

# GEOPHYSICS®

## Petrofacies classification using machine learning algorithms

|                    |   |
|--------------------|---|
| Journal:           | <i>Geophysics</i>   |
| Manuscript ID      | GEO-2019-0439.R2  |
| Manuscript Type:   | Machine Learning and data analytics for geoscience applications |
| Keywords:          | artificial intelligence, reservoir characterization, lithology  |
| Area of Expertise: | Case Histories  |

SCHOLARONE™  
Manuscripts

## PETROFACIES CLASSIFICATION USING MACHINE LEARNING ALGORITHMS

Adrielle A. Silva<sup>1</sup>, Mônica W. Tavares<sup>1</sup>, Abel Carrasquilla<sup>1</sup>, Roseane Misságia<sup>1</sup>, Marco Ceia<sup>1</sup>

<sup>1</sup> Darcy Ribeiro Northern Rio de Janeiro State University, Petroleum Engineering and Exploration Laboratory, Macae-RJ, Brazil. E-mail: adrielle@lenep.uenf.br, monicawtavares@gmail.com, abelgonca@gmail.com, rose@lenep.uenf.br, marco@lenep.uenf.br.

Original paper date of submission: 1st July 2019

## ABSTRACT

Carbonate reservoirs represent a large portion of the world's oil and gas reserves, exhibiting specific characteristics that pose complex challenges to the reservoirs characterization, production and management. Therefore, the evaluation of the relationships between the key parameters, such as porosity, permeability, water saturation and pore size distribution, is a complex task considering only well-log data, due to the geological heterogeneity. Hence, the petrophysical parameters are the key to assess the original composition and post-sedimentological aspects of the carbonate reservoirs. The concept of reservoir petrofacies was proposed as a tool for the characterization and prediction of the reservoir quality as it combines primary textural analysis with laboratory measurements of porosity, permeability, capillary pressure, photomicrograph descriptions and other techniques, which contributes to understanding the post-diagenetic events. This work proposes a workflow to petrofacies classification of a carbonate reservoir from Campos Basin in southeastern Brazil, employing the following machine learning methods: Decision Tree (DT), Random Forest (RF), Gradient Boosting (GB), K-Nearest Neighbors (KNN) and Naïve Bayes (NB). The dataset comprised 1477 wireline data from two wells (A3 and A10) that had petrofacies classes already assigned based on core descriptions. It was divided into two subsets, one for training and one for testing the capability of the trained models to assign petrofacies. The supervised-learning models have used labelled training data to learn the relationships between the input measurements and the petrofacies to be assigned. Additionally, we present a comparison of the models' performance using the testing set according to accuracy, precision, recall, and F1-score evaluation metrics. Then, the proposed approach has proved to be valuable allies for the petrofacies classification, especially for analyzing a well-logging database with no prior petrophysical information.

## INTRODUCTION

Carbonate rocks often display characteristics that distinguish them from other rocks. Some carbonate rocks consist of grains that have undergone mechanical transport in the depositional environment; others are geochemically generated on-site; and some, from the first moment of deposition, undergone chemical and mechanical diagenetic transformations. Most of the carbonate reservoirs consist primarily of limestones and dolostone, being the vast majority of carbonate sediments produced and accumulated in a submerged marine platform (the pre-salt microbialites are formed in salt lagoons) under appropriate temperatures and water conditions. Therefore, evaluating the origin, composition, and post-sedimentological aspects (i.e., textures and structures) of these rocks can facilitate their petrophysical characterization and thus result in a more accurate description of a carbonate reservoir.

Considering that these rocks represent a large portion of the world's oil and gas reserves, being able to define the types of fluids and the amount of hydrocarbons these reservoirs contain is of fundamental importance to the oil industry. Additionally, an essential component of reservoir characterization is the description of the pore system, including its geometry, topology, and effective volume; these factors determine the production potential of the reservoir (Kopaska-Merkel and Friedman, 1989).

Reservoir petrofacies are defined by the combination of the main attributes that affect reservoir quality, including depositional structures, textures, primary composition, diagenetic processes and products, and pore types. De Ros and Goldberg (2007) demonstrate that petrofacies could be used as a tool for reservoir characterization and modeling since the combination of primary textural and compositional aspects with specific diagenetic processes

## Geophysics

4

and products corresponds to defined value ranges of porosity and permeability; as well as characteristic log and seismic signatures.

Petrofacies are usually defined by using core samples associated with wireline logs. However, recovering a core sample is costly as it requires stopping the drilling process as well as conducting a subsequent laboratory analysis. Several authors propose apply machine learning (ML) algorithms to evaluate petrofacies/lithofacies distribution in oil wells, to minimize costs and increase the accuracy (Baldwin et al., 1990; Avseth and Mukerji, 2002; Qi and Carr, 2006; Silva et al., 2015; Sebtosheikh and Salehi, 2015; Xie et al., 2017; Bhattacharya and Mishra, 2018; Dell'Aversana, 2019).

The main objective of this study is to classify the petrofacies of an Albian carbonate reservoir by ML supervised algorithms. The petrofacies labels were defined based on geological interpretation and a mathematical model that integrates lithofacies, electrofacies and petrophysical laboratory measurements (Tavares, 2018). The dataset containing the labelled petrofacies was addressed to train the models. We present a comparison of five ML approaches, as Decision Tree (DT), Random Forest (RF), Gradient Boosting (GB), K-Nearest Neighbors and Naïve Bayes (NB), to petrofacies classification using data from wireline measurements of two oil wells (A3 and A10). The classification results of the ML approaches determine the models that could potentially assist in the characterization of reservoirs, especially when no petrophysical information is available.

## GEOLOGIC SETTING

This study focuses on the Albian carbonate reservoir located in the Campos Basin (Figure 1), an important and prolific oil area along the Brazilian continental margin, currently accounting for more than 45% of the country's national oil production (ANP, 2019). It is located in

southeastern Brazil with approximately 100,000 km<sup>2</sup> of area, bordered by the state of Espírito Santo to the north and Alto do Cabo Frio to the south. This basin contains a carbonate formation with significant lithological variation, mainly composed of grainstones, packstones, oncolites, peloids, oolites, and rare bioclasts. The carbonaceous stacking cycles consist of an upward overlapping of wackestones, oolytic/oncolytic packstones and oolitic/oncolytic grainstones.

[Figure 1]

We have used data from A3 and A10 wells to create the ML models because only these wells have petrophysical data available. Figure 2 presents the location of these wells showing the facies distribution obtained from core plugs sampled within 93m from A3 and 162m from A10. The photomicrographs from plug samples were associated with their energy zone: oolitic and oncolitic grainstones from high-energy zone (red frames), oncolitic peloidal grainstones and oncolitic bioclastic packstones from moderate-energy zone (orange frames) and, peloidal bioclastic packstones and wackestones from low-energy zone (blue frames) (Okubo, 2014).

[Figure 2]

## PETROFACIES MODELING

The electrofacies are a set of log responses that characterizes a sediment and allows it to be distinguished from others (Serra and Abbott, 1982). Because of the importance and applicability of studies that contemplate the electrofacies characterization using geophysical logs and direct geological data, this theme has been increasingly studied within the geological modeling of reservoirs (Serra, 1984; Doveton, 1994; Rider, 2002). Although the petrofacies

definition is similar to the electrofacies, it includes a description of the petrophysical laboratory characterization, such as gas porosity, permeability tests, and capillary pressure. Consequently, petrofacies characterization takes account the porous geometry, fluid saturation and its distribution, both associated to the permoporous characteristics that resulted from depositional control to identify the conditions of fluids storage and flow (Doveton et al., 2000). Thus, the petrofacies allow distinguishing intervals of similar textural and mineralogical composition, diagenesis processes and products, and pore types and distribution (Rushing et al., 2008).

