MACHINE LEARNING & DATA ANALYTICS FOR SUBSURFACE IMPLEMENTATION RSE043

COURSE OVERVIEW

In this course, we will teach participants various machine learning techniques and tools necessary to train, test, and apply a model from scratch. We will use open source and easy to learn language such as Python to build workflows using different libraries (pre-built code samples) to perform various tasks in an interactive session. In other words, we make the entire coding experience and set up easy, pleasant, and straightforward.

There is no need to have any experience in machine learning, data science, linear algebra, or coding. We will teach every step of the way, anyone in the organization who has the slightest passion for implementing AI, ML and deciphering essential information from the data is welcome to attend. We will also accommodate those who have a higher level of experience with implementing AI and ML using python.

COURSE OBJECTIVES

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

- Understand machine learning fundamentals and its applications in the oil and gas industry.
- Use Python for implementing machine learning workflows and data analysis.
- Manipulate and visualize data using Pandas, Matplotlib, Seaborn, Plotly, and Cufflinks.
- Preprocess data for machine learning, handling outliers, missing data, and scaling.
- Build and evaluate supervised machine learning models (linear regression, logistic regression, K-nearest neighbors, decision trees, random forests, support vector machines, neural networks).
- Apply dimensionality reduction and clustering algorithms for unsupervised learning.
- Select and fine-tune machine learning models using cross-validation and grid search.
- Solve oil and gas industry problems with machine learning (e.g., frac pump failure detection, rock properties prediction, pipeline leak detection).
- Analyze and interpret machine learning model results for subsurface decision-making.
- Work on customized exercises with real-world oil and gas datasets.

WHO SHOULD ATTEND

- Reservoir Engineers
- Petroleum Engineers
- Geoscientists
- Field Engineers

COURSE DURATION

5 Working Days

COURSE OUTLINES

1. Introduction
    - Machine Learning and Python
        - What is Machine Learning?
        - Practical applications of machine learning in various industries
        - The future of ML and artificial intelligence (AI)
        - Why Python?
2. Getting started with Python
    - Python installation (Anaconda installation)
    - Jupyter Notebook interface and functionalities
    - NumPy
        - NumPy arrays, indexing, and selection
        - NumPy operations
        - NumPy example problems and solutions
    - Pandas
        - Pandas series
        - Data frames
        - Groupby
        - Merging, Concatenation, and joining
        - Data input and output
        - Pandas example problems and solutions
    - Data visualization with Matplotlib and Seaborn
        - Distribution plots
        - Categorical plots
        - Grids
        - Matrix plots
        - Regression plots
        - Seaborn example problems and solution
    - Plotly and Cufflinks data visualizations
3. Data preprocessing
    - Supervised vs. unsupervised learning

- Data preprocessing for machine learning
  - Outliers
  - Missing data
  - Scaling
  - Data Transformation
  - Feature selection
  - Date Preprocessing problems and solutions

4. **Model Building**
   - Data Preparation
   - Training
   - Testing
   - Validation
   - Results Analysis

5. **Unsupervised machine learning: Dimensionality Reduction**
   - Principal component analysis (PCA)
   - PCA example problems and solutions
   - Clustering
     - K-means Clustering
     - K-means clustering example problem and solution
     - Density-based spatial clustering of applications with noise (DBSCAN)
     - Hierarchical Clustering

6. **Supervised Learning**
   - Introduction to predictive model characteristics
     - Generalization
     - Overfitting and underfitting
     - Correctness
     - Classification or Regressions
   - Linear Regressions
     - Linear regression using SciKit-learn package
     - Linear regression example problem and solution
   - Logistic Regression
     - Logistic regression and classification using SciKit-learn package
     - Logistic regression classification example problem & solution

- K-Nearest Neighbor (KNN)
  - K-nearest neighbor using SciKit-learn package
  - K-nearest neighbor example problem and solution
- Decision Trees (DT) and Random Forest (RF)
  - Decision trees and random forest using SciKit-learn package
  - Decision trees and random forest example problem and solution
- Support Vector Machine (SVM)
  - Support vector machine using SciKit-learn package
  - Support vector machine example problem and solution
- Neural Networks
  - Introduction and basic architecture of neural network
  - Feed forward
  - Feed Backward
  - Loss Function
  - Backpropagation
  - Hidden Layer
  - Activation Function
  - ANN example problem and solution

7. **Model Evaluation**
   - Model selection
   - Cross Validation
   - Grid Search
8. **Oil & Gas Applications & Case Studies**
   - Frac pump failure detection (predictive pump maintenance)
   - Troublesome frac completions stage identification
   - Rock properties prediction
   - Bit failure prediction
   - Synthetic data creation
   - Pipeline leak detection
   - Liquid loading detection and automation
   - Screen-out prediction
   - Friction reducer (FR) optimization
9. **Customized exercise based on client's data set**