N595: Introductory Machine Learning with Applications for Petroleum Engineers and Geoscientists

Instructor(s): Deepak Devegowda

Format and Duration
Classroom - 3 Days
Virtual - 5 Sessions

Summary

Business impact: This course enables participants to develop an understanding of data-driven workflows from data processing and data QA/QC to the application of various algorithms to gain insights from data that are not readily available with physics-based models. The course will provide participants with: (1) An overview of machine learning terminology and workflows (2) A discussion of use-cases covering a broad spectrum of disciplines in the oil and gas industry (3) Practical hands-on workflows in Python for several use-cases using a diverse set of machine learning algorithms.

Learning Outcomes

Participants will learn to:

1. Discuss the subsurface applications of machine learning.
2. Perform data clean-up, outlier detection and handling, and visualization.
3. Apply a structured approach to unsupervised clustering and cluster evaluation.
4. Perform supervised methods for classification and regression.
5. Critically evaluate the various machine learning algorithms.

Training Method

A classroom or virtual classroom course. Each session begins with a lecture to provide the introduction, mathematical foundations, and theory, followed by examples and Python-based practical exercises.

Software and Datasets: Jupyter notebooks written in Python will be provided to participants with real field datasets. Participants will execute specific tasks in Python to aid in their learning experience.

Who Should Attend

The course is designed for geoscientists and petroleum engineers with little to no prior experience in machine learning, seeking to incorporate machine learning tools in their E&P workflows. It is expected that participants have a working knowledge of programming and Python. A Python self-assessment rubric is provided in advance of the course; If participants need a Python refresher, a series of pre-course self-paced videos and exercises will prepare participants with the necessary background for the machine learning course.

Course Content

Introduction

- Terminology related to data-driven modelling

N595: Introductory Machine Learning with Applications for Petroleum Engineers and Geoscientists

Instructor(s): Deepak Devegowda

Format and Duration
Classroom - 3 Days
Virtual - 5 Sessions

- Key concepts related to supervised, unsupervised, and reinforcement learning
- A discussion of use cases relevant to the engineering and geoscience disciplines

Data Processing

- Basic probability, univariate, and multivariate statistics
- Exploratory data analysis, visualization and plotting multivariate data
- Data imputation for numerical and categorical data
- Outlier handling

Machine Learning Workflows

- Transformations: Scaling, dimensionality reduction
- Unsupervised learning: K-means and K-Prototype clustering, variants of Hierarchical Clustering, DBSCAN, Gaussian Mixture Models
- Supervised learning:
    - Model Evaluation, Cross Validation, and Parameter Selection
    - Classification: Decision Trees, Random Forests, Gradient Boosting, K-Nearest Neighbours, Support Vector Machines, Neural Networks
    - Regression: K-Nearest Neighbours, Neural Networks, Decision Trees and Random Forests, Multilinear Regression
    - Feature selection, ranking, and elimination
- Use cases in the E&P industry