Tavares (2018) defined the petrofacies of both wells A3 and A10 using 13 mercury injection capillary pressure ( $PC_{HG}$ ) samples. The pore throat radii were calculated for the accumulated saturation of 35% MICP ( $R_{35}$ ) and an empirical relationship between throat radii ( $R_g$ ), porosity ( $\phi$ ), and absolute permeability ( $k$ ) was established using the following equation (Winland, 1972):

$$\log(R_g) = c_1 \cdot \log(k) + c_2 \cdot \log(\phi) + c_3, \quad (1)$$

where  $c_1$  and  $c_2$  are the permeability and porosity coefficients, respectively, and  $c_3$  is the intercept with y-axis. Pittman (1992) modified Winland's (1972) equation, using other coefficients, but both found an accumulated saturation of the Pearson correlation coefficient  $R^2$  larger than 0.90. Table 1 shows the coefficients of this equation in accord with these two authors and, the modification of the coefficients made by Tavares (2018) through MLR.

#### [Table 1]

The  $PC_{HG}$  samples were used to verify the Winland's model and assess the basic petrophysics coefficients. Therefore, the experimental permeability and porosity assisted in

predicting the permeability from well-logs data, enabling the model upscaling through multilinear regression (equation 2). The Tavares (2018) equation's choice was based on estimate an absolute permeability value  $k$  that better fit the experimental  $k$ .

$$\log(k) = 1.293\log(R_{35}) + 0.063\log(\phi) + 1.238. \quad (2)$$

Frequently, large pore throat size is related to better petrofacies and reservoir qualities, showing homogeneous pore structure, and a linear porosity/permeability relationship. Hereafter, three units were established to express the permoporous quality and petrofacies of the rock, supported by the lithofacies and electrofacies supervision, as shown in Table 2. Figure 3 shows mercury injection capillary pressure curves in which was chosen only the curves capable of distinguishing the three different energies of the depositional environments (Kolodzie, 1980).

Figures 4(a) and 4(b) show the relationship between permeability and porosity using the laboratory measurements of the cored samples for comparison between textures and the three clustered petrofacies. In this figure, permeability values higher than 100mD (3.2 microns) were presented by facies with best reservoir qualities; whereas intermediate facies showed permeability between 10mD and 100mD (0.6 microns) and facies with poor qualities, values less than 10mD.

[Table 2]

[Figure 3]

[Figure 4]

## MACHINE LEARNING APPROACHES

In ML and statistics, classification refers to the problem of identifying to which set of categories a new observation belongs, based on a training set containing observations, the category membership of which is known. This paper addresses the petrofacies classification problem by applying a simple and ensemble data classifier algorithms including DT, RF, GB, K-NN and NB classifiers. When selecting the algorithm, we took into account the distinction between mathematical and statistical fundamentals, which helped us to evaluate the information from different perspectives.

### Decision Tree

A Decision Tree (DT) is a machine learning method used for classification problems (Quinlan, 1986). As with many other ML techniques, the learning in DT is performed using a dataset of already classified instances to build a DT model in order to classify unseen data. This classifier has a tree-shaped structure, and each internal node represents a test on a feature; each leaf node represents a class label, and branches represent conjunctions of features that lead to said class labels. The paths from the root (the top-most node) to the leaves represent classification rules. In a DT, all possible scenarios of a problem are considered based on certain transparent conditions. The process of classification is observable in a white box model, which makes the classification procedure explicit. The DT model is advantageous because of its comprehensive and transparent nature, computational simplicity, and ease of application.

**Random Forest**

Random Forest (RF) is an ensemble ML algorithm that is generally used for both classification and regression problems. RF applies the technique of bootstrap aggregation to DT learners, which involves building multiple tree classifiers and making predictions based on the averaged predictions of several independent trees (Breiman, 1996; 2001). The RF classification algorithm has two phases. First, it generates training data for each tree by sampling with replacement several samples equal to the number of samples in the original dataset. Second, the RF algorithm classifies the DTs and implements a simple vote for each; the classification vote that occurs most frequently is the result of the prediction. One of the main concepts in the RF algorithm is that only a subset of all features is considered for splitting each node in each DT. This results in trees with distinct predictors at the top split, leading to a more accurate average yield in decorrelated trees.

**Gradient Boosting**

Gradient Boosting (GB) is used to improve the prediction of binary outcomes using weighted ensembles of DTs, called base-learners (Freund and Schapire 1996, 1997). In GB machines, the learning procedure is an iterative process. It begins with an initial estimation of a function using a constant offset that fits the data poorly. The fit is improved in each successive iteration by adjusting a base-learner to the negative gradient of a pre-specified loss function. The prediction function is estimated with the negative gradient, and the function approaches the true form of the relationship with successive iterations (Maloney et al., 2012).

### K-Nearest Neighbors

The K-Nearest Neighbors (KNN) model predicts the class value of an unlabeled example by analyzing its K neighboring examples (Cover and Hart, 1967). During classification, individual test samples are compared locally to K neighboring training samples in variable space. Neighbors are commonly identified using a Euclidian or Manhattan distance metric (Yu et al., 2011). Predictions are based on a majority vote cast by neighboring samples, each neighbor votes for its class value and their vote is usually weighted according to the distance metric adopted. The result of the voting is a probabilistic class prediction, and the sample can be labeled with the most probable value of class. As high K-Neighbors value can lead to overfitting and model instability, appropriate values must be selected for a given application (Hastie et al., 2009).

### Naïve Bayes

Naïve Bayes (NB) is a probabilistic ML model used for classification problems. It is based on Bayes' s theorem under the assumption that for a given class the inputs are independent of one another (Sammut and Webb, 2011). NB reduces the problem of discriminating classes to finding class-conditional marginal densities, which represent the probability that a given sample is one of the possible target classes. NB usually performs well against other alternatives unless the data contains correlated inputs (Hastie et al., 2009).

### Evaluation metrics

To test and evaluate the performance of the algorithms, we used four standard metrics: accuracy, precision, recall and F1-score. Accuracy, calculated by equation 3, is the percentage of correctly classified instances over the total number of instances:

$$Accuracy = \frac{1}{N} \sum_{i=1}^N I(f(x_i) = y_i), \quad (3)$$

where  $f(x_i)$  is the predicted class of a test sample and  $y_i$  is the true class of this sample. Consider that  $I(\text{true}) = 1$  and  $I(\text{false}) = 0$ .

