3. Write a programme in Python to predict the traffic on a new mode of transport.
4. Write a programme in Python to predict the class of user.
5. Write a programme in Python to identify the tweets which are hate tweets and which are not.
6. Write a programme in Python to predict the age of the actors.
7. Mini project to predict the time taken to solve a problem given the current status of the user.