Ministry of Higher Education and Scientific Research

Amar Thelidji- Laghouat University

FACULTY: TECHNOLOGY

DEPARTMENT: PROCESS ENGINEERING

MASTER THESIS (L M D)

Presented by: Walid Boussebci

BRANCH : Hydrocarbons

OPTION : Gas Engineering

Topic

Application of machine learning algorithms for porosity prediction based on well logging data

Jury de soutenance :

Nom et Prénom	Grade	Qualité
MERIGUI Khaled	MAA	Président
		Examinateur
HADJADJ Asma	MCB	Examinateur
YOUCEFI Mohamed Riad	MCB	Rapporteur

Promotion : JUIN 2024

Acknowledgement

Upon completion of this research proposal, I want to offer the heartfelt appreciation to a lot of individuals. Without their aid. I might not have been successful in achieving this thesis.

It would be impossible to achieve it alone without the support of my family. I would like to extend my sincere appreciation to them. I want to write a certain letter to my parents and sister. They have been constant companions throughout my journey on the final year paper.

My supervisor, Dr. Mohamed Riad YOUCEFI is the one person I cannot thank enough. I am very lucky to have a lecturer who directed me to return when I strayed. He stood with me and messages of support how to manage many crises and complete my study paper.

Best friends helped me throughout these difficult times, especially Ferial Bedjeladjel, Fatima and Abdul karim, this study could not have been done without the support and help from them. Their words of support and participations have assisted me in finishing this particular study. I treasure our connection and am grateful for their faith towards me.

"Last but not least, I wanna thank me, I wanna thank me for believing in me, I wanna thank me for doing all this hard work, I wanna thank me for having no days off. I wanna thank me for... for never quitting, I wanna thank me for always being a giver and tryna give more than I receive, I wanna thank me for tryna do more right than wrong, I wanna thank me for just being me at all times".

Table of Contents

List of Figures

List of Tables

Nomenclature

AAPE	Average absolute percentage error.
ANN	Artificial neural network.
CCL	Casing collar locator.
DT	Sonic transit time.
ELM	Extreme machine learning.
FFNN	Feedforward neural network.
FL	Fuzzy logic.
GR	Gamma-ray.
GST	Gamma-ray spectroscopy tool.
HI	Hydrogen index.
LBFGS	Limited-memory BROYDEN-FLETCHER-GOLDFARB-SHANNO.
LLS and LLD	Shallow and Deep lateroresistivity logs.
LM	Levenberg-Marquardt.
LWD	Logging while drilling.
MICP	Mercury injection capillary pressure.
ML	Machine learning.
MLP	Multilayer perceptron.
MSFL	Microspherical focused resistivity log.
NPHI	Neutron porosity.
PE	Photoelectric.
PEF	Photoelectrical factor log.
PHIT	Total porosity.

R^2	Coefficient of determination.
RBF	Radial basis function.
RHOB	Bulk density.
RHOZ	Photoelectric log.
RLLD	Deep lateral log resistivity.
RMSE	Root mean square error.
RXOZ	Shallow resistivity.
SGD	Stochastic gradient descent.
SVM	Support vector machine.
SVR	Support vector regression.
TDT	Thermal decay tool.

Introduction

1. Background of the study

Petroleum reserves are heterogeneous geological systems with great internal complexity. There are many forms of heterogeneity in rock properties in petroleum reserves. Understanding the shape and spatial distribution of these variations is important in evaluating petroleum reservoirs [1,2]. Porosity is the main variables for reservoir characterization in order to estimate the volume of hydrocarbons and their flow patterns to improve field production. Porosity is the fraction of pore space volume to the total volume of a rock. It is often related to the measure of fluid storage capacity inside rock. Total or absolute porosity includes all the pore space, while effective porosity describes only the interconnected pore space [3,4].

Porosity is a key property of oil reservoirs that shows the ability of rocks to store fluids such as oil, water, and gas. Therefore, determining porosity is a critical task in estimating reserves, production, and development of oil reservoirs. The porosity can be measured well logging measures of physical reservoir properties [5]. The downhole logging tools for determining rock porosity are costly. In addition, the logging operation might be affected by the hole conditions due to the mud contamination [6,7]. Porosity can be measured using the lab measurement such as mercury displacement, gravimetric, and Boyle's measurement [8]. However, these techniques take much time and cost for coring the rock sample and lab testing and they might provide a complete log for the rock porosity [9]. Thus, it is very important to seek any robust approaches and procedures that allow us to reliably estimate porosity in wells.

The porosity measurement could be done by applying machine learning (ML) algorithms. The applications of machine learning (ML) techniques provided huge contributions for dealing with petroleum data in different disciplines. ML tools such as artificial neural networks (ANNs), fuzzy logic (FL), Extreme Machine Learning (ELM), Radial Basic Function (RBF), and support vector machines (SVMs) provided high performance and accurate prediction results [10,11]. The implementation of algorithms contributed to solve several technical issues such as estimation of key drilling parameters [12–14], predicting the drilling fluid properties [15,16], reservoir fluid properties [17–19], rock properties [20,21], and geo-mechanical properties [22].

2. Literature Review

Many papers have investigated the application of machine learning algorithms to predict the porosity. Among these models, Rafik & Kamel [23] have predicted the porosity using a feedforward neural network (FFNN) to predict permeability and porosity from well log data in the Hassi R'Mel Field, Algeria. They trained the FFNN on a dataset of 927 data points that were collected from seven wells. The authors have considered six inputs including GR, RLLD, DT, NPHI, RHOB, and SW to develop a robust model that can predict the porosity. They found that their models can predict the porosity with an average error of 3%.

The model built in that study was robust and exhibit high performance. This suggests that FFNNs can be a very accurate tool for predicting porosity from well log data. In addition, the fact that the authors used a relatively modest dataset to train and test their FFNN. This reveals that FFNNs may be able to learn from limited data, which is important for many real-world applications.

There is also other research by the authors Ahmadi & Chen [24] who compared the performance of machine learning methods for estimating permeability and porosity of oil reservoirs in the Northern Persian Gulf using petro-physical logs. They used a dataset of 857 data points for the training and 366 for testing the models from seven wells, with inputs of sonic transit time (DT), density tool reading (NPHI), bulk density (RHOB), PHIT (total porosity), and outputs the matched permeability and porosity in a core laboratory. They trained and tested a variety of machine learning models, including artificial neural networks (ANNs) and support vector machines (SVMs).

Their results showed that both ANNs and SVMs can be used to accurately estimate permeability and porosity, with average errors of less than 1%. However, ANNs were found to be slightly more accurate than SVMs, especially for estimating permeability.

This study demonstrated that machine learning can be a powerful tool for estimating permeability and porosity of oil reservoirs, even with limited data. The study used a relatively small dataset, so it would be interesting to see how the models perform on larger datasets.

This is a well-written and informative study that provides valuable insights into the use of comparison of machine learning methods for estimating permeability and porosity of oil reservoirs via petro-physical logs.

Another outstanding research was conducted by Helle et al. [25]. The authors applied ANN with simple three-layer network to predict porosity and permeability from wireline logs in an offshore eastern Canada field. They trained the ANNs on a dataset that consists of sonic, density, and resistivity as inputs parameters and porosity output parameter, and for modeling the permeability they considered the density, gamma ray, neutron porosity and sonic as inputs.

For porosity prediction, the ANN achieved a mean difference between predicted porosity and helium porosity from core plugs of less than 0.01 fractional units. For permeability prediction, the ANN achieved a mean difference of approximately 400 mD.

The authors concluded that ANNs are a promising tool for predicting porosity and permeability from wireline logs. This is a well-written and informative study that provides valuable insights into the use of ANNs for predicting porosity and permeability from wireline logs.

There was also another distinguished research conducted by Gholami et al. [26] who used a support vector machine (SVM) to predict permeability in the Iranian South Pars field. They trained the SVM using a dataset of 125 data wells, and testing data of 50 wells, with inputs of sonic log DT, gamma ray log GR, compensated neutron porosity log NPHI, density log ROHB, photoelectric factor log PEF, microspherical focused resistivity log MSFL, and shallow and deep lateroresistivity logs LLS and LLD. Outputs of core permeability datasets were obtained from 3 wells of Kangan and Dalan gas reservoirs. The SVM achieved a correlation coefficient of 0.97 between predicted and actual permeability.