Accuracy alone generally is not sufficient to measure the classifier performance, especially when the datasets have some classes that occur much more frequently than others. To fully evaluate the effectiveness of a model, it is necessary also to examine precision and recall. Precision, determined by equation 4, represents the number of class members classified correctly over the total number of instances classified as class members:

$$Precision = \frac{TP}{TP + FP}, \quad (4)$$

TP is the number of true positives, and FP is the number of false positives. Recall measures the percentage of actual positive samples that were classified as positive, as shown in equation 5:

$$Recall = \frac{TP}{TP + FN}. \quad (5)$$

The F1-score (equation 6), represents a harmonic mean of both precision and recall. The regular mean treats all values equally, whereas the harmonic mean gives much more weight to low values. As a result, the classifier will only obtain a high F1-score if both recall and precision are high.

$$F_1 = \frac{TP}{TP + \frac{FN + FP}{2}}. \quad (6)$$

8  
9  
10 METHODOLOGY  
11  
1213  
14 Scikit-learn  
15  
1617  
18  
19 The ML models in this paper are built using the scikit-learn (Pedregosa et al., 2011) in  
20 Python3. The scikit-learn is a Python module that integrates a wide range of ML algorithms for  
21 both supervised and unsupervised problems. This study uses “DecisionTree”,  
22 “RandomForestClassifier”, “GradientBoostingClassifier”, “KNeighborsClassifier” and  
23 “GaussianNB” to create the supervised learning classifiers.  
24  
2526  
27 Dataset  
28  
2930  
31  
32 The data analysis and exploration was combined with 13 mercury injection capillary  
33 pressure measurements, 493 basic petrophysical evaluations of porosity (L) and permeability  
34 (kL) (154 in A3 and 339 in A10), total porosity (T) derived from the RHOB log and,  
35 permeability k<sub>E</sub> estimated by a multiple linear regression described in Petrofacies section. The  
36 data spacing of k and φ core samples is 30/30 cm for each meter of rock (Tavares, 2018).  
37  
3839  
40  
41 The dataset used in this study consists of 1,477 two wells measurements (A3 and A10),  
42 which include the well logs shown in Table 3. The depth interval recorded for well A3 was  
43 1,775-1,920m (726 samples) and for A10, 1,750-1,900m (751 samples). The log measurements  
44 spacing was 20cm (Petrobras, 2012).  
45  
4647  
48  
49 The data from both wells were combined to compose the entire dataset used for the  
50 classification. By doing so, the training set was increased, and the accuracy of the classifiers was  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

improved since the two wells are in neighboring regions with the same geological characteristics.

Table 4 shows the total of samples from each petrofacies.

[Table 3]

[Table 4]

As a pre-processing step, data normalization was applied to increase the data input consistency, because some ML algorithms (like KNN) usually obtain better results when attributes are in similar units and scales.

## Feature selection

Feature selection consisted of choosing a subset of features for improving prediction accuracy or decreasing the size of the data structure without significantly decreasing the prediction accuracy of the classifier (Koller and Sahami, 1996). Reducing the dimension based on feature selection can speed up the learning process, improve prediction accuracy and reduce overfitting.

Some features are more powerful for classification, whereas others may be redundant and thus should not be included in the learning and classification processes. One method that should be used to select the most relevant input variables is calculating the relative importance of the features. Breiman et al. (2001) proposed that a feature's importance should be evaluated based on the reduction of model accuracy using ensembles of randomized trees. This analysis, based on Gini impurity index, is known as “Gini importance” and is given by equation 8:

$$\text{Importance}(X_m) = \frac{1}{N_t} \sum_T \sum_{t \in T: v(S_t)} p(t) \Delta_i(S_t, t), \quad (8)$$

1  
2  
3  
4  
5  
6  
7  
where the relative importance of a feature  $X_m$  for predicting the target  $y$  is calculated by adding  
the weighted impurity decreases  $p(t)\Delta_i(S_t, t)$  for all nodes  $t$  where  $X_m$  is used, averaged over all  
 $N_t$  trees in the forest. The  $p(t)$  is the proportion  $\frac{N_t}{N}$  of samples reaching  $t$ , and  $v(S_t)$  is the feature  
used in the split  $S_t$ .

1  
2  
3  
4  
5  
6  
7  
8  
9  
9  
Based on a tree-based ensemble method, we produce a five-fold cross-validation test to  
determine the most relevant features in the dataset; these features were then chosen to build the  
classifiers. Figure 5 presents a summary of the predictive power of each variable for the dataset  
used for the classification. As is already known, GR log is a good predictor of the petrofacies in  
this area due to the presence of various granites and sediments, all of which have distinct  
radioelement signatures. Using this information, we consider at each depth  $d$  of each well, a set  
of three scalar features: GR, DRDN and RHOB.

1  
2  
[Figure 5]

## 3 4 5 Data visualization

6  
7  
8  
A scatter matrix plot (Figure 6) was generated to visualize the variations between the log  
measurements in the dataset and find any pattern of common well-logs associated with  
petrofacies. Each pane shows the relationship between two of the variables on the x and y axes,  
with a density plot showing the distribution of each class along the diagonal. Each point is  
colored according to its petrofacies (P1, P2 or P3). Based on these plots, it is not evident what  
relationships exist between the measurements and petrofacies labels. However, the plot shows  
that all petrofacies clusters overlap, which is typical of the subsurface dataset and not easy to

1  
2  
3 separate using traditional statistical techniques. This plot justifies the application of ML to better  
4 analyze a multi-variate dataset and extract geologically meaningful features.  
5  
6

7 [Figure 6]  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

Figure 7 shows the data visualization obtained by radial coordinate visualization (RadViz). RadViz is a visualization technique based on Hooke's law that maps a set of n-dimensional points onto a plane; each feature, called dimensional anchors, is placed around the circle attached to one data point through imaginary springs. The data point is plotted where the sum of all spring forces equals zero according to Hooke's law ( $F = Kx$ ), where the force is proportional to the distance  $x$  to the anchor point. The value  $K$  for each spring is the value of the feature for the data point; higher anchor values "pull" the data points closer to the anchor on the circumference (Hoffman et al., 1997). Figure 7(a) shows RadVizs of all attributes from the dataset (11 well-logs), resulting in an uninformative visualization since it shows all classes overlapping. However, in Figure 7(b), it is evident that for the three most relevant well logs, the classes are better separated. From the position of the points in this figure, we can observe that some samples of P3 lie close to the positions of GR attribute, which means that P3 is more influenced by GR than by DRDN or RHOB, whereas P1 and P2 tend to be more influenced by RHOB and DRDN measurements.

[Figure 7]

## Dataset Split

The entire dataset was divided into two subsets: the training/validation and the blind test dataset. The training/validation dataset was first used to train the model. The blind test set was only used for the final model assessment. In our research, the blind test dataset comprised 20%

of the entire dataset. The remaining data (training/validation set) were devoted to performing a cross-validation evaluation and then used to fit the final model. Table 5 shows the dataset distribution after splitting, where it is possible to observe that there are much fewer P1 samples compared with others, resulting in an imbalanced class dataset. To solve that imbalance and improve the learning rate of the model, we over-sampled P1 and P2 data points.

[Table 5]

## Over-sampling

One dataset is considered imbalanced when the important class (minority class) has significantly fewer samples than the majority class. When the minority data points are more important than the majority ones, and the main goal is to classify those minority data points correctly, standard ML usually results in hyperplanes that favor the majority class and thus generalize poorly.

Class imbalance presents a significant challenge for classification algorithms whenever the cost loss for the minority class is higher compared with that of the majority class. Over-sampling methods can improve classifiers by identifying more instances in the minority class. The synthetic minority over-sampling technique (SMOTE) algorithm (Chawla et al., 2002) over-samples the minority class by introducing artificial minority samples between a given minority data point and its nearest neighbors.

Therefore, we applied SMOTE to balance the sample rate of the P1 and P2 class. The over-sampling increases the error and slightly reduces the accuracy during the training/validation but makes the model easier to train and generalize. Figure 8 shows the total samples from each class and a P1 and P2 class distribution before and after the SMOTE. The SMOTE resultant dataset was considered the new training/validation set for the following steps. After using SMOTE, the areas with many minority instances got more new minority instances, and the sparse areas are still sparse.

