

Introduction to Data Science for PI System Professionals

Course Description

This course introduces PI System professionals to the fundamentals of data science, focusing on how to harness PI data for analytics, predictive modeling, and AI-driven insights. Participants will learn essential concepts, tools, and techniques to transform raw operational data into actionable intelligence.

Duration : 5 Days

Pre-requisites

- Familiarity with PI System data structures and interfaces
- Basic knowledge of statistics and mathematics
- Exposure to scripting languages (Python, R, or SQL)
- Understanding of industrial operations and data collection

Learning Objectives

By the end of this course, participants will be able to:

- Explain the role of data science in industrial operations
- Apply statistical and machine learning techniques to PI data
- Build pipelines for data preparation, cleaning, and transformation
- Visualize and interpret PI data using modern analytics tools
- Design predictive models for anomaly detection and forecasting
- Integrate data science workflows with PI System environments

Content Coverage

Day 1: Foundations of Data Science & PI System Context

Module 1: Introduction to Data Science

- What is data science?
- Key components: statistics, machine learning, AI

- Data science lifecycle

Module 2: PI System Data Landscape

- PI System architecture overview
- Types of data: time-series, events, asset data
- Challenges in industrial data science

Module 3: Data Science Use Cases in PI Systems

- Operational efficiency
- Predictive maintenance
- Energy optimization

Day 2: Data Preparation & Exploration

Module 4: Data Acquisition & Cleaning

- Extracting PI data for analysis
- Handling missing values and noise
- Data normalization and scaling

Module 5: Exploratory Data Analysis (EDA)

- Statistical summaries of PI data
- Identifying patterns and anomalies
- Correlation and causation in operational datasets

Module 6: Hands-on Lab

- Connecting Python/R to PI data
- Performing EDA on sample datasets

Day 3: Statistical & Machine Learning Techniques

Module 7: Statistical Methods for PI Data

- Descriptive and inferential statistics
- Time-series analysis basics
- Hypothesis testing

Module 8: Machine Learning Fundamentals

- Supervised vs unsupervised learning
- Regression and classification models
- Clustering for operational data

Module 9: Hands-on Lab

- Building regression models with PI data
- Applying clustering to sensor datasets

Day 4: Advanced Analytics & Visualization

Module 10: Predictive Modeling

- Forecasting with time-series models
- Anomaly detection in industrial processes
- Case studies: predictive maintenance

Module 11: Data Visualization & Reporting

- Visualizing PI data with Python libraries (Matplotlib, Seaborn)
- Dashboards with Power BI and PI Vision
- Communicating insights effectively

Module 12: Hands-on Lab

- Creating dashboards for PI data monitoring
- Implementing anomaly detection workflows

Day 5: Integration, Governance & Capstone Project

Module 13: Integrating Data Science with PI Systems

- Embedding models into PI workflows
- Real-time monitoring and alerting
- Automation with scripting and APIs

Module 14: Governance & Best Practices

- Data privacy and compliance in industrial analytics
- Model validation and explainability
- Scaling solutions across enterprise environments

Module 15: Capstone Project

- End-to-end data science workflow using PI data