Gamal and Elkatatny [9] developed a new model based on an ANN for predicting rock porosity from only drilling parameters that include weight on bit, torque, standpipe pressure, drill string rotation speed, rate of penetration, and pump rate. The authors employed two data sets for training the ANN model (3767 data points) and the second one for validating the developed ANN model (1676 data points). The developed model showed a correlation coefficient (R) between the predicted and actual porosity values of 0.97 and 0.92 with average absolute percentage errors (AAPE) of 6.2 and 9.3% for training and testing, respectively.

3. Motivation and Contribution (Thesis objectives)

In this thesis, we display an outline of the application of machine learning for porosity expectation within the oil and gas industry. Based on logging data, Multi-Layer Perceptron (MLP) and Support vector machine (SVM) were applied in this study to build a ML models for predicting the porosity.

The main objective of this project is to provide a comprehensive summary of relative performance and evaluate algorithms, comparing MLP and SVM in reservoir characterization for porosity prediction based on their applications. The relevant details of their relative performance will assist in selecting appropriate supervised machine learning models for future reservoir characterization studies. This review also aims to differentiate between the algorithms and clarify which one is more suitable based on the error rate, as well as explore the use of hybrid machine learning techniques when there are scattered data points. Furthermore, it emphasizes the need to enhance accuracy and reduce error by exploring hybrid machine learning techniques, as they offer better options for improving prediction accuracy and overcoming limitations of individual models in this field.

Overall, this thesis points to a comprehensive outline of the use of machine learning strategies for porosity expectation within the oil and gas industry, highlighting its potential to upgrade effectiveness, exactness, and knowledge in store assessment and administration workflows.

Chapter I. Porosity: Key Concepts and Determination Methods

Introduction

In this chapter, we embark on an exploration of porosity, a fundamental concept in reservoir engineering. We will delve into its definition, significance in reservoir characterization, various types, and methods for determination, encompassing both core analysis and well logging techniques. Through this comprehensive understanding, we aim to provide the reader a comprehensive understanding of the importance of modeling the porosity which is done in our thesis

1. Definition of Porosity

The porosity of a rock is a measure of its storage capacity (volume of voids capable of retaining fluids as depicted in Figure I-1. This property represents the ratio of void volume to total rock volume.

It is determined mathematically by the following generalized relationship:

$$\emptyset = \frac{\text{void volume}}{\text{total volume of rock}} \qquad \text{(I. 1)}$$

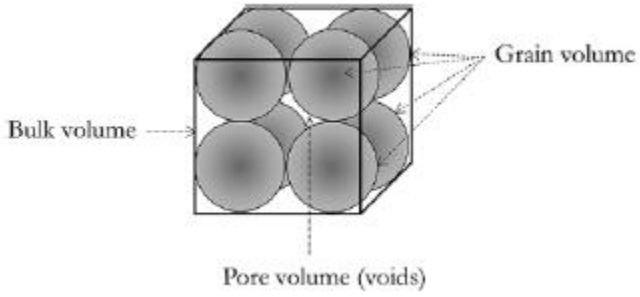

Figure I-1. Principle of porosity

2. Types of Porosity

During rock sedimentation, certain voids that have developed remain interconnected, while certain porous spaces are completely isolated by cementation.

This leads to two types of porosity:

- Absolute porosity

- Effective porosity

2. 1. Absolute porosity

Absolute porosity is defined as the ratio of total rock pore space to total rock volume. Absolute porosity is generally expressed mathematically by the following relationships:

$$\emptyset_a = \frac{\text{void volume}}{\text{total volume of rock}} \quad \text{(I. 2)}$$

Or by:

$$\emptyset_a = \frac{\text{total rock volume} - \text{grain volume}}{\text{total volume of rock}} \quad \text{(I. 3)}$$

2. 2. Effective porosity

Effective porosity is the ratio of interconnected pore space to total rock volume. Effective porosity is generally expressed mathematically by the following relationship:

$$\emptyset_e = \frac{\text{volume of interconnected pores}}{\text{total volume of rock}} \quad \text{(I. 4)}$$

Effective porosity is the value used in all reservoir calculations, as it represents the interconnected pore space containing recoverable hydrocarbon fluids.

3. Determination of Porosity using Core Analysis

This section delves into the world of direct methods used in laboratories to meticulously measure rock porosity. These methods provide highly accurate data on a rock's pore space, offering valuable insights for reservoir characterization. We will explore two common techniques:

3. 1. Mercury Injection Capillary Pressure (MICP) Method

The setup typically consists of a high-pressure chamber containing the rock sample, a volume pump for injecting mercury, pressure gauges to monitor pressure changes, and a data acquisition system to record the volume of mercury injected at each pressure increment [27].

3. 2. Boyle's Law

The Boyle Law, which is a fundamental gas law, finds application in the laboratory for measuring the porosity of rock samples. This method offers a straightforward approach to determine the pore volume within a rock by utilizing a known chamber volume and the ideal gas law. Here's a simplification of the process: Imagine a sealed chamber with a precisely known volume (Vc). This chamber acts as the container for our experiment. Additionally, we'll need a dry rock sample and a measured amount of gas, typically helium due to its inert properties [27].

4. Determination of Porosity using Well Logging

4. 1. Basic log types

4. 1. 1. Logging While Drilling (LWD)

Historically, petrophysicists focused primarily on wireline logging, which involves collecting data using tools running into the wellbore on a cable from a winch post-drilling. Yet, innovations in drilling and logging technology have enabled the gathering of log data through tools integrated into the drilling assembly itself. These instruments can either relay data to the surface in real-time or store it in downhole memory for retrieval upon resurfacing of the assembly [5].

While LWD tools introduce complexities and incur additional costs during drilling operations, their deployment becomes warranted under certain circumstances:

- Immediate access to real-time data is essential for operational decisions, like steering a well along a specific formation or identifying coring points and casing setting depths.
- Obtaining data before hole instability or invasion compromises measurements.
- Ensuring data preservation in scenarios where hole integrity is at risk.
- Navigating challenging trajectories, such as in horizontal wells, where wireline acquisition proves challenging.

LWD operations typically provide a range of data types, including:

- GR (Gamma Ray): Indicates natural gamma ray emissions from the formation.
- Density: Captures formation density via gamma ray Compton scattering utilizing a radioactive source and gamma ray detectors. This measurement may also incorporate a photoelectric effect (Pe) assessment.
- Neutron Porosity: Estimates formation porosity based on the hydrogen index (HI), utilizing gamma rays emitted during the capture of injected thermal or epithermal neutrons from a source within the tool string.
- Sonic: Measures the transit time of compressional sound waves through the formation.
- Resistivity: Provides formation resistivity data across multiple investigation depths, using an induction-type wave resistivity tool [5].

4. 1. 2. Wireline Open and Cased hole logging

This section dives into the world of wireline logging, a crucial practice employed after a drilling section is complete. With the drill bit retrieved, a valuable window opens for gathering detailed information about the formation before it's encased or abandoned. Two primary methods are used for this data acquisition [5]:

Some of the key tools used in wireline **open hole logging** include:

- Gamma ray: Measures natural radioactivity, aiding in differentiating rock types.
- Density: Provides information about formation density.
- Neutron porosity: Estimates porosity based on the formation's hydrogen content. Newer versions offer advantages like reduced dependence on salinity.
- Full-waveform sonic: Analyzes sound wave propagation to assess formation properties beyond just basic compressional velocity.
- Resistivity: Measures formation's ability to conduct electricity, helping determine fluid saturation. These come in laterolog (for water-based mud) and induction (for oil-based mud) varieties.

When a hole has been **cased** and a completion string run to produce the well, certain additional types of logging tools may be used for monitoring purposes. These include:

- Thermal decay tool (TDT): Estimates water saturation by measuring formation's chlorine content.