[Figure 8]

## Optimize parameters

ML applications require the selection of one or more parameters to optimize the classifier training in each specific application (Kovacevic et al., 2009; Yu et al., 2012). There are two types of parameters: some learned from the training data and the specifics of a learning algorithm. In contrast to the parameters learned during training, model hyper-parameters are defined before training, governing the implementation aspects of the model. These parameters are commonly optimized separately.

The hyper-parameter values should be optimized carefully because the optimal values are dependent on the dataset. One of the approaches for optimizing hyper-parameters is the Grid Search, which was also available in scikit-learn package. This is an exhaustive, brute-force search method in which a list of values for different hyper-parameters were specified; finally, we evaluate the model performance for each combination of hyper-parameters with a final aim of obtaining the optimal set of parameters.

A trial-and-error approach using a Grid Search algorithm (Raschka and Mirjalili, 2017) was applied to determine the best model parameters for each of the classifiers. The set of parameters that showed the best accuracy was selected for the training phase. Table 6 shows the parameter values that were used to build the classifier models.

### Cross-validation

Cross-validation involves partitioning the dataset into a training set, used to train the model, and an independent set used to evaluate the model. In K-fold cross-validation, the original sample is randomly partitioned into equal K sized subsamples called folds. In each iteration, a single subsample was retained for validating the trained model, and the remaining K-1 subsamples were used as training data.

[Table 6]

The training/validation set was first partitioned into ten parts of (approximately) equal size. Then, a sequence of models was trained. The first model was trained using the first fold as the validation set, and the nine remaining folds as the training set. It was repeated ten times, each using a different fold as the validation set. The estimated accuracy was averaged across all ten trials to obtain the total effectiveness of the model. Figure 9 shows the spread of classification accuracy rates for the classification methods resulting from the ten-fold cross-validation procedure with the default model parameters (Figure 9a) and with the tuned parameters (Figure 9b). For the majority of techniques, the maximum, minimum and mean values were higher for models with tuned parameters. NB has no parameter to optimize, presenting the same result for both tests. The RF model achieved the highest absolute accuracy (96.23%), followed by the

## Geophysics

19

KNN and GB models, which had maximum accuracy values approximately of 96.25% and 94.78%, respectively. The DT model presents a lower performance compared to RF, GB and KNN. It showed a small mean accuracy and a set of suspected outliers (open circles) in both tests. Otherwise, NB presented the lowest accuracy and a larger boxplot width, which means that the accuracy for this method varied more during cross-validation. For this model, more than half of the cross-validation tests achieved accuracy below 85%.

The purpose of cross-validation in the model building phase was to estimate the performance of this final model using new data. Each ML algorithm was fitted using the entire training/validation set to build the final model.

[Figure 9]

### Learning curves

To examine the effect of training and validation dataset size on the performance of the proposed classifiers and analyze the bias and variance to achieve generalization, we perform many simulations varying the number of samples that composes the dataset. It is helpful to know how our model benefits from additional training data and whether the model suffers from high bias or high variance error. Figure 10 illustrates the learning curve of the DT, RF, GB, KNN, and NB models. The orange lines show the average accuracy on the training folds, and the green lines show the average accuracy on the validation fold for different sizes of the initial training set. For RF and GB, we observe that as the training set size increases, the model cannot fit perfectly anymore the training set, decreasing the training accuracy. However, the model fitted better the validation set because it was trained with more data. Thus, the validation score

increased as more training points were added. The small gap between training and validation curves indicate that DT, RF and GB models presented an overall good bias-variance tradeoff. The KNN model presented an overfitting situation with 200 training samples; the model specializes in the training dataset, leading to a poor generalization and consequently, low validation accuracy. Nevertheless, this model tended to improve the validation accuracy by increasing the samples in the training set. Additionally, the NB model presented the training and validation curves close to each other at a relatively low score compared with the other models, showing an underfitting caused by high bias. It usually occurs when models are too simple that may not capture the complexity of a dataset.

[Figure 10]

## DISCUSSION

To evaluate the accuracy of the classifiers, we performed petrofacies classification using the blind test dataset and compared the predicted and the actual petrofacies. Figures 11 and 12 present the classification results for well A3 and A10, respectively. The measured well logs (GR, DRDN, and RHOB) are plotted on the left-hand side of the figure, whereas the defined and predicted petrofacies categories are plotted on the right-hand side; each use the same depth information.

The results from both well datasets show that tree-based ensemble methods (GB and RF) are the best petrofacies classifier in the blind-testing domain. The best predictive power should be associated with the combination of multiple weak classifiers (DT). For well A3 (Figure 11), RF performs better than GB, excepting for depth of 1840m. Otherwise, for A10 (Figure 12), GB

showed the best correlation. Note that in some depths (as 1785, 1822, 1850 and 1864m) all the methods differ from the target but agreed in the classification result. In this case, the petrofacies classification before ML methods application maybe could be revised by a geologist/interpreter in order to understand the reason for all those methods committing the same error. Table 7 shows the average accuracy, precision, recall, and F1-score of each method considering the blind dataset.

[Figure 11]

The confusion matrix represents the classification model performance in tabular form (Congalton and Green, 1998). The number of rows and columns in the matrix are equal to the total number of classes. The values in each cell of the matrix represent the counts of class predictions for the samples associated with a particular class; the main diagonal elements denote the number of correct decisions. In an ideal case using a perfect classifier, the confusion matrix should become an identity matrix with main diagonal elements equal to one and off-diagonal elements equal to zero. The comparison of the confusion matrix in Figure 13 shows that overall, our solution is more focused on diagonal elements and reduces the values of off-diagonal elements, indicating an overall improvement in prediction accuracy.

[Figure 12]

[Table 7]

[Figure 13]

The confusion matrix shows that each model has a distinct behavior for each petrofacies classes. All the models exhibit relatively high classification accuracies for P3 compared with the

other petrofacies, which could be associated with the homogeneous characteristics of P3 and the fact that more original samples from P3 are present in the dataset. The precision scores for the P3 prediction of DT, GB, RF and KNN classifiers are above 92%, whereas NB only presented 83%. Since there are significant variations in the lithology properties of the P1 and P3 classes, all techniques differentiated them with high efficiency. Furthermore, P2, which has a set of transitional characteristics, is more frequently, classified as P1 or P3. The difficulty in distinguishing the P1 and P3 samples (true) from the P2 samples (predicted) may be related to samples from transitional regions that present characteristics of P2 (e.g., texture, grain size, pore shape) but a range of permeability similar to P1 or P3. GB and RF each exceeded 88% in the accuracy score regarding P2 classification, whereas both DT and KNN showed approximately 85%, and NB achieved 80%. Although NB presented the lowest score for the P3 class, its accuracy regarding P1 classification was much closer to those of the ensemble models. KNN presented a good accuracy for P1 and P3 classification, concentrating the majority of errors in P2 classification. GB and RF achieved similar classification accuracies among the petrofacies. These results indicate that the choice between those models may be situational. Both are ensemble learning methods and make the predictions by combining the outputs of individual trees (DT). They differ in how they built the trees and in the order and method in which the results are combined. GB builds trees one at a time; each new tree helps correct errors made by the previously trained tree. RF trains each tree independently, using a random data sample retired of the entire dataset. This randomness makes the model less likely to overfit than GB. Even though KNN is also prone to overfitting, it presented a great performance in P1 and P3 prediction, achieving the highest accuracies.

Figure 14 compares the training and predicting computation times of the five methods. GB training generally requires more time because the trees are built sequentially. It presented the highest computation times for both training and predicting process, whereas RF was faster than GB on training and slower on prediction. RF could be chosen instead of GB when computation training time is an issue, primarily because the RF algorithm can be parallelized, thus decreasing the execution time.

