**WELL LOG FACIES ANALYSIS & CLASSIFICATION USING MACHINE LEARNING TECHNIQUES**



Name :- L Akhilesh

Roll No :- 200107045

Submission Date :- 30/03/2024

Course :- CL 653; AI-ML In Chemical Engineering

# **I) Project Overview: -**

INTRODUCTION:

The project aims to develop a machine learning model to classify geological facies in the Hugoton and Panoma Fields, North America's largest gas fields. Accurate classification of these facies is crucial for reservoir characterization, as they impact reservoir quality and production performance. The dataset comprises wireline log measurements and geologic constraining variables, matched with core samples collected at half-foot intervals in well locations.

OBJECTIVES:

* Develop a machine learning model to accurately classify geological facies in the Hugoton and Panoma Fields using wireline log measurements and geologic constraining variables.
* Enhance reservoir characterization and support decision-making in reservoir development by providing precise facies classification.

# **II) Description Of the Project: -**

GEOLOGICAL FACIES CLASSIFICATION:

In geology and petroleum engineering, **geological facies** refer to distinct rock units with specific characteristics, such as sedimentary features, lithology, depositional environment, and other geological properties. Facies classification is a fundamental task in reservoir characterization, as different facies exhibit varying porosity, permeability, and hydrocarbon saturation properties, which directly influence reservoir quality, connectivity, and production performance.

Wireline logging is a well-logging method that uses a variety of downhole tools to measure the physical properties of rocks and fluids in the subsurface.

The dataset contains well-logs from the largest gas fields in North America, where core samples have been collected every half foot and matched with the logging data at the well location.

By addressing these aspects, this project aims to develop an accurate and reliable machine learning model for classifying geological facies in the Hugoton and Panoma Fields, thereby improving reservoir characterization, production optimization, and decision-making in petroleum engineering applications.

# **III) Block Diagram/Flowchart of Process & Model Implementation: -**

## **Modelling Steps:**

#### **Data Collection and Preprocessing:**

* **Data Source: Hugoton and Panoma Fields Well-Logs and Core Samples**
  + **Well-Logs Data**
  + **Core Samples Data**
* **Data Preprocessing**
  + **Data Cleaning**
  + **Data Transformation**
  + **Feature Engineering**

#### **Machine Learning Model Development:**

* **Feature Selection**
  + **Correlation Analysis**
  + **Feature Importance**
* **Model Training**
  + **Model Selection**
  + **Hyperparameter Tuning**
  + **Cross-Validation**
* **Model Evaluation**
  + **Classification Metrics**
  + **Confusion Matrix**

#### **Model Interpretation and Visualization:**

* **Facies Classification Results**
  + **Facies Prediction**: Use the trained machine learning model to predict the geological facies labels for the test dataset.

## **Well Logging and Sampling Process:-**

* Well Drilling: Drilling of wells into the reservoir to extract rock cores and perform wireline logging.
* Core Sampling: Collection of core samples from the wells at regular intervals (every half foot) for laboratory analysis and facies classification.
* Wireline Logging: Measurement of physical properties of the rocks and fluids using downhole tools to obtain wireline log data.

# **IV) Data Source : -**

<https://github.com/akhiroxxx/AI-ML-Project-CL>

## **Data Characteristics :-**

* **GR (Gamma Ray)**
* **ILD\_log10 (Resistivity Log)**
* **PE (Photoelectric Effect Log)**
* **DeltaPHI (Porosity Index)**
* **PNHIND (Neutron-Density Porosity)**
* **NM\_M (Nonmarine-Marine Indicator)**
* **RELPOS (Relative Position)**

# **V) Description of Data : -**

## **Nature of Data :-**

The dataset provided for this project is primarily **steady-state** data. In the context of reservoir engineering and well-logging, steady-state data refers to measurements taken under conditions where the properties of the reservoir and the flow of fluids (oil, gas, water) are relatively constant over time.

## **Preprocessing of Data :-**

* Setting up columns into categorical data to improve model performance and interpretation.
* Standard preprocessing can be used; where data is standardized to zero mean and unit variance. Scikit preprocessing can be made use of.
* Also data splitting for test set.

# **VI) Strategies for AI/ML Model Development : -**

For this classification problem, several ML algorithms can be considered. For this multi-class classification problem models like the ones below can be used.

1. Random Forests Classification:

* Can capture complex non-linear relationships and interactions between the input features and the target variable.
* Can provide improved accuracy and robustness by reducing overfitting and variance.

1. SVM Classification:

* SVM is effective in high-dimensional spaces and can efficiently perform multi-class classification.
* It works well for both linearly separable and non-linearly separable data by using different kernel functions.

1. Gradient Boosting Machines:

* An ensemble learning technique that builds multiple weak learners (typically decision trees) sequentially and combines them to improve the model's accuracy and generalization capability.

1. Logistic Regression:

* It is efficient, easy to implement, and provides probabilities for outcomes.
* It assumes a linear relationship between the features and the log-odds of the target.