- Gamma ray spectroscopy tool (GST): Determines elemental composition, aiding in fluid saturation evaluation.
- Production logging: Measures flow contributions from different formation zones.
- Cement bond log: Evaluates the quality of the cement bond between casing and formation.
- Casing collar locator (CCL): Identifies casing collars and perforation intervals within the well [5].

4. 2. Calculating the porosity

4. 2. 1. Determination of porosity by gamma ray logging

Gamma-ray logging is a method for measuring formation density. It works by irradiating the formation with gamma rays and measuring the intensity of the returning radiation [27]. Denser materials attenuate the gamma rays more, so a weaker response indicates a higher density formation. This technique can be used to estimate porosity, as porosity is inversely proportional to density.

It is possible to estimate porosity by means of the formula [27]:

$$\phi = \frac{D_g - D}{D_g - D_\phi} \qquad \text{(I. 5)}$$

Where:

D : formation density according to the gamma log.

D_g: matrix density (grain density).

D_ϕ: density of fluid in pores.

4. 2. 2. Determination of porosity by Neutron Logging

Neutron Logging is a well logging technique for estimating porosity. It highlights that neutron logs measure the formations' response to high-energy neutrons. Formations with a higher concentration of hydrogen atoms will slow down and absorb more neutrons. Since water has more hydrogen than oil or gas, the neutron log primarily detects the amount of hydrogen present in the formation. As a first approximation we have [27]:

$$\log \phi_N = AN + B \quad \text{(I. 6)}$$

Where:

N: API neutron deflection.
A, B: constants.

4. 2. 3. Determination of porosity by sonic

The porosity affects the velocity of sound in rocks. It highlights that sound travels slower in rocks with higher porosity. This is particularly evident in reservoir rocks, where sound velocity can drop by 60% as porosity increases from 3% to 30% [27].

The estimation of porosity using acoustic logging. This equation is [27]:

$$\frac{1}{V_{\log}} = \frac{\phi}{V_f} + \frac{1-\phi}{V_m} \quad \text{(I. 7)}$$

where:

ϕ: porosity
$V_{\log}$: velocity measured in the rock by the logging tool
V_f: velocity in the fluid contained in the pores
V_m: velocity in the matrix

4. 2. 4. Determination of porosity by resistivity

A geophysical measurement is performed to assess the electrical resistance (resistivity) of the rock formation close to the borehole. This likely reflects the formation being filled with filtrate, a fluid introduced during the drilling process, rather than the original brine. Consequently, if the formation is water-bearing, this measurement provides valuable information about the properties of the formation water [27].

If the formation is water-bearing, it can be deduced that:

$$(FF) = \frac{R_{xo}}{R_{mf}} \quad \text{(I. 8)}$$

where: FF: Formation factor.

R_{xo}: shallow formation resistivity and R_{mf}: Resistivity of the Mud Filtrate.

Chapter II. A review of machine learning algorithms and their implementation in Python

Introduction

In this chapter, we delve into the fundamentals of Multilayer Perceptron (MLP) and Support Vector Machines (SVM), pivotal algorithms in machine learning. Understanding their principles is essential for implementing effective porosity models. Through Python implementation, we aim to showcase their practical applications and comparative performance in our thesis.

1. Support vector machines

Support vector machines (SVMs) are powerful supervised learning algorithms developed by Vapnik [28] in the 1990s for supervised classification.

1. 1. Supervised learning

Supervised learning is the branch of machine learning concerned with learning to make predictions based on the knowledge provided by N labeled observations, i.e. accompanied by the value to be predicted. In this case, we use a data set containing N observations $\{\vec{x}_i\}_{i=1,\dots,N}$ described in space E, accompanied by their responses $\{\vec{y}_i\}_{i=1,\dots,N}$ described in space F to construct a prediction function $f : E \rightarrow F$ such that, for any pair $(\vec{x}_i, \vec{y}_i) \in E \times F$, $f(\vec{x}_i) \approx \vec{y}_i$.

A supervised learning problem in which the space of variables forms discrete, finite sets whose labels y correspond to categories to be identified, is called a supervised classification problem. The function f associates each observation x with one of the classes. The boundaries between classes are called separation boundaries. If the target y takes only two values $y = \{0,1\}$ or $y = \{-1, +1\}$ then we speak of binary classification, otherwise if; $|y| > 2$ the classification is said to be multiclass. Figure II.1 illustrates a classification problem.

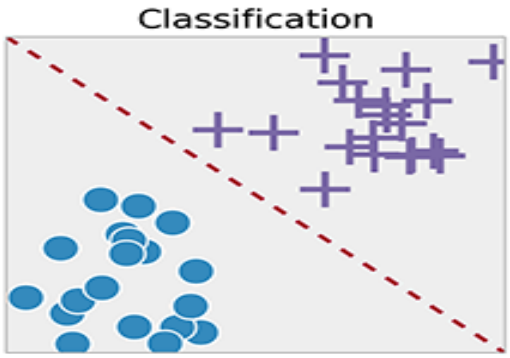

Figure II-1. Classification problem

1. 2. Prediction with SVM

In the case of discrimination of a dichotomous variable, support vector machines aim to construct a decision function associating each observation with its class, and minimizing the probability of error for each observation. Assuming we have N training observations divided into two linearly separable classes, denoted (x_i, y_i) with $y_i \, \epsilon \, \{-1,1\}$, the Figure II. 1 and Figure II. 2 reveals that there are an infinite number of separating hyperplanes that make no error in classifying the training data and that verify for any i between 1 and N:

$$\begin{cases} w \cdot x_i + b > 0 & if \quad y_n = 1 \\ w \cdot x_i + b < 0 & if \quad y_n = -1 \end{cases} \qquad \text{(II. 3)}$$

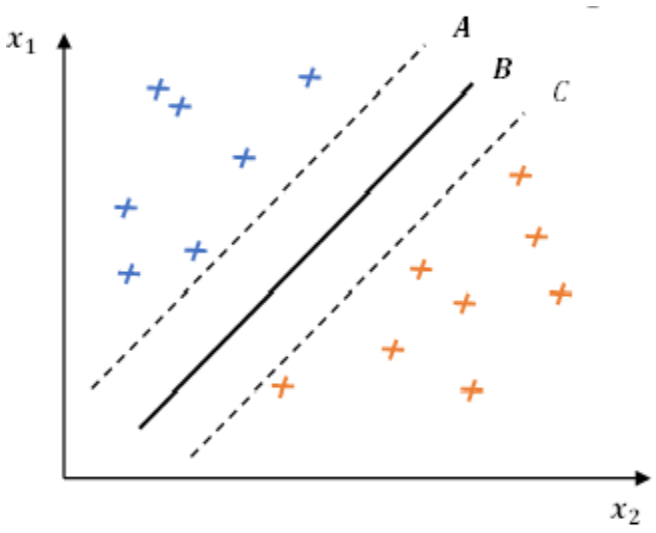

Figure II-2. Infinite hyperplanes separate the two classes: negative points and null points

It is obvious that the three hyperplanes described in Figure II. 2 offer exactly the same classification scores, but the separating hyperplane B seems the most preferable as it is equidistant from the nearest orange-colored observation and the nearest blue-colored observation. In addition to the optimal hyperplane B, we can define the separating hyperplane A and the hyperplane C. These two hyperplanes are parallel to hyperplane B and located at a distance m on either side. Knowing that the optimal hyperplane B we're looking for has equation $w \cdot x_i + b = 0$, the equation of hyperplane A is written as follows:

$$w \cdot x_i + b = M \quad \text{(II. 4)}$$

Since A is symmetrical to B with respect to C, its equation is :

$$w \cdot x_i + b = -M \quad \text{(II. 5)}$$

We can set M=1 so that all the positive points in our training set satisfy $w \cdot x_i + b \geq 1$. Similarly, all negative observations satisfy $w \cdot x_i + b \leq -1$.

