# DATA SCIENCE FOR ENGINEERS

**PROF. RAGHUNATHAN RENGASAMY**
Department of Chemical Engineering
IIT Madras

**PROF. SHANKAR NARASIMHAN**
Department of Chemical Engineering
IIT Madras

**PRE-REQUISITES :** 10 hrs of pre-course material will be provided, learners need to practice this to be ready to take the course.

**INTENDED AUDIENCE :** Any interested learner

**INDUSTRIES APPLICABLE TO :** HONEYWELL, ABB, FORD, GYAN DATA PVT. LTD.

## COURSE OUTLINE :

This course will Introduce: R as a programming language, mathematical foundations required for data science, the first level data science algorithms, data analytics problem solving framework, practical capstone case study Course will also describe a flow process for data science problems (Remembering), Classify data science problems into standard typology (Comprehension), Develop R codes for data science solutions (Application) , Correlate results to the solution approach followed (Analysis), Assess the solution approach (Evaluation) , Construct use cases to validate approach and identify modifications required (Creating).

## ABOUT INSTRUCTOR :

Prof. Rengaswamy was a professor of Chemical Engineering before joining at IIT Madras as a professor and Co-director of the Process Control and Optimization Consortium at Texas Tech University, Lubbock, USA. He was also a Professor at Clarkson University, USA and Assistant Professor at IIT Bombay. His major research interests are in the areas of fault detection and diagnosis and development of data science algorithms for manufacturing industries. Prof. Shankar Narasimhan is currently a professor in the department of Chemical Engineering at IIT Madras. His major research interests are in the areas of data mining, process design and optimization, fault detection and diagnosis and fault tolerant control. He has co-authored several important papers and a book titled Data Reconciliation and Gross Error Detection: An Intelligent Use of Process Data which has received critical appreciation in India and abroad.

## COURSE PLAN :

**Week** 1: Course philosophy and introduction to R

**Week** 2: Linear algebra for data science

**Week** 3: Statistics

**Week** 4: Optimization

**Week** 5: Typology of data science problems and a solution framework

**Week** 6: Simple linear regression and verifying assumptions used in linear regression, Multivariate linear regression, model assessment, assessing importance of different variables, subset selection

**Week** 7: Classification using logistic regression

**Week** 8: Classification using kNN and k-means clustering