PetroSkills

**Unsupervised Machine Learning and Clustering**

MODULE

About the Skill Module

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.

See online learning demo
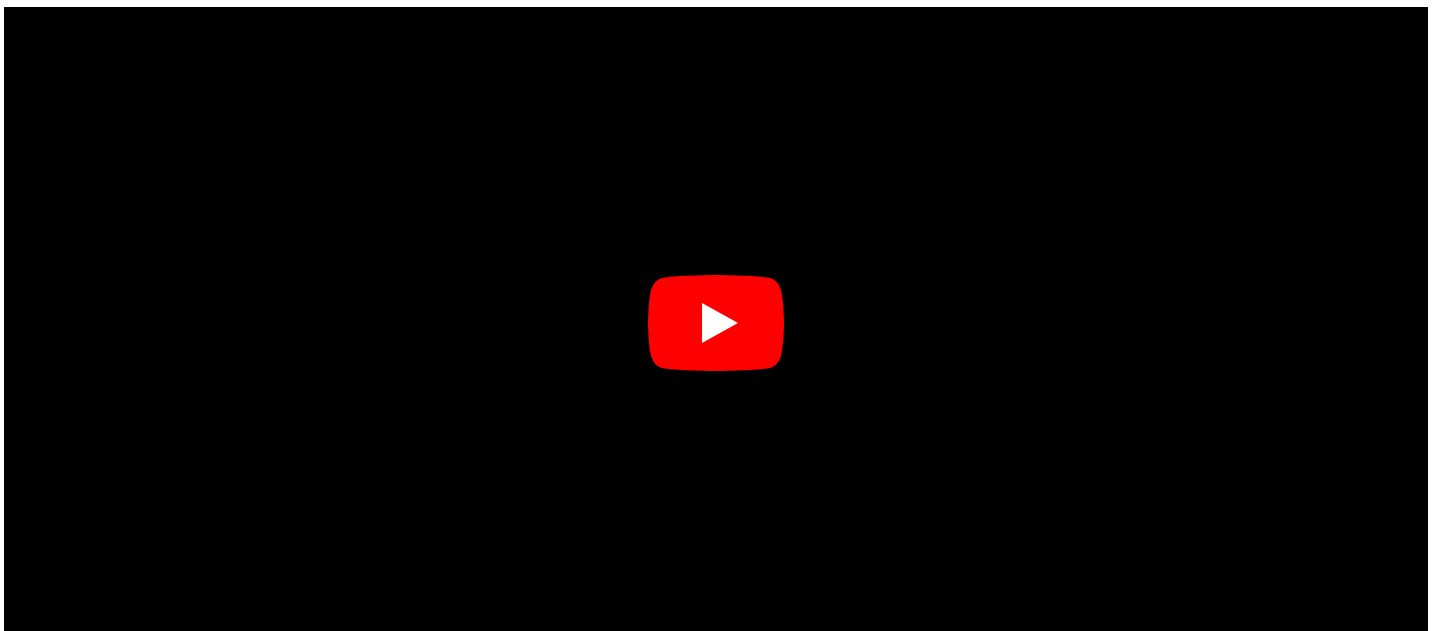
Target Audience

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

You Will Learn

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 is applied in the oilfield

Product Details

Categories: Upstream

Disciplines: Data Management, Science and Analytics

Levels: Basic

Product Type: Individual Skill Module

Format: On-Demand

Duration: 2 hours (approx.)

$250.00