We're therefore looking here to maximize the classifier margin m (the distance between the decision boundaries and the nearest training observation), which is equivalent to maximizing the distance between the blue-colored observation x_+^* and the orange-colored observation x_-^* located at a distance m from the separating hyperplane B. This notion of margin provides a good criterion for choosing classifiers, assuming that maximizing their margin is identical to maximizing their chances of generalization, and therefore minimizes the probability of classification error. The orthogonal projection of the distance vector $x_+^* - x_-^*$ onto the hyperplane B is given by:

$$(x_+^* - x_-^*)\frac{w}{\|w\|} = \frac{1}{\|w\|} \quad \text{(II. 6)}$$

The problem to be solved by support vector machines is formulated as follows [29]:

$$\begin{cases} \min\limits_{w,b} \frac{1}{2}\|w\|^2 \\ under\ constraint \quad y_i(w^T \cdot x_i + b) \geq 1 \end{cases} \tag{II. 7}$$

1. 3. SVM implementation in Python

Having understood the basics of SVM, this section explains how to implement it in Python. This is very simple and direct with the Scikit Learn SVM package [30], as described below:

1.3.1 Importing the data set

```
import pandas as pd
data = pd.read_csv(r'C:\file location\filename.csv')
```

1.3.2 Splitting and adjusting training and test data

```
from sklearn.model_selection import train_test_split
training_set, test_set = train_test_split(data, test_size = 0.2, random_state = 1)
X_train = training_set.iloc[:,0:2].values
Y_train = training_set.iloc[:,2].values
X_test = test_set.iloc[:,0:2].values
Y_test = test_set.iloc[:,2].values
```

1.3.3 SVR function initialization

```
from sklearn.ensemble import RandomForestRegressor
Regression_Model = SVR(C=, cache_size=,epsilon=, gamma=, kernel=, ….)
Regression_Model.fit(X_train, y_train)
```

1.3.4 Predicting targets for the test set

```
Y_pred = Regression_Model.predict(X_test)
```

2. Artificial neural network

Artificial neural network (ANNs) is one of the various branches of computational intelligence, which has been applied successfully in many fields such as in petroleum engineering [31–33], biology and biomedicine [34,35], modeling and design [36,37], finance [38]. Its architecture was inspired by adapting the biological neurons conception of the human

brain. The most well-known ANNs is Multi-Layer Perceptron (MLP). This network is an efficient algorithm for discovering existing patterns of data, approximating nonlinear function, and linking inputs and outputs variables of a complex system [39].

2. 1. Multi-layer perceptron (MLP)

MLP network is composed of three kinds of layers: input layer, hidden layer, and output layer, each of these layers contains a specific number of neurons which represent the processing elements. Hidden layers are answerable for establishing the existing relationship between the input and the output variables of the system [40,41]. The number of neurons in the first and the last layer is similar to the number of input and output variables, respectively, while the hidden layer number and their corresponding magnitude of neurons are defined empirically using trial and error [42]. Basically, applying MLP with only one hidden layer is enough for the modeling of relatively simple systems. However, if the system presents a higher complexity, more than one hidden layer is required for reaching an appropriate solution [43]. Each hidden layer neuron is connected to the whole neurons in the following and previous layers via an associated weight vector. The MLP output is generated by applying the transfer function between the input and the hidden layer and between the hidden layer and the output layer. The input value of the transfer function is calculated by summation of multiplied node's value of preceding layer to their specific weight factors added to a bias. Various transfer functions can be employed, some of which are shown below:

Tansig transfer function:

$$f(x) = \frac{2}{1 + e^{-kx}} - 1 \quad \text{(II. 8)}$$

Logsig transfer function:

$$f(x) = \frac{1}{1 + e^{kx}} \quad \text{(II. 9)}$$

Purelin transfer function:

$$f(x) = x \quad \text{(II. 10)}$$

Generally, Tanh and Sigmoid functions are applied for the hidden layer, while linear function is employed for the output layer.

In order to obtain an appropriate relationship between the inputs and the outputs of a system, back-propagation learning algorithms are used to efficiently train MLP networks by adjusting the connection weights and biases. Levenberg-Marquardt (LM) algorithm is widely employed for MLP modeling. More information about this algorithm can be found in the prior published works [44,45].

2. 2. MLP implementation in Python

Having understood the basics of MLP, this section explains how to implement it in Python. This is very simple and direct with the Scikit Learn MLP package [30], as described below:

2.2.1 Importing the data set

```
import pandas as pd
data = pd.read_csv(r'C:\file location\filename.csv')
```

2.2.2 Splitting and adjusting training and test data

```
from sklearn.model_selection import train_test_split
training_set, test_set = train_test_split(data, test_size = 0.2, random_state = 1)
X_train = training_set.iloc[:,0:2].values
Y_train = training_set.iloc[:,2].values
X_test = test_set.iloc[:,0:2].values
Y_test = test_set.iloc[:,2].values
```

2.2.3 MLP function initialization

```
from sklearn.neural_network import MLPRegressor
Regression_Model = MLPRegressor(hidden_layer_sizes=(), activation=, solver=, learning_rate_init=,…)
Regression_Model.fit(X_train,y_train)
```

2.2.4 Predicting the targets for the test set

```
predicted_test=Regression_Model.predict(X_test)
```

Chapter III. Application of machine learning for predicting porosity

Introduction

Porosity prediction plays a crucial role in various industries, particularly in reservoir engineering. Porosity, the measure of void spaces in a rock or soil, directly impacts the fluid storage capacity and permeability of subsurface formations, making it a fundamental parameter in resource exploration and production. Traditionally, porosity estimation relied heavily on physical measurements and empirical relationships, which often lacked precision and generalizability across different geological settings.

However, the advent of machine learning techniques has revolutionized porosity prediction by leveraging the power of data-driven models to capture complex relationships between input variables and porosity. In this presentation, we explore the application of machine learning, specifically Multilayer Perceptron (MLP) and Support Vector Regression (SVR) models, for predicting porosity based on a set of input features commonly used in the field of geosciences.

The input features utilized in two models include gamma ray (GR), photoelectric (PE), neutron porosity (NPHI), shallow resistivity (RXOZ), and bulk density (RHOZ). These measurements, acquired through various logging tools and techniques, provide valuable information about the subsurface lithology and fluid content, serving as proxies for porosity estimation.

Our primary objective is to evaluate the performance of MLP and SVR models in predicting porosity and to determine the relative importance of input features in the predictive process. Through comprehensive experimentation and statistical analysis, we aim to demonstrate the efficacy of machine learning approaches in porosity prediction and elucidate the strengths and weaknesses of MLP and SVR models in this context.

1. Methodology

1. 1. Parameter selection

Model performance is highly influenced by the input features used for training and validation, and it is essential to find out which features are the most relevant to predict porosity. In this study, the model's input features typically include various geological measurements such as gamma ray (GR), photoelectric (PE), neutron porosity (NPHI), Shallow Resistivity (RXOZ) and bulk density (RHOZ), which serve as proxies for understanding the rock's composition and structure. The choice of this inputs parameters to model the porosity was based on the previous studies [23]. The MLP and SVR learn from these features to discern underlying patterns and relationships that influence porosity.

The relationship between the five key's input parameters and porosity is pivotal in geoscience and petroleum engineering. These parameters collectively offer valuable insights into subsurface rock properties, including rock physical properties. By analysing their variations and interdependencies, can better understand reservoir characteristics, aiding in efficient hydrocarbon exploration and production strategies.

To investigate the strong relationship between the selected input parameters and the porosity, the Pearson correlation coefficient R was employed in this study to measure the relative importance between the input variables (GR, PEFZ, NPHI and RHOZ) and the porosity. This coefficient measures the linear relationship strength between one variable and its value vary between 1 and -1. The R coefficient value equal or close to 0 indicates that there is a weak correlation between these variables. In contrast, relative value to 1 or -1 implies a robust linear relationship [15]. Figure III-1 shows that the four variables are strong functions of porosity with R values equal to (GR =0.1493, PEFZ= -0.4589, NPHI=0.8684 and RHOZ= -0.8213) respectively. The positive values of correlation coefficient (GR and NPHI) reveal an increasing porosity.