1. K Nearest Neighbors:

* is a non-parametric, instance-based learning algorithm that stores all available cases and classifies new cases based on a similarity measure

1. Naïve Bayes:

* is a probabilistic classifier based on Bayes' theorem with an assumption of independence between features.

## **Evaluation Strategy :-**

1. Accuracy:
   * Appropriateness: Accuracy provides an overall measure of the model's correctness in predicting the geological facies and is suitable for balanced datasets.
2. Precision, Recall, and F1-Score:
   * Appropriateness:
     + Precision: To evaluate the model's ability to correctly identify the facies and minimize false positives.
     + Recall (Sensitivity): To evaluate the model's ability to correctly identify the facies and minimize false negatives.
     + F1-Score: To assess the balance between precision and recall and provide a harmonic mean of the two metrics, which is particularly useful for imbalanced datasets.
3. Confusion Matrix

# **VII) Deployment Strategy: -**

The deployment strategy involves integrating the trained machine learning model with existing systems through API development and database integration, developing an interactive dashboard to visualize and analyse the model predictions and results, providing user training and support, implementing monitoring and logging mechanisms for model maintenance, regular retraining and updates of the model to adapt to the changing reservoir conditions and data patterns, and deploying the model, API, and dashboard on cloud platforms for scalability, availability, and accessibility.

By following this deployment strategy, the machine learning model can be effectively deployed in a real-world environment, integrated with existing systems, and maintained and updated over time to support reservoir characterization, exploration, and decision-making in petroleum engineering applications.

# **VIII) Scalability & Performance Optimization: -**

The scalability strategy involves designing a scalable model architecture, utilizing distributed computing and parallel processing techniques, provisioning and optimizing the computational resources, implementing efficient data pipelines and storage solutions, and leveraging cloud computing and on-premises infrastructure to handle increased demands and larger datasets efficiently.

The performance optimization strategy includes algorithmic optimizations, feature engineering and selection, hyperparameter tuning, leveraging GPU acceleration and high-performance computing, utilizing optimized software solutions and libraries, and implementing quantization and pruning techniques to enhance the model's efficiency, speed, accuracy, and performance in classifying geological facies and supporting reservoir characterization and decision-making in petroleum engineering applications.

# **IX) Use of Open-Source Tools: -**

The utilization of open-source tools and libraries, including Scikit-learn, TensorFlow, Keras, PyTorch, Pandas, NumPy, Matplotlib, and Seaborn, will significantly contribute to the development, training, evaluation, deployment, and optimization of the machine learning model for classifying geological facies.

These tools and libraries offer comprehensive functionalities, algorithms, and utilities for data preprocessing, feature engineering, model development, training, evaluation, optimization, visualization, and deployment, supporting the efficient and effective implementation and integration of the machine learning model with existing systems, applications, and workflows in petroleum engineering applications.

# **X) Purpose & Use Case: -**

The development and deployment of the machine learning model for classifying geological facies in petroleum engineering applications will significantly contribute to enhancing the accuracy, efficiency, and reliability of reservoir characterization and exploration, improving decision-making, risk assessment, and optimization of exploration and production operations, fostering innovation, collaboration, and digital transformation in the petroleum engineering industry, and promoting environmental protection, conservation, and sustainability.

The project's value and relevance to real-world challenges are justified by its potential impacts on improving reservoir performance, recovery efficiency, and economic viability, reducing exploration and production costs, increasing hydrocarbon recovery, maximizing profitability and sustainability, fostering innovation, collaboration, and digital transformation, and promoting environmental protection, conservation, and stewardship in the petroleum engineering domain.

# **XI) Conclusion: -**

The development and deployment of the machine learning model for classifying geological facies in petroleum engineering applications will significantly contribute to enhancing the accuracy, efficiency, and reliability of reservoir characterization and exploration, improving decision-making, risk assessment, and optimization of exploration and production operations, fostering innovation, collaboration, and digital transformation in the petroleum engineering industry, and promoting environmental protection, conservation, and sustainability.

The project's value and relevance to real-world challenges are justified by its potential impacts on improving reservoir performance, recovery efficiency, and economic viability, reducing exploration and production costs.

# **XII) References :-**

1. Dubois, D., Grader, A., Barry, P.H., et al. (2007). Hugoton and Panoma Fields, Kansas. Kansas Geological Survey, Open-file Report 2007-29.
2. University of Kansas. (n.d.). Class Exercise on Hugoton and Panoma Fields.
3. Scikit-learn Documentation. (n.d.). Machine Learning in Python.
4. TensorFlow Documentation. (n.d.). An end-to-end open-source machine learning platform.
5. Keras Documentation. (n.d.). The Python Deep Learning Library.
6. PyTorch Documentation. (n.d.). An open source machine learning framework.
7. Pandas Documentation. (n.d.). Data Structures and Data Analysis Tools.
8. NumPy Documentation. (n.d.). The fundamental package for scientific computing with Python.
9. Matplotlib Documentation. (n.d.). Visualization with Python.

10.Seaborn Documentation. (n.d.). Statistical Data Visualization.