[Figure 14]

## CONCLUSION

In this study, we applied five different ML models to petrofacies classification problem using well-log data obtained from two exploratory oil wells in the Campos Basin, southeastern Brazil. The dataset comprised eleven well measurements (features) in which we select the three most relevant (GR, DRDN, and RHOB) through the feature importance analysis. All instances (wells A3 and A10) were subsequently combined, normalized, and split into training/validation (80%) and testing (20%) datasets. The training/validation set presented an imbalanced distribution of the petrofacies class and, to boost the prediction power of the minority class (P1), we applied the SMOTE over-sampling algorithm, reducing this imbalance by changing the P1 and P2 sample distribution of the training set. A set of the optimal parameters to build each ML model was defined by a hyper-parameter tuning technique called grid search. The over-sampled dataset was used in ten-fold cross-validation for each ML model, training and evaluation the model simultaneously. Then, the final fitted model was applied to predict unseen data (blind test set), and the obtained evaluation metrics, accuracy, precision, recall, and F1-score, were compared to measure the performance of each classifier. In overall, GB achieved the best

1  
2  
3 performance and showed to be robust, presenting a small variation in the accuracy score.  
4  
5 However, when the data size is too large, RF might be preferred to reduce the computation time  
6 of the training phase. NB presented the lowest testing accuracy, possibly associated with the  
7 model's more straightforward hypothesis function, causing that some feature interactions were not  
8 well modelled. We analyzed the confusion matrix of the models to measure how effectively they  
9 determined each petrofacies class. The five models presented high uncertainty associated with P1  
10 samples because this petrofacies is the most challenging to find characteristic patterns since it is  
11 related to rocks with higher permeability, such as grainstones, which have significant variations  
12 in mineralogy, grain size distribution and texture.

13  
14  
15  
16  
17  
18 The following steps were taken in this analysis: feature selection, dataset visualization,  
19 dataset split (training/validation and test), over-sampling of the minority classes, hyper-  
20 parameters optimization, ten-fold cross-validation (making the necessary parameters adjustments  
21 until achieved the desired accuracy), and testing the model. Considering that carbonate rocks  
22 present a complex porous system and that petrophysical data based on core plugs are limited, this  
23 study can be considered a method for characterizing carbonate reservoirs through the integration  
24 of core and well data. This is a case study applied to Brazilian oilfields; however, the  
25 applicability of such methods may vary based on the location, reservoir type, depth,  
26 heterogeneity, and other reservoir parameters. Each method needs to be adapted to the  
27 training/testing dataset to be used.

## REFERENCES

ANP, 2019, National Agency of Petroleum, Natural Gas and Biofuels: Bulletin of Production of Oil and Natural Gas, External Circulation, 101, Development and Production Superintendence (SDP), January 2019, Rio de Janeiro. (In Portuguese).

Avseth, P. and T. Mukerji, 2002, Seismic lithofacies classification from well logs using statistical rock physics: *Petrophysics*, **43**, 70–81.

Baldwin, J. L., Bateman, R. M. and C. L. Wheatley, 1990, Application of a neural network to the problem of mineral identification from well logs: *The Log Analyst*, **31**, 279-293.

Bhattacharya, S. and S. Mishra, 2018, Applications of machine learning for facies and fracture prediction using Bayesian Network Theory and Random Forest: Case studies from the Appalachian basin, USA: *Journal of Petroleum Science and Engineering*, **170**, 1005–1017.

Breiman, L., 1996, Bagging predictors: *Machine Learning*, **24** (2), 123–140.

Breiman, L., 2001, Random forests: *Machine Learning*, **45** (1), 5–32.

Bruhn, C. H. L., Gomes, J.A.T., Lucchesi, C.D. and P.R.S Johann, 2003, Campos basin: reservoir characterization and management - Historical overview and future challenges: Paper OTC 15220 presented at the Offshore Technology Conference, Houston.

Chawla, N., Bowyer, K., Hall, L. and W. Kegelmeyer, 2002, SMOTE: Synthetic Minority Over-sampling Technique: *Journal of Artificial Intelligence Research*, **16**, 341-378.

## Geophysics

26

Congalton, R.G. and K. Green, 1998, Assessing the accuracy of remotely sensed data: principles and practices. New York: Lewis Publishers. 137p.

Cover, T.M., Hart, P.E., 1967. Nearest neighbor pattern classification: IEEE Transactions Information Theory IT, **13** (1), 21–27.

De Ros, L.F., and K. Goldberg, 2007, Reservoir petrofacies: a tool for quality characterization and prediction: Proceedings of the AAPG Annual Convention and Exhibition. Long Beach, 1–6.

Dell'Aversana, P., 2019, Comparison of different Machine Learning algorithms for lithofacies cladrassification from well logs: Bollettino di Geofisica Teorica ed Applicata, **60** (1), 69-80.

Doveton, J. H. 1994. Geologic log analysis using computer methods: AAPG Computer Applicationsin Geology, American Association of Petroleum Geologists, **2**.

Doveton, J. H., Watney, W.L. and W.J. Guy, 2000, Integrated analysis of reservoir petrofacies in platform carbonates of Kansas: techniques and case studies, in Johnson, K.S., ed., Platform carbonates in the southern Midcontinent, 1996 symposium: Oklahoma Geological Survey Circular, **101**, 223-225.

Freund, Y., and R.E. Schapire, 1996, Experiments with a new boosting algorithm. Machine Learning: Proceedings of the Thirteenth International Conference, 148–156.

1  
2  
3 Freund, Y., and R.E Schapire, 1997, A decision-theoretic generalization of on-line  
4 learning and an application to boosting: Journal of Computer and System Sciences, **55** (1), 119–  
5 139.  
6  
7  
8

9  
10 Hastie, T., Tibshirani, R., and J. Friedman, 2009, The Elements of Statistical Learning:  
11 Data Mining Inference and Prediction: Springer.  
12  
13  
14

15 Hoffman, P., Grinstein, G., Marx, K., Grosse, I., and E. Stanley, 1997, DNA visual and  
16 analytic data mining: Proceedings of the IEEE Visualization. 437–441.  
17  
18  
19

20 Koller, D. and M. Sahami, 1996, Toward optimal feature selection. In: Proceedings of  
21 ICML-96, 13th international conference on machine learning, 284–292.  
22  
23  
24

25 Kolodzie, S., 1980, Analysis of pore throat size and use of the Waxman-Smits equation to  
26 determine OOIP in Spindle Field, Colorado: SPE Annual Technical Conference and Exhibition,  
27 Dallas, Texas, USA.  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

Kopaska-Merkel, D.C., and G.M. Friedman, 1989, Supplementary data to “Petrofacies analysis of carbonate rocks: example from the lower Paleozoic Hunton Group of Oklahoma and Texas”: AAPG Bulletin, **73**(11), 1289–1306.

Kovacevic, M., Bajat, B., Trivic, B., and R. Pavlovic, 2009, Geological units classification of multispectral images by using Support Vector Machines: Proceedings of the International Conference on Intelligent Networking and Collaborative Systems, IEEE, 267–272.

## Geophysics

28

Maloney, K.O., Schmid, M. and D.E. Weller, 2012, Applying additive modelling and gradient boosting to assess the effects of watershed and reach characteristics on riverine assemblages: *Methods in Ecology Evolution*, **3** (1), 116-128.

Okubo, J., 2014, Albian limestones in the Campos Basin: facies, diagenesis and depositional model: M.S. thesis, UNESP, Rio Claro – SP, Brazil. (In Portuguese).

