

Introduction to Data Engineering

Learn via: **Classroom / Virtual Classroom / Online**

Duration: **1 Day**

Overview

The first challenge in the machine learning life cycle is understanding the problem or opportunity; the next is acquiring, understanding, and preparing data for the modeling phase. This second step is estimated take more than 50% of the time allotted for a machine learning project. This course addresses how to translate the problem statement, identify data sources, explore data for relationships and patterns, identify the starting inputs for the model, prepare data, and validate it for the model fitting process.

Prerequisites

There are no prerequisites for this course.

What You Will Learn

- Understand the data science project methodology
- Understand data source identification (i.e., sources aligned with the problem model)
- Evaluate data findings to determine and validate modeling techniques
- Review feature selection techniques
- Understand data preparation techniques (cleansing, formatting, and blending approaches)
- Plan for data pipelines (proactive and reusable data preparation)
- Understand data visualization techniques for data understanding and data preparation