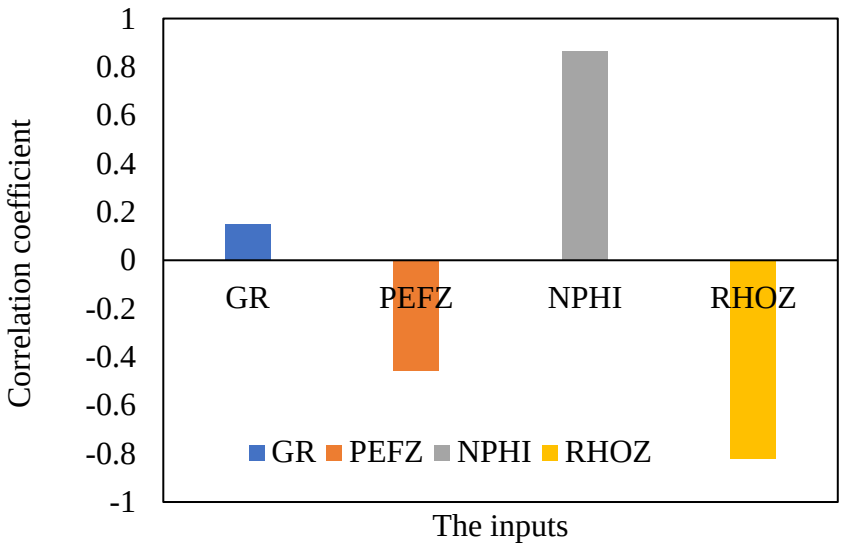

Figure III-1. **Relative importance between input and output parameters**

1. 2. Data collection, processing, and normalization

The well-logs and reservoir petrophysical properties datasets used to develop the Machine Learning (ML) models for predicting reservoir porosity ∅ were collected from well-A located in Libya field. The well-logs data are employed as input variables while the core porosity is employed as the target (output).

To check the validity of the collected data and remove the suspected outliers, data processing was carried out to remove data points that contain unrealistic values such as null and negative values. Such random values are mainly caused by logging sensor malfunctions that can occur during well logging operation.

In addition, data sets which consistently resulted in poor predictions were considered to be invalid and therefore were removed. After such a filtering process, a total of 292 data sets were used to develop the ML models: 80% of the filtered data are used to build (training) the models, and the other 20% are used to check their accuracy (predictive test).

2. Model development

2. 1. MLP model development

The process of developing a MLP model for porosity estimation involves a systematic exploration of the neural network's architecture to optimize its performance. Leveraging the scikit-learn library in Python [30], the development begins by defining a loop that iterates through a large number of trials, typically ranging from 1 to 1000. At each iteration, random combinations of key parameters are selected, including the number of neurons in the hidden layer, activation function type, and algorithm solver. The range of neurons tested spans from 5 to 30, covering a broad spectrum of network complexities. Activation functions explored include the logistic function, $(x) = \frac{1}{1+e^{-x}}$, and the hyperbolic tangent (tanh) function, $f(x) = \frac{e^{x}-e^{-x}}{e^{x}+e^{-x}}$, each bringing distinct nonlinearity properties to the model. Additionally, different solver algorithms such as Adam, Limited-memory Broyden–Fletcher–Goldfarb–Shanno (LBFGS), and Stochastic Gradient Descent (SGD) are investigated to optimize the model's convergence. Following parameter selection, the MLP model is trained on a subset of the available data, commonly 80%, with the remainder reserved for evaluation. Once trained, the model's performance is assessed on the test set using an evaluation metric such as absolute percentage error. After exhaustive iterations, the optimal combination is identified at iteration 594, where the MLP architecture consists of 15 neurons in the hidden layer, employs the hyperbolic tangent (tanh) activation function, and utilizes the LBFGS solver algorithm. This configuration yields a testing error of 3.371, showcasing its efficacy in accurately estimating porosity. The chosen model configuration, along with the predicted porosity values on both training and testing datasets, is then saved for further analysis and validation.

2. 2. SVR model development

The process of developing a Support Vector Regression (SVR) model for porosity estimation involves a rigorous exploration of various hyperparameters to enhance predictive accuracy. Utilizing the scikit-learn library in Python [30], the development initiates with a systematic search through a considerable range of parameter combinations. A loop is designed to iterate over numerous trials, typically ranging from 1 to 50. At each iteration, different

combinations of key parameters are randomly selected, including the choice of kernel type, regularization parameter (C), and gamma (γ).

For the kernel type, options such as linear, polynomial, radial basis function (RBF), and sigmoid are evaluated to identify the most suitable nonlinear mapping strategy. The regularization parameter (C) is varies between 0.1 and 100, influencing the trade-off between minimizing the training error and maximizing the margin of the decision function. Similarly, the gamma parameter, controlling the influence of individual training samples, is explored within the range of 0.1 to 1.

After exhaustive trials, the optimal combination of hyperparameters is determined. In this case, the SVR model attains its peak performance at iteration 29, where the selected kernel type is RBF, with a regularization parameter (C) of 99.40 and a gamma value of 0.9604174. This configuration demonstrates the model's ability to effectively capture the nonlinear relationships within the data.

Subsequently, the SVR model is trained on a subset of the available data, typically 80%, with the remaining portion reserved for evaluation. The performance of the trained model is then assessed on the test set using a suitable evaluation metric, such as mean absolute error or root mean squared error. The chosen SVR configuration, along with the predicted porosity values on both training and testing datasets, is saved for further analysis and validation.

3. Model assessment

3. 1. Statistical and graphical error analysis

Throughout this thesis, the root mean square error (RMSE), Average Absolute Percentage Error (AAPE), and the coefficient of determination (R^2) were employed as statistical indexes to evaluate the prediction accuracy and the effectiveness of the constructed models. In brief, the model prediction is satisfactory if RMSE and AAPE are close to 0, respectively, while R^2 is found close to 1.

The expressions used to calculate RMSE, AAPE, and R^2 are defined below [46]:

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(\emptyset_{\text{real},i} - \emptyset_{\text{predicted},i}\right)^2} \quad (1)$$

$$AAPE = \frac{1}{N}\sum_{i=1}^{N}\left|\frac{\emptyset_{\text{real},i} - \emptyset_{\text{predicted},i}}{\emptyset_{\text{real},i}}\right| \times 100 \quad (2)$$

$$R^2 = 1 - \frac{\sum_i^N \left(\emptyset_{\text{real},i} - \emptyset_{\text{predicted},i}\right)}{\sum_i^N \left(\emptyset_{\text{predicted},i} - \overline{\emptyset}\right)} \quad (3)$$

Where, N represents the number of data points, $\emptyset_{\text{real}}$ is the measured porosity value, $\emptyset_{\text{predicted}}$ is the porosity value estimated by the developed models and $\overline{\emptyset}$ is the average of the measured porosity values.

Additionally, graphical evaluations were created to depict the behavior of the proposed porosity model. A scatter plot in which all the calculated porosity values by the constructed models were sketched against the real porosity values. A 45° straight line is drawn on the cross plot between the experimental and estimated data points. The closer the sketched data points to this line, the higher is the reliability of the model. Moreover, the histogram plot of residuals which represent the differences between the measured and predicted porosity values was also employed in this study to assess the performance of the established porosity models.

3. 2. Results discussion

The obtained statistical index values for the developed porosity models are presented in From this table, it can be seen that both MLP and SVR exhibited a great performance with RMSE, AAPE, and R^2 values of 1.37297, 4.69307, and 0.89525, respectively, for the MLP model and RMSE, AAPE, and R^2 values of 1.520521, 4.508296, and 0.883217, respectively, for the SVR model. These obtained values statistical index reveal the high performance of the built models and that the SVR slightly outperforms MLP.

Figure III-2 depicts the measured porosity values against the obtained predicted porosity values for both MLP and SVR models. As it can be observed from this figure, the most of calculated data lay around the unit slope line for both SVR and MLP models. This tight distribution of the data around the unit line proves the perfect accuracy and robustness of the established ML models. Thus, these models yield better stability over the real porosity alone.

Table III-1. Results of MLP and SVR models

Model	Index	Training	Testing	All
MLP	RMSE	1.55849	0.63089	1.37297
	AAPE	5.02349	3.37138	4.69307
	R^2	0.8837	0.94584	0.89525
SVR	RMSE	1,570001	1,322601	1,520521
	AAPE	4,624138	4,044925	4,508296
	R^2	0,87536	0,911683	0,883217

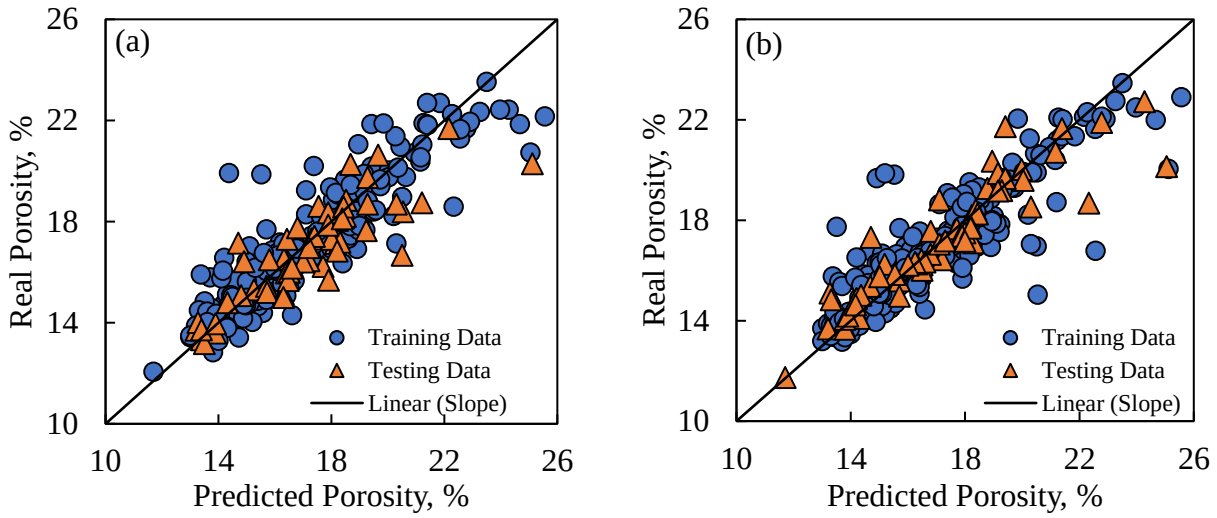

Figure III-2. Cross plots of real porosity versus predicted of porosity: subplot(a). MLP model, subplot(b). SVM model

The histogram for the MLP model (a) shows a lower relative error distribution than the SVR model (b) in **Figure III-3.** Histograms of relative error distribution between (a). MLP model and (b). SVR model This means that the MLP model generally makes predictions that are closer to the actual values than the SVR model.

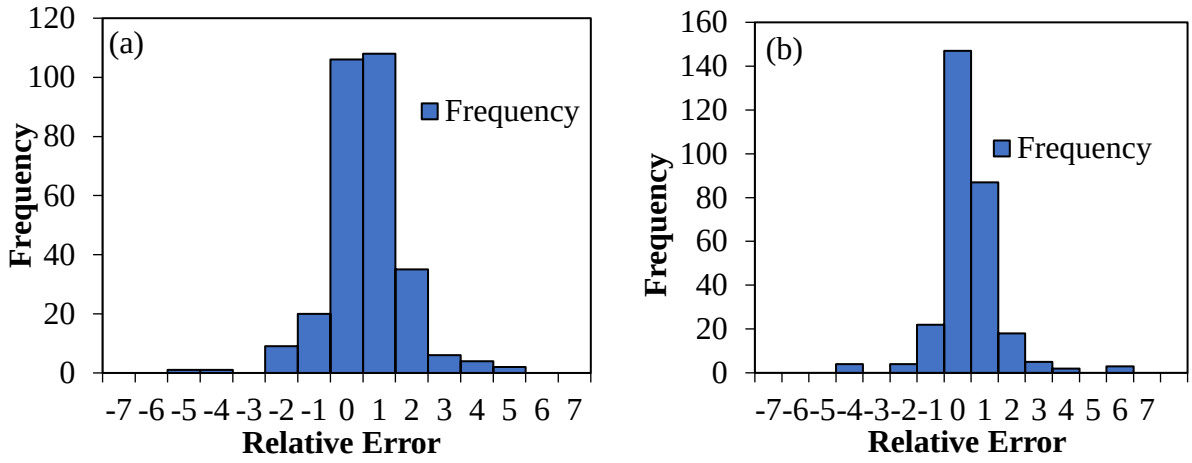

Figure III-3. Histograms of relative error distribution between (a). MLP model and (b). SVR model

3. 3. Trend analysis

As discussed previously. porosity is a strong function of the employed inputs (NPHI, RHOZ and PEFZ). In the current study, trend analysis is performed to testify the developed MLP model ability to follow this experimental data trend with the variation of input variables. The SVR model outputs are depicted for different range of NPHI, RHOZ, and PEFZ in **Figure III-4. Trend** prediction capability of the RHOZ (a), PEFZ (b) and NPHI (c)

According to this figure, it is apparent that the experimental and the predicted values are almost identical for different input parameters. This finding confirms that the proposed MLP model can accurately predict the porosity under various conditions and perfectly capture the expected physical trends.

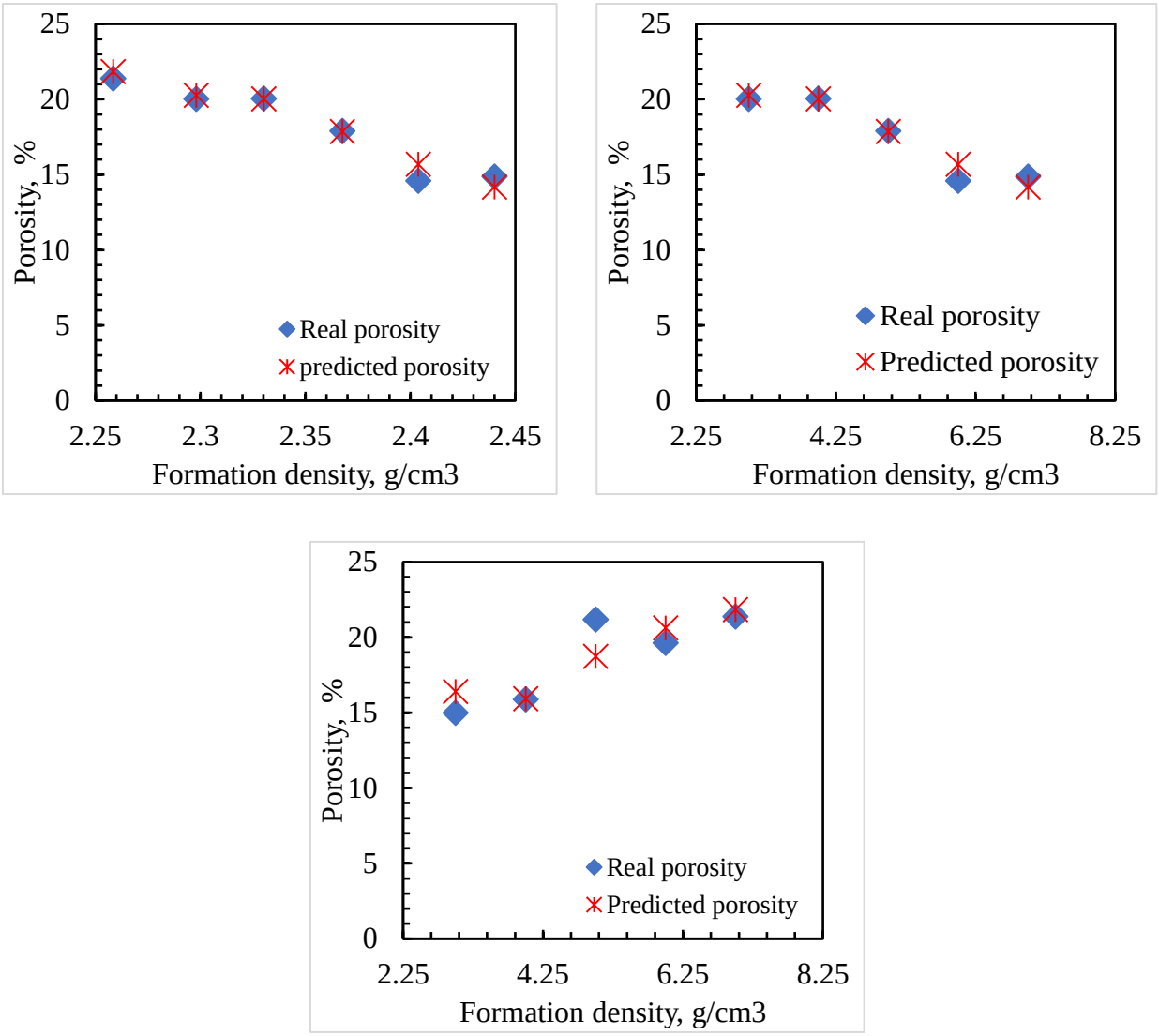

Figure III-4. Trend prediction capability of the RHOZ (a), PEFZ (b) and NPHI (c)

Conclusion

In this study, MLP and SVR were applied to build a new accurate model for predicting porosity. These models were using 292 data points collected from Libya field. Several statistical parameters and graphical assessment were considered to check the performance of the established porosity models.

The main conclusions that can be drawn from this study are summarized as follows:

1. Statistical analysis revealed that both MLP and SVR models achieved good performance, with SVR outperforming slightly the MLP model.

2. Trend analysis confirmed that the proposed model effectively captured the expected physical between porosity and the input variables.
3. The results of this study confirmed the huge contributions of ML algorithms for dealing with petroleum data.
4. Beside the high performance achieved, the data size employed in this study is too small. More data need to be collected in further works to build a new porosity model with a better generalization capacity.

Overall, this study highlights the effectiveness of machine learning, particularly SVR and MLP, in accurately predicting porosity within the studied context. These findings can be valuable for geoscientists and reservoir engineers for accurate reservoir characterization and exploration activities.

Conclusions and Perspectives

In this thesis, we investigated the application of machine learning algorithms, namely Multi-Layer Perceptron (MLP) and Support Vector Regression (SVR), for predicting porosity from well logging data. Both models achieved good performance in predicting porosity, with SVR exhibiting a slight edge based on the employed statistical analysis. The established models effectively captured the relationship between porosity and the input well logging variables, demonstrating the potential of machine learning for accurate porosity prediction in reservoir characterization.

However, it is crucial to acknowledge the limitation of the current model's generalizability due to the relatively small dataset (292 data points) used in this study. Future research should focus on incorporating more data to enhance the model's robustness and applicability across diverse geological settings.

The finding of this study indicated that the machine learning algorithms can be emerged as a powerful technique for estimating the porosity based on well logging data. This alternative approaches can contribute significantly to the oil and gas field by saving both time and money spent on conducting laboratory and core analysis.

Future works can be done for predicting other reservoir properties and construct a white box program that can allow a precise reservoir characterization based on machine learning algorithms and the collected well data. This include the relative permeability, capillarity pressure and other fluid properties such as CO_2 solubility and oil viscosity.

Based on the finding of this study, the following recommendations can be considered for further works in order to achieve a good prediction accuracy of reservoir properties:

- Huge data should be collected for training the ML models. This imply that the company should facilitate the data diffusion through online sharing platform.
- Many other recent algorithms should be investigated and compared with the most well-known ML algorithms during the training process.

References

[1] Suyun HU, Bin BAI, Shizhen TAO, Congsheng B, Zhang T, Yanyan C, et al. Heterogeneous geological conditions and differential enrichment of medium and high maturity continental shale oil in China. Pet Explor Dev 2022;49:257–71.

[2] Sun P, Xu H, Zhu H, Jia L, Hu X, Fang H, et al. Investigation of pore-type heterogeneity and its control on microscopic remaining oil distribution in deeply buried marine clastic reservoirs. Mar Pet Geol 2021;123:104750.

[3] Sanni M. Petroleum engineering: principles, calculations, and workflows. vol. 237. John Wiley & Sons; 2018.

[4] Ahmed T. Reservoir engineering handbook. Gulf professional publishing; 2018.

[5] Darling T. Well logging and formation evaluation. Elsevier; 2005.

[6] Bonnecaze RT, Sharma MM, Butler JE, Arboleda G. High resolution downhole measurements of porosity and fluid saturation while core drilling. SPE Annu. Tech. Conf. Exhib., SPE; 2002, p. SPE-77561.

[7] Kane JA, Jennings JW. A method to normalize log data by calibration to large-scale data trends. SPE Annu. Tech. Conf. Exhib., SPE; 2005, p. SPE-96081.

[8] McPhee C, Reed J, Zubizarreta I. Core analysis: a best practice guide. Elsevier; 2015.

[9] Gamal H, Elkatatny S. Prediction model based on an artificial neural network for rock porosity. Arab J Sci Eng 2022;47:11211–21.

[10] Sircar A, Yadav K, Rayavarapu K, Bist N, Oza H. Application of machine learning and artificial intelligence in oil and gas industry. Pet Res 2021;6:379–91.

[11] Hegde J, Rokseth B. Applications of machine learning methods for engineering risk assessment–A review. Saf Sci 2020;122:104492.

[12] Gamal H, Elkatatny S, Abdulraheem A. Rock drillability intelligent prediction for a complex lithology using artificial neural network. Abu Dhabi Int. Pet. Exhib. Conf., SPE; 2020, p. D021S030R003.

[13] Hegde C, Pyrcz M, Millwater H, Daigle H, Gray K. Fully coupled end-to-end drilling optimization model using machine learning. J Pet Sci Eng 2020;186:106681.

[14] Boukredera FS, Youcefi MR, Hadjadj A, Ezenkwu CP, Vaziri V, Aphale SS. Enhancing the drilling efficiency through the application of machine learning and optimization algorithm. Eng Appl Artif Intell 2023;126:107035.

[15] Youcefi MR, Hadjadj A, Bentriou A, Boukredera FS. Real-Time Prediction of Plastic Viscosity and Apparent Viscosity for Oil-Based Drilling Fluids Using a Committee Machine with Intelligent Systems. Arab J Sci Eng 2021:1–14.

[16] Elkatatny S, Mousa T, Mahmoud M. A new approach to determine the rheology parameters for water-based drilling fluid using artificial neural network. Soc Pet Eng - SPE Kingdom Saudi Arab Annu Tech Symp Exhib 2018, SATS 2018 2018. https://doi.org/10.2118/192190-ms.

[17] Amar MN, Ghriga MA, Hemmati-Sarapardeh A. Application of gene expression programming for predicting density of binary and ternary mixtures of ionic liquids and molecular solvents. J Taiwan Inst Chem Eng 2020.

[18] Menad NA, Noureddine Z, Hemmati-Sarapardeh A, Shamshirband S, Mosavi A, Chau K.

Modeling temperature dependency of oil-water relative permeability in thermal enhanced oil recovery processes using group method of data handling and gene expression programming. Eng Appl Comput Fluid Mech 2019;13:724–43.

[19] Amar MN, Zeraibi N, Redouane K. Optimization of WAG process using dynamic proxy, genetic algorithm and ant colony optimization. Arab J Sci Eng 2018;43:6399–412.

[20] Al Khalifah H, Glover PWJ, Lorinczi P. Permeability prediction and diagenesis in tight carbonates using machine learning techniques. Mar Pet Geol 2020;112:104096.

[21] Erofeev A, Orlov D, Ryzhov A, Koroteev D. Prediction of porosity and permeability alteration based on machine learning algorithms. Transp Porous Media 2019;128:677–700.

[22] Alloush RM, Elkatatny SM, Mahmoud MA, Moussa TM, Ali AZ, Abdulraheem A. Estimation of geomechanical failure parameters from well logs using artificial intelligence techniques. SPE Kuwait Oil Gas Show Conf., SPE; 2017, p. D031S010R002.

[23] Rafik B, Kamel B. Prediction of permeability and porosity from well log data using the nonparametric regression with multivariate analysis and neural network, Hassi R'Mel Field, Algeria. Egypt J Pet 2017;26:763–78.

[24] Ahmadi MA, Chen Z. Comparison of machine learning methods for estimating permeability and porosity of oil reservoirs via petro-physical logs. Petroleum 2019;5:271–84.

[25] Bhatt A, Helle HB. Committee neural networks for porosity and permeability prediction from well logs. Geophys Prospect 2002;50:645–60.

[26] Gholami R, Shahraki AR, Jamali Paghaleh M. Prediction of hydrocarbon reservoirs permeability using support vector machine. Math Probl Eng 2012;2012.

[27] Monicard RP. Properties of reservoir rocks: core analysis. Editions Technip; 1980.

[28] Vapnik VN. The nature of statistical learning. Theory 1995.

[29] YOUCEFI Mohamed Riad. Utilisation des réseaux de neurones dans l'optimisation des paramètres de forage des puits des hydrocarbures en temps réel. UNIVERSITE M'HAMED BOUGARA-BOUMERDES, 2021.

[30] Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: Machine learning in Python. J Mach Learn Res 2011;12:2825–30.

[31] Gharbi RBC, Mansoori GA. An introduction to artificial intelligence applications in petroleum exploration and production. J Pet Sci Eng 2005;49:93–6. https://doi.org/10.1016/j.petrol.2005.09.001.

[32] Vaferi B, Eslamloueyan R, Ayatollahi S. Application of Recurrent Networks to Classification of Oil Reservoir Models in Well-testing Analysis. Energy Sources, Part A Recover Util Environ Eff 2015;37:174–80. https://doi.org/10.1080/15567036.2011.582610.

[33] Agwu OE, Akpabio JU, Alabi SB, Dosunmu A. Artificial intelligence techniques and their applications in drilling fluid engineering: A review. J Pet Sci Eng 2018;167:300–15. https://doi.org/10.1016/j.petrol.2018.04.019.

[34] Witek-Krowiak A, Chojnacka K, Podstawczyk D, Dawiec A, Pokomeda K. Application of response surface methodology and artificial neural network methods in modelling and optimization of biosorption process. Bioresour Technol 2014;160:150–60. https://doi.org/10.1016/j.biortech.2014.01.021.

[35] Amato F, López A, Peña-Méndez EM, Vaňhara P, Hampl A, Havel J. Artificial neural networks in medical diagnosis. J Appl Biomed 2013;11:47–58. https://doi.org/10.2478/v10136-012-0031-x.

[36] A.H. Zaabab, Q.-J. Zhang MN. A neural network modeling approach to circuit optimization and statistical design. Adv Sci Lett 2016;22:555–6. https://doi.org/10.1166/asl.2016.6861.

[37] Tighilt Y, Bouttout F, Khellaf A. Modeling and design of printed antennas using neural networks. Int J RF Microw Comput Eng 2011;21:228–33. https://doi.org/doi:10.1002/mmce.20509.

[38] Aiken M, Krosp J, … MV-J of EU, 1995 undefined. A neural network for predicting total industrial production. IDEA Gr Publ n.d.

[39] Kumar M, Yadav N. Multilayer perceptrons and radial basis function neural network methods for the solution of differential equations: A survey. Comput Math with Appl 2011;62:3796–811. https://doi.org/10.1016/j.camwa.2011.09.028.

[40] Hu X, Weng Q. Estimating impervious surfaces from medium spatial resolution imagery using the self-organizing map and multi-layer perceptron neural networks. Remote Sens Environ 2009;113:2089–102. https://doi.org/10.1016/j.rse.2009.05.014.

[41] Rebai N, Hadjadj A, Benmounah A, Berrouk AS, Boualleg SM. Prediction of natural gas hydrates formation using a combination of thermodynamic and neural network modeling. J Pet Sci Eng 2019;182:106270.

[42] Orhan U, Hekim M, Ozer M. EEG signals classification using the K-means clustering and a multilayer perceptron neural network model. Expert Syst Appl 2011;38:13475–81. https://doi.org/https://doi.org/10.1016/j.eswa.2011.04.149.

[43] Hemmati-Sarapardeh A, Ghazanfari M-H, Ayatollahi S, Masihi M. Accurate determination of the CO2-crude oil minimum miscibility pressure of pure and impure CO2 streams: A robust modelling approach. Can J Chem Eng 2016;94:253–61. https://doi.org/10.1002/cjce.22387.

[44] Hemmati-Sarapardeh A, Varamesh A, Husein MM, Karan K. yy Modeling and data assessment. Renew Sustain Energy Rev 2018;81:313–29. https://doi.org/10.1016/j.rser.2017.07.049.

[45] Menad NA, Noureddine Z. An efficient methodology for multi-objective optimization of water alternating CO2 EOR process. J Taiwan Inst Chem Eng 2019;99:154–65. https://doi.org/10.1016/j.jtice.2019.03.016.

[46] Fattah MA, Karim KH. Performance of linear models in predicting cation exchange capacity of calcareous soils. Iraqi J Agric Sci 2021;52:1489–97.

Abstract

In the oil and gas industry, porosity is a crucial parameter for determining the viability of a reservoir. Traditionally, porosity is measured by collecting core samples from a well, which is a time-consuming and expensive process. This study aims to investigate the effectiveness of two Machine learning (ML) algorithms, including Support Vector regressor (SVR) and Multi-Layer Perceptron (MLP) for porosity prediction. A dataset of well logs, consists of 292 data points collected from Libya field, was utilized to train and test the ML models. The well logs includes core porosity measurement as a target variable and five input geological measurements such as gamma ray (GR), photoelectric (PE), neutron porosity (NPHI), shallow resistivity (RXOZ) and bulk density (RHOZ). The obtained results reveal that the Both MLP and SVM exhibited good performance in predicting porosity, with the SVR model achieving slightly better results. The MLP model yielded a Root Mean Square Error (RMSE) of 1.3729, Average Absolute Percentage Error (AAPE) of 4.6930, and coefficient of determination (R^2) of 0.8952. The SVR model achieved an RMSE of 1.52052, AAPE of 4.5082, and R^2 of 0. 88321. Furthermore, the findings of this study demonstrate the potential of machine learning algorithms, particularly SVR, for accurate porosity prediction using well logging data. This approach offers a more efficient and cost-effective alternative to traditional core analysis methods for reservoir characterization in the oil and gas industry.

الملخص

في صناعة النفط والغاز، تعتبر المسامية عاملاً حاسماً لتحديد صلاحية المكمن. تقليديا، يتم قياس المسامية عن طريق جمع العينات الأساسية من البئر، وهي عملية تستغرق وقتا طويلا ومكلفة. تهدف هذه الدراسة إلى التحقق من فعالية للتنبؤ بالمسامية. تم استخدام مجموعة MLP خوارزمية SVR ، بما في ذلك خوارزمية (ML) خوارزميتين للتعلم الآلي بيانات من سجلات الآبار، تتكون من 292 نقطة بيانات تم جمعها من حقل ليبيا، لتدريب واختبار نماذج التعلم الآلي. تشتمل ، سجلات البئر على قياس المسامية الأساسية كمتغير مستهدف وخمسة قياسات جيولوجية مدخلة مثل أشعة جاما SVM و MLP تكشف النتائج التي تم الحصول عليها أن كلا من. والكهروضوئية، ومسامية النيوترونات، والمقاومة والكثافة قدره RMSE عن خطأ MLP لنتائج أفضل قليلاً. أسفر نموذج SVR أظهرا أداءً جيدًا في التنبؤ بالمسامية، مع تحقيق نموذج قدره 0.8952. حقق نموذج (R^2) قدره 4.6930، ومعامل التحديد AAPE 1.3729، ومتوسط النسبة المئوية للخطأ المطلق تبلغ 0.88321. علاوة على ذلك، توضح نتائج هذه R^2 تبلغ 4.5082، و AAPE تبلغ 1.52052، و RMSE قيمة SVR ، للتنبؤ الدقيق بالمسامية باستخدام بيانات تسجيل الآبار. يوفر هذا SVR الدراسة إمكانات خوارزميات التعلم الآلي، وخاصة .النهج بديلاً أكثر كفاءة وفعالية من حيث التكلفة لطرق التحليل الأساسية التقليدية لتوصيف المكامن في صناعة النفط والغاز