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., et al. 2011, Scikit-learn: machine learning in python: *Journal of Machine Learning Research*, **12**, 2825.

Petrobras. 2012. Oilfield A Report. Petrobras - UENF Agreement. (In Portuguese).

Pittman, E., 1992, Relationship of porosity and permeability to various parameters derived from mercury injection capillary pressure curves for sandstone: *AAPG Bulletin*, Tulsa, OK, USA, **76** (2), 191-198.

Qi, L. and T.R Carr, 2006, Neural network prediction of carbonate lithofacies from well logs, Big Bow and Sand Arroyo Creek fields, Southwest Kansas: *Computers & Geosciences*, **32**, 947–964.

Quinlan, J.R., 1986, Induction of Decision Trees: *Machine Learning*, **1**, 81-106.

Raschka, S. and V. Mirjalili, 2017, Python Machine Learning, 2 ed., Packt Publishing Ltd.

Rider, M., 2002, The Geological Interpretation of Well Logs. Second edition. [S.I.]: Rider French Consulting Ltd.

## Geophysics

29

Rushing, J.A., Newsham, K.E. and T.A. Blasingame, 2008, Rock typing - Keys to understanding productivity in tight-gas sands: In SPE Unconventional Reservoirs Conference, Society of Petroleum Engineers.

Sammut, C., and G. I. Webb, 2011, Encyclopedia of Machine Learning: Springer Science & Business Media.

Sebtosheikh, M. A., and A. Salehi, 2015, Lithology prediction by support vector classifiers using inverted seismic attributes data and petrophysical logs as a new approach and investigation of training data set size effect on its performance in a heterogeneous carbonate reservoir: Journal of Petroleum Science and Engineering, **134**(1), 143 – 149.

Serra, O. 1984. Fundamentals of well-log interpretation: The acquisition of logging data. Developments in Petroleum Science, Elsevier.

Serra, O. and H.T. Abbott, 1982, The Contribution of Logging Data to Sedimentology and Stratigraphy. Society of Petroleum Engineers.

Silva, A. A., Lima Neto, I.A., Missagia, R., Ceia, M.A., Carrasquilla, A. and Archilha, N., 2015, Artificial neural networks to support petrographic classification of carbonate-siliciclastic rocks using well logs and textural information: Journal of Applied Geophysics, **117** (1), 118–125.

Tavares, M. W., 2018, Litofacies, electrofacies and petrofacies in the characterization of an Albian carbonate reservoir using data mining and artificial intelligence: P.h.D. thesis, UENF/LENEP, Macaé – RJ, Brazil. (In Portuguese).

1  
2  
3 Winland, H., 1972, Oil accumulation in response to pore size changes, Weyburn field,  
4 Saskatchewan: Amoco Production Research Report, F72-G25.  
5  
6

7  
8  
9 Xie, Y., Zhu, C., Zhou, W., Li, Z., Liu, X., and M. Tu, 2017, Evaluation of machine  
10 learning methods for formation lithology identification: A comparison of tuning processes and  
11 model performances: *Journal of Petroleum Science and Engineering*, **139** (27), 182–193.  
12  
13

14  
15 Yu, L., Porwal, A., Holden, E.J., and M.C. Dentith, 2012, Towards automatic  
16 lithological classification from remote sensing data using support vector machines *Computer &*  
17 *Geosciences*, **45**, 229-239.  
18  
19

20  
21 Yu, F., Lu, Z., Luo, H., and P. Wang, 2011, Three-dimensional model analysis and  
22 processing: Springer Science & Business Media.  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

## LIST OF FIGURES

Figure 1 - Location of Albian carbonate reservoirs in the Campos Basin, offshore Brazil. Albian carbonates reservoirs are indicated by the dashed red line (modified from Bruhn et al., 2003).

Figure 2 - The generalized location of the Wells A3 and A10 on the carbonate ramp depositional model with representative "depositional energy" designations based on characteristic lithofacies. The photomicrographs from plug samples of each lithofacies and their associated energy zone: oolitic and oncoidal grainstones of the high-energy zone (red frames), oncoidal peloidal grainstones and oncoidal bioclastic packstones of the moderate-energy zone (orange frames) and, peloidal bioclastic packstones and wackestones of the low-energy zone (blue frames) (modified from Okubo, 2014).

Figure 3 - Porosity (%) and permeability (mD) measured in the laboratory, mercury injection capillary pressure tests and the curves of the study model of Tavares (2018) for wells A3 and A10. Depositional classification according to ambient energy of PchG samples: Pink triangle = high energy, black square = moderate energy, and purple circle = low energy.

Figure 4 - Experimental porosity (%), permeability (mD), and curves of mercury injection capillary pressure (throat radii values of 0.6 microns and 3.2 microns) showing: (a) the textural facies, (b) petrofacies classification for A3 and A10 wells. Depositional classification, according to PCHG samples: Pink triangle = high energy, black square = moderate energy, and purple circle = low energy (Tavares, 2018).

Figure 5 - Importance of different petrophysical log-based predictors for both Well 3 and Well 10. Y-axis refers to the score of importance. GR, DRDN and RHOB are ranked higher than others.

Figure 6 - Scatter matrix plot of selected well logs, GR (API), RHOB (g/cc) and DRDN, color-coded according to petrofacies. The plot reveals overlapping pattern of petrofacies clusters.

Figure 7 - Radial coordinate visualization (RadViz) of P1, P2 and P3 in relation to (a) all 11 well-log measurements and to (b) GR, DRN and RHOB measurements. Each feature placed around the circle is attached to one data point through imaginary springs. The data point is plotted where the sum of all features' spring forces equals zero according to Hooke's law.

Figure 8 - (a) Class distribution before SMOTE; (b) Class distribution after SMOTE; (c) P1 and P2 samples before SMOTE; (d) P1 and P2 samples after SMOTE.

Figure 9 - Boxplots of ten-fold cross-validation accuracies using the default parameters (a) and the optimized parameters (b). The orange lines represent the median accuracy achieved, the width is related to range of accuracies and the open circles represent the supposed outliers.

Figure 10 – Influence of training dataset size. Learning curves for (a) DT, (b) RF, (c) GB, (d) KNN and (e) NB. The orange lines show the training accuracy and the green solid lines show the cross-validation scores which evaluates the model on the validation set. The shaded areas represent the standard deviation of the accuracies after running the model multiple times with the same number of training data.

Figure 11 - Well logs and petrofacies classification results from Well A3.

Figure 12 – Well logs and petrofacies classification results from Well A10.

Figure 13 – Normalized confusion matrix of (a) DT, (b) RF, (c) GB, (d) KNN and (e) NB. The rows represent the actual petrofacies, whereas the columns represent the predicted petrofacies. The values of the main diagonal represent the proportion of correctly predicted petrofacies. Mistakenly are represented by the off-diagonal elements.

Figure 14 – Computation time of machine learning methods.

8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
LIST OF TABLES

Table 1 - Multilinear regression coefficients of Winland (1972), Pittman (1992) and Tavares (2018).

Table 2 - Petrofacies established for Well A3 and Well A10 (Tavares, 2018).

Table 3 - Well logs measurements used in this study. DRDN normalizes the RHOB (g/cc) and PHIN (%) logs, highlighting sandstone or carbonate from other reservoir lithologies.

Table 4 - Total of samples for each petrofacies.

Table 5 - Total of samples of each petrofacies for training/validation and blind dataset.

Table 6 - Hyper-parameters optimized for each machine learning technique.

Table 7 - Performance metrics for the test dataset.

