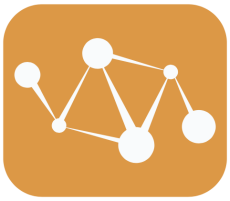

DataScienceWorks
Training Institute

Machine Learning Engineering

Available from **Feb 2022**

Instructor:	Dr. Nayyar Zaidi	**Time/Place:**	TBA
		Duration:	1-2 Days
		Price:	Ask for the quote
		Contact:	info@datascience-works.com

Description:

This comprehensive course provides you with a complete introduction to Machine Learning Engineering. Machine Learning Engineering is the marriage of Big Data and Machine Learning Analysis, and one of the game changers in todays analytics world. The course is ideal for someone with basic understanding of Machine Learning and Databases. It provides an overview of NoSQL, the need for NoSQL, standard NoSQL database. The course develop the theory behind big data platforms and later delve into Apache Spark. It covers various cloud computing platforms such as GCP, AWS and Azure. Importantly, it discusses the engineering aspects of various Machine Learning modelling, e.g, evaluating and deploying the models as well as monitoring their performance.

Training Learning Outcome (TLO):

After the training, you are expected to:

- Have a good understanding of big data technologies and their underlying algorithms.
- Be able to comprehend, communicate, disseminate complex Big Data concepts.
- Have a basic to expert understanding of inner workings of various Big Data technologies.
- Be able to engineer a Machine Learning solution to a given problem.

Target Audience:

The training is for any aspiring or seasoned data scientists, and is perfect for:

- Computer Scientists and I.T Professionals,
- Engineers (Electrical, Mechanical, Industrial, etc),
- First year Ph.D. students in any field looking to break in Data Science,
- Post-doc fellows and Early Career Researchers in any field.

Duration:

The course is expected to be delivered in 2 days (9-5pm), but the duration can be adjusted based on audience experience and background.

Outline:

The following outline is tentative, and can be customized based on audience demand.

	Day 1	Day 2	Day 3
Session 1	Introduction	Machine Learning Engineering I	Introduction to GCP
Session 2	Big Data Engineering	Machine Learning Engineering II	Introduction to AWS
Session 3	Large Scale ML Platforms	Machine Learning with GPUs	Introduction to Azure
Session 4	MLE Lab 1	MLE Lab II	MLE Lab III

Table 1: 2 Days – Training Outline (Day 3 is optional).

Let us delve deep into the details (outline of topics covered) of each session in the following.

Introduction:

- Machine learning, Artificial Intelligence, Statistics, Data Mining and More
- Linear/Logistic Regression
- Optimization
  - Gradient Descent, Stochastic Gradient Descent, etc.
- Model Selection
  - Regularization
  - Feature Engineering

Big Data Engineering:

- Databases - Management, Query, Indexing, Control
- Relational Algebra
- Brief overview of Business Intelligence
  - Data Cubes, Marts
  - Data Warehouses
- Introduction to Big Data
  - SQL vs. NoSQL
- Parallel Query Processing
- Google Big Table, GFS, HDFS, Hadoop
- Hadoop Echosystem
- NoSQL Databases
  - Hbase
  - MongoDB

Large Scale ML Platforms:

- Introduction Apache Spark
- Spark Internals
  - Diving Deep into Spark
  - Spark runtime and application architecture
  - Spark RDDs
  - Spark as distributed file systems
- Spark SQL architecture
  - Spark SQL
- Data Processing with Spark
- Introduction to Yarn
- Introduction to Streaming Architecture
  - Kafka

Machine Learning with GPUs (E)

- Introduction to CUDA
- GPUs Set-up
- Running Deep Learning Models on GPUs

Machine Learning Engineering I:

- Introduction
- Cloud Architecture
  - Cloud Computing
  - Available Resources: AWS, Google Cloud, etc.
- Enviroments
  - Docker
  - Kubernetes
- Model Training
  - TF Serving
  - Training Models on GPUs
  - Training Models across Multiple Devices
- Model Deployment

Machine Learning Engineering II:

- Introduction
- Building Models
- Bad Data
- Evaluating Models
- Error Analysis
- Handling Distribution Shift
- Analysis in Production

About the Instructor:

Dr. Nayyar Zaidi is the lead Data Scientist at DataScienceWorks and a Senior Lecturer of Computer Science at Deakin University. He received the B.S. degree in computer science and engineering from the University of Engineering and Technology, Lahore, in 2005, and the Ph.D. degree in Artificial Intelligence from Monash University, Melbourne, Australia, in 2011. He worked as a Research Fellow, a Lecturer, and a Research Fellow,

from 2011 to 2013, from 2013 to 2014, and from 2014 to 2017, respectively, at the Faculty of Information Technology, Monash University. From 2017 to 2019, he worked as Research Scientist at Credit AI (Trusting Social) Melbourne Lab. His research interests include feature engineering, data generation explainable models and ethical AI. He is also interested in practical (applied) data science, machine learning engineering, and data science training.

About this Document: