PetroSkills

Introduction to Machine Learning/Data Analytics for Subsurface Engineering and Geoscience Applications - IMLD - eLearning course

COURSE

About the Course

The interpretation of rich, heterogeneous and even real-time data has become possible because of recent advances in machine learning and the broader availability of computational power. The oil and gas industry is harnessing the power of this data-driven revolution to create actionable insights from real-time production, drilling, and completions data, SCADA data streams, 3D and 4D seismic, well data such as cores, well-logs, thin-sections and SEM images and even the advent of newer data type such as DTS/DAS measurements.

This blended course introduces the concepts of exploratory data analyses, machine learning workflows, and most importantly, data analytics and machine learning use cases for subsurface applications.

This program is comprised of the following PetroAcademy® Skill Modules™. Each module averages approximately 4 hours of self-paced online learning activities.

There is also an in-classroom instructor-led version of this course - see details

Target Audience

Geoscientists, petrophysicists, engineers, or anyone interested in subsurface engineering and geoscience applications of machine learning and data analytics

You Will Learn

- Essential terminology specific to data analytics and machine learning
- Data type and reporting protocols in the oil and gas industry
- Practical approaches to ensuring and verifying data quality
- Exploratory data analyses to visualize and quantify relationships as well as identifying outliers
- The basic principles of common machine learning tools in the petroleum industry
- Unsupervised learning
- Supervised learning
- Reinforcement learning
- Use cases of subsurface geoscience and engineering data-driven applications
- Recognize and address pitfalls of data-driven methods in the oil and gas industry

Course Content

Introduction to Data-driven Workflows

This module introduces data-driven modeling, including its connection to machine learning. We will examine the rising applications of machine learning in different sectors of the economy and how this impacts daily life. Learners will then see how the principles and effects of machine learning are transforming work in the oilfield, focusing on the various applications of data-driven modeling and where this can make operations more efficient and profitable.

You will learn how to:

- Define and describe machine learning
- Discuss the adoption of machine learning and data-driven modeling in our industry, including potential strengths and obstacles
- Identify the modes of machine learning and what distinguishes each
- Recognize the main types of supervised learning
- Conceptualize applications of supervised learning
- Describe unsupervised learning and what distinguishes it from supervised learning
- Conceptualize applications of unsupervised learning
- Identify different data types
- Recognize sampling methods and their pitfalls
- Be able to interpret various measures of univariate statistics:
  - Measures of central tendency
  - Measures of spread
  - Visual representations of data
  - Handling of outliers

Supervised Machine Learning

This skill module introduces supervised learning as a key type of machine learning that drives data-driven analysis across economic sectors and impacts the experiences of consumers. The skill module focuses on the emerging uses of supervised learning in the oil and gas industry as an important complement to other forms of analysis, as well as subject matter expertise, in solving diverse problems and providing reliable data streams.

The skill module begins by placing supervised learning among the three forms of machine learning and explains its distinguishing qualities. The two key forms of supervised learning, regression and classification, are examined in detail with diverse examples from daily life to the technical work in the oilfield. The skill module discusses how supervised learning models are trained to fit data and subsequently validated for deployment. The validation procedure discusses procedures to balance model complexity and model predictability to avoid overfitting and to obtain optimal model performance. The skill module discusses data pre-processing steps including exploratory data analysis, scaling, and an assessment of correlation. Significant emphasis on the appropriate choice of performance metrics for regression and classification problems is also provided. Finally, the skill module reviews emerging uses of supervised learning in the oilfield. A case study approach shows basic and more complex applications, including studies from leading experts in the field.

You will learn how to:

- Distinguish between two forms of supervised learning: regression and classification
- Recognize use cases for regression and classification
- Identify why an iterative approach is essential in supervised learning
- Recognize covariance and correlation as key aspects of data pre-processing, and track their importance for supervised learning
- Recognize a generalized workflow for supervised learning
- Identify and explain the steps involved in exploratory data analysis
- Recognize the need for, and some of the nuances involved in, handling outliers
- Recognize the purpose of scaling
- Distinguish between Standard and Min-Max scaling
- Identify how to apply both the Standard and Min-Max methods
- Recognize how performance metrics are used to evaluate regression models
- Build awareness of the need to fit models and metrics to the specifics of datasets and data problems
- Identify approaches to more complex cases of model evaluation
- Recognize that there is no universal algorithm that can be effectively used

Unsupervised Machine Learning and Clustering

This skill module introduces unsupervised learning as a key type of machine learning that streamlines the extraction of information from raw data that can be very high dimensional, noisy, and heterogeneous. The skill module begins by placing unsupervised learning among the three forms of machine learning and explaining its distinguishing qualities. Unsupervised data analyses are shown to primarily comprise two goals: either pattern identification or dimensionality reduction. In the case of pattern identification, the objectives can be two-fold. The most common application is to condense large datasets into meaningful clusters that contain data points that share similar characteristics.

A second application is related to anomaly detection. This skill module shows that this can be challenging when dealing with multivariate data. In either case, tuning the algorithm to choose the appropriate number of clusters and balancing cluster homogeneity with inter-cluster differences is important. The skill module also discusses data pre-processing steps, including exploratory data analysis and scaling. A discussion of one of the approaches to clustering is provided to enable the participant to see unsupervised learning in action. Finally, the skill module reviews the uses of supervised learning in the oilfield. A case study approach shows basic and more complex applications, including studies from leading experts in the field.

You will learn how to

- Increase awareness of the purposes and benefits of unsupervised learning
- Dig into how unsupervised learning works, including clustering and dimensionality reduction
- Assess the requirements for proper clustering or grouping of data
- Recognize how unsupervised learning and clustering are applied in the oilfield

Product Details

Categories: Upstream

Disciplines: Data Management, Science and Analytics   Petrophysics

Levels: Foundation

Product Type: Course

Formats Available: On-Demand

Instructors: Deepak Devegowda

On-Demand Format

| Course | On-Demand (Available Immediately) $1,100.00