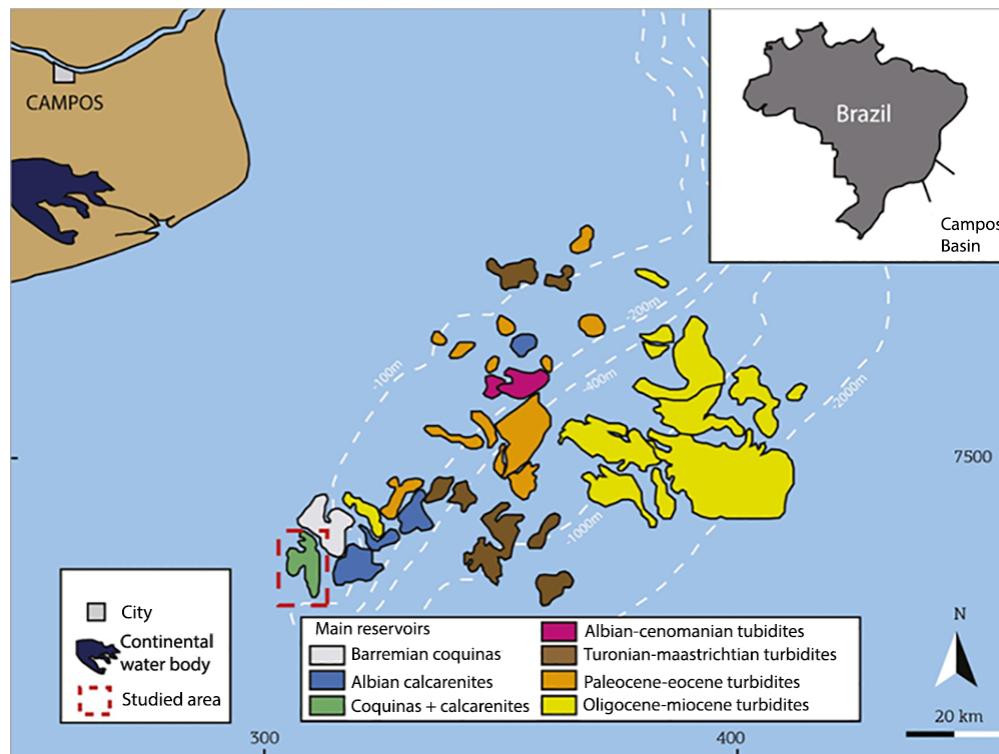


Figure 1: Location of Albian carbonate reservoirs in the Campos Basin, offshore Brazil. Albian carbonates reservoirs are indicated by the dashed red line (modified from Bruhn et al., 2003).

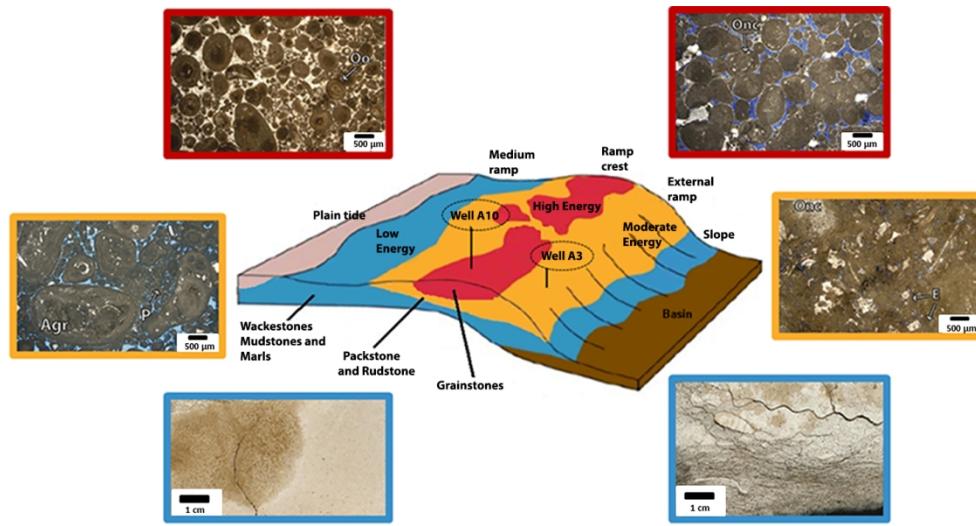


Figure 2: The generalized location of the Wells A3 and A10 on the carbonate ramp depositional model with representative "depositional energy" designations based on characteristic lithofacies. The photomicrographs from plug samples of each lithofacies and their associated energy zone: oolitic and oncocalcareous grainstones of the high-energy zone (red frames), oncocalcareous peloidal grainstones and oncocalcareous bioclastic packstones of the moderate-energy zone (orange frames) and, peloidal bioclastic packstones and wackestones of the low-energy zone (blue frames) (modified from Okubo, 2014).

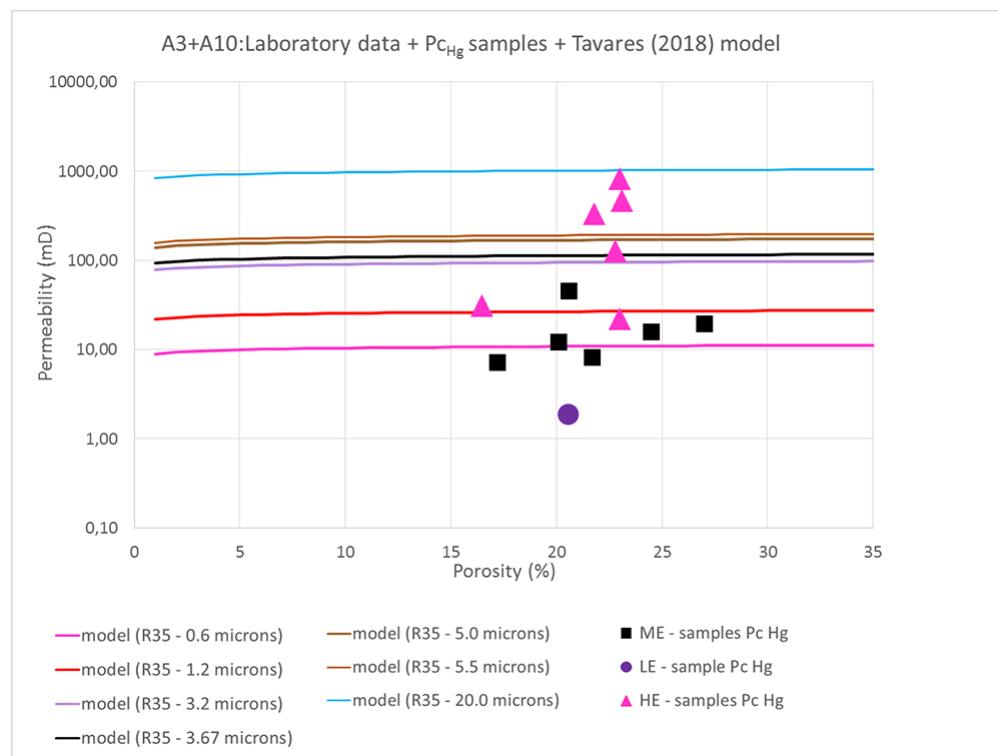


Figure 3: Porosity (%) and permeability (mD) measured in the laboratory, mercury injection capillary pressure tests and the curves of the study model of Tavares (2018) for wells A3 and A10. Depositional classification according to ambient energy of  $P_{CHg}$  samples: Pink triangle = high energy, black square = moderate energy, and purple circle = low energy.

