

Data Science and Data Analytics with NASCOMM Certification

Step 1: Training of Faculty

Dr B Kanmani, ETE, and **Dr Smitha V Shenoy, MBA,** successfully completed the Online Short course on Data Science and Analytics conducted by the Knowledge Network of Indian Institute of Technology Gandhinagar, under the TEQIP-III initiative (MHRD and Government of Gujarat). IIT Gandhinagar, from 11th to 23rd January 2021. The schedule of the two week program is as below:

Day	9 - 10:30 AM	10:30 - 11 AM	11 AM - 1 PM	1 - 2 PM	2- 3:30 PM	3:30 - 4 PM	4 - 5 PM	5 PM - 6 PM
1	Introduction: What is Machine Learning	Break	Python Data Science ecosystem Lab 1	Lunch	Python Data Science ecosystem Lab 2	Break	Python Data Science ecosystem Tutorial 1	
2	Python Data Science ecosystem Lab 3		Python Data Science ecosystem Lab 4		Python Data Science ecosystem Tutorial 2		Python Data Science ecosystem Tutorial 3	
3	Linear Regression 1		Linear Regression 2		Linear Regression Lab		Linear Regression Lab tutorial	
4	Linear Regression theory tutorial		Metrics		Probability		Probability tutorial	
5	Logistic Regression		Logistic regression		Logistic regression theory tutorial		Logistic regression lab tutorial	
6	MLPs		MLPs		MLP Lab		MLP lab tutorial	
7	MLP theory tutorial		Naive Bayes		Naive Bayes Lab		Naive Bayes tutorial	
8	KMeans		Kmeans theory tutorial		KMeans lab		KMeans lab tutorial	
9	SVM 1		SVM 2		SVM theory tutorial		SVM lab	
10	SVM lab tutorial		Decision Trees		Decision Trees		Decision Trees tutorial	
11	Decision Trees lab		Decision Trees lab tutorial		Bias variance		Regularisation	
12	Bias variance and regularisation tutorial		Regularisation lab		Ask Me Anything Sessions		Feedback and closing	

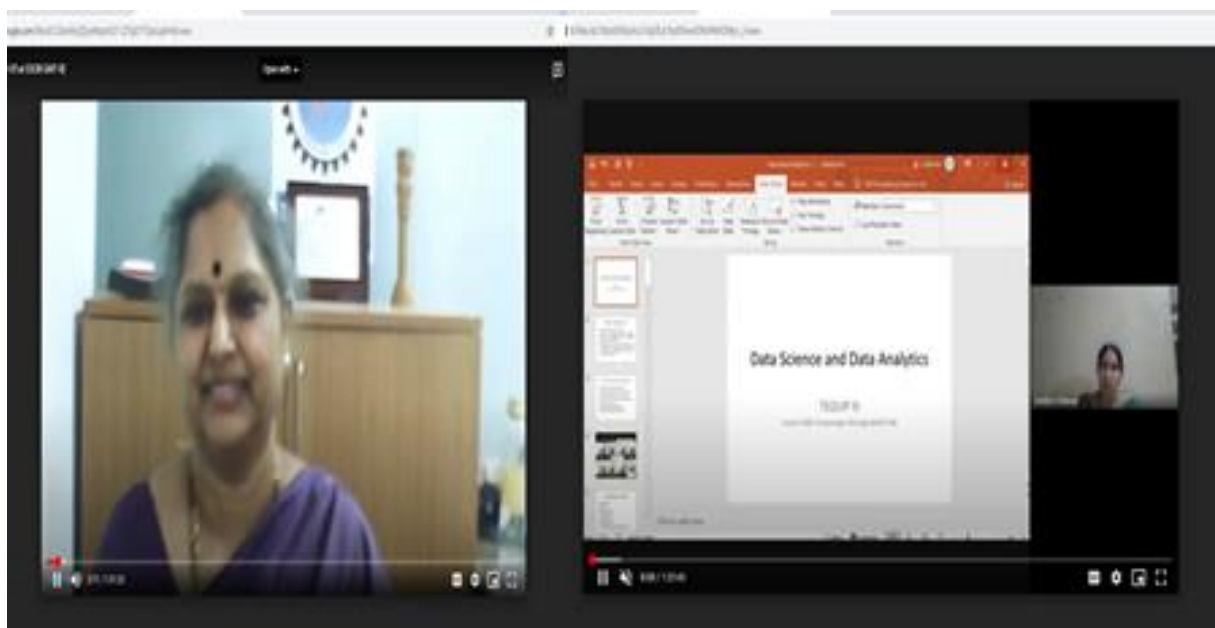
The program was well planned and conducted by experienced resource persons. The course content was shared with all participants of the program, and is available on notion:

<https://www.notion.so/b05dbad55b9f44338a0020044edbc129?v=a6c704dd09524254a84004b12895bc6e>

Step 2: Training of students by the Trained Faculty

As per the TEQIP mandate, Faculty who are trained on Future Skills, conducted by IITs in their respective area through ONLINE mode, had to conduct to further impart job-oriented training to the students of their institutes (at least a batch of 50) to make them industry ready for complying to the future skill requirements of IT-ITes industries, and target a minimum of 50 student registrations, and atleast 25% to qualify the program.

Accordingly, from March 1 to March 27, 2021, **Dr B Kanmani, ETE**, and **Dr Smitha V Shenoy, MBA**, successfully conducted the TEQIP sponsored 60 hours online course on Future Skills on, 'Data Science and Data Analytics', for 180 students registered from BE, MBA, MCA and MTech programs, culminating in the online exam conducted by NASCOMM on 29th March, 2021.



The course content of the 60 hours training to offered to students is as below:

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.

Step 3: Evaluation by NASCOMM

Out of the 180 registered, only 97 took the examination by NASCOMM. It is interesting to observe that out of these 97, who have taken the test, 7 cleared with Gold, 14 cleared with Silver and 15 cleared with Bronze Certification from NASCOMM.

Sample certificate



NASSCOM Assessment of Competence—Technology TEQIP III Mid-Term Assessment

SCORE CARD

Registration ID	NACT0321KAAAA00239
Name	Sai Varnitha R
Date of Birth	12/6/2001
Name of College	BMS College of Engineering, Bangalore, Karnataka Bangalore
Test Location	BMS College of Engineering, Bangalore, Karnataka Bangalore
Test Date	Mar 13 2021

Data Science—TEST SCORE

Skill	Max Score	Your Score	% Score	College Percentile
Introduction to Data Science	20	20	80%	100%
Mathematical Foundations	80	42	56%	20%
Total Score	100	62	62%	70%

Your overall score is 62%

To motivate the committed students, all students who have taken the end test by NASCOMM, received a Certificate of successful completion by TEQIP and IEEE TEMS. Sample certificate included below:



Step 4: Introducing an Open Elective Course in the Curriculum

An additional Outcome, of the having attended the above course by IIT Gandhinagar, is that the UG Curriculum of the Electronics and Telecommunication Engineering Program, has now introduced an Open Elective Course in the VII semester to be offered from August 2021 onwards, with title: **Data Science and Machine Learning**, with course content based on the recommended content by NASCOMM, after approval by the BOS (Board of Studies), held on 27 March, 2021.

(<https://bmsce.ac.in/Syllabus/TE/UG/UG%20Syllabus%202018-22.pdf>)

Course Title	DATA SCIENCE FOR MACHINE LEARNING				
Course Code	19ET7OEDM	Credits	3	L:T:P	3:0:0
UNIT I				[8 hours]	
Introduction to Data Science. Functions (polynomial functions, exponential, sinusoidal); integration and differentiation of functions. To plot the functions, and perform integration/differentiation using Python					
UNIT II				[8 hours]	
Probability: Dependence and Independence, Conditional Probability, Bayes's Theorem, Poisson Distribution, Binomial Distribution Random Variables, Continuous Distributions, The Normal Distribution, The Exponential distribution, The Uniform Distribution, The Central Limit Theorem To plot the given distribution, and compute the Mean, variance, standard deviation using Python					
UNIT III				[8 hours]	
Introduction to Machine learning concepts – Bias/variance, over-fitting and train/test splits. Types of Machine learning – Supervised, Unsupervised, Semi-supervised, Classification and Regression algorithms, Linear Regression, Logistic Regression algorithms, the concept, and implementation using Python.					
UNIT IV				[8 hours]	
The Naïve Bayes Classifier for Discrete and Continuous Input; Decision Trees (for Discrete and Continuous Input and Output); the concept, and implementation using Python.					
UNIT V				[8 hours]	
Kmeans Clustering, Regularization (lasso, ridge). Support vector machines (SVM), Analysis of Time Series; Bagging and Boosting (to balance bias and variance) and random forest. Introduction to Neural Networks. Data Science Applications: To implement two or more relevant Machine Learning models on a given data set, and make a comparative study.					
Unit Choice: Unit III and Unit V					
Course Outcomes					
CO1	Ability to understand the data science concepts				
CO2	Ability to apply the knowledge of Engineering mathematics and programming skills to develop efficient machine algorithms in data science				PO1(3)
CO3	Ability to analyse the regression and classification models				PO2(2)
CO4	Ability to design a solution for data science application				PO3(3)
CO5	Ability to work as an individual and thereby conduct experiments using matlab/python for a given application/problem statement.				PO5(3) PO9 (3)
CO6	Develop, test, analyse and demonstrate applications using python through a mini-project				PO4(3) PO5(3) PO11(3)

TEXT BOOKS:

Data science from scratch (first principles with python) by Joel Grus, O'Reilly, April 2015, 1st edition.

REFERENCE BOOKS:

Doing data science (straight talk from the front line) by Rachel Schutt and Cathy O'Neil, O'Reilly, October 2013, 1st edition.

MOOCs: Machine Learning by Andrew NG, Coursera,
<https://www.coursera.org/learn/machine-learning>