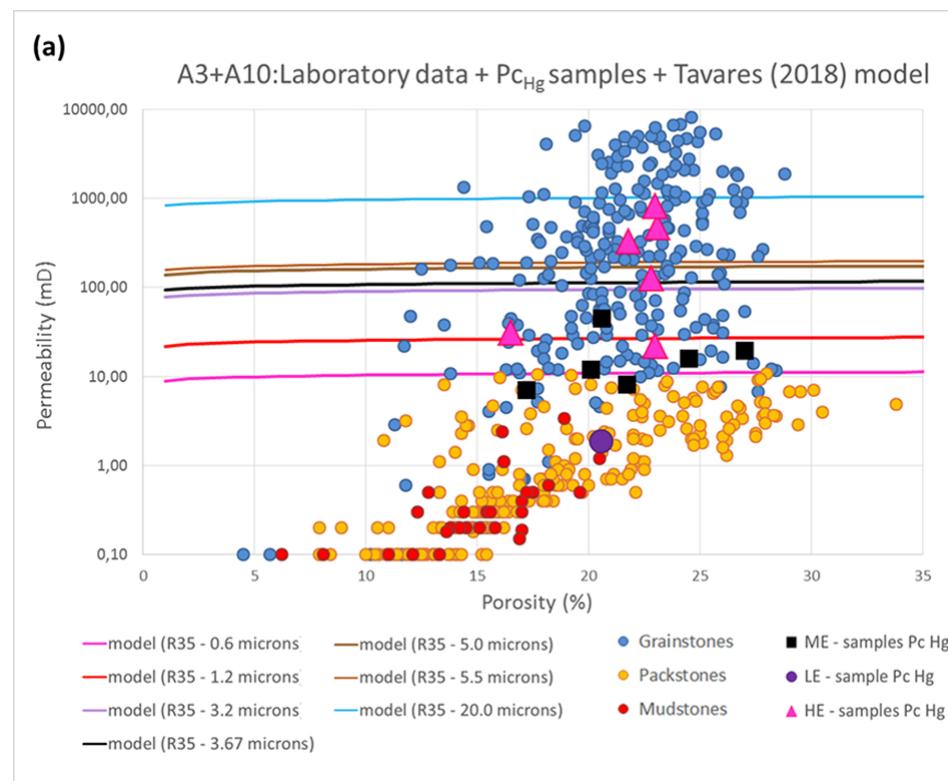


Figure 4(a): Experimental porosity (%), permeability (mD), and curves of mercury injection capillary pressure (throat radii values of 0.6 microns and 3.2 microns) showing: (a) the textural facies, (b) petrofacies classification for A3 and A10 wells. Depositional classification, according to PCHG samples: Pink triangle = high energy, black square = moderate energy, and purple circle = low energy (Tavares, 2018).

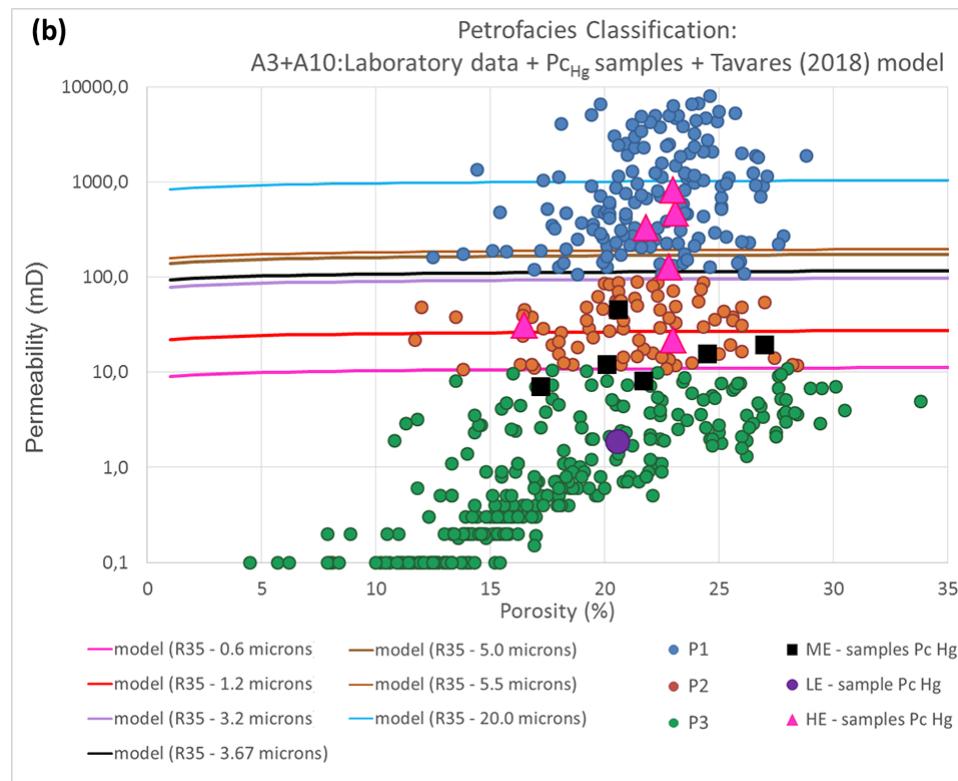


Figure 4(b): Experimental porosity (%), permeability (mD), and curves of mercury injection capillary pressure (throat radii values of 0.6 microns and 3.2 microns) showing: (a) the textural facies, (b) petrofacies classification for A3 and A10 wells. Depositional classification, according to PCHG samples: Pink triangle = high energy, black square = moderate energy, and purple circle = low energy (Tavares, 2018).

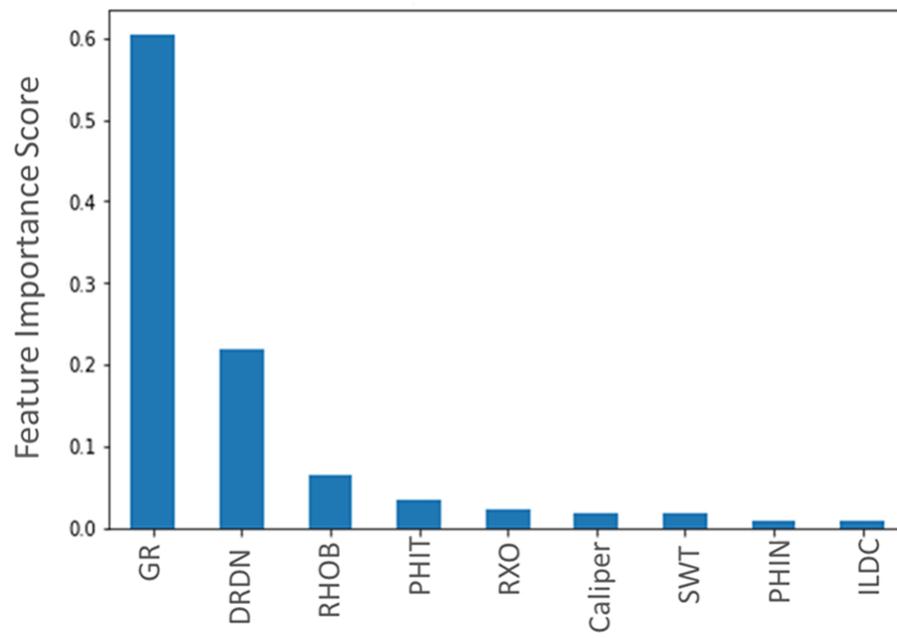


Figure 5: Importance of different petrophysical log-based predictors for both Well 3 and Well 10. Y-axis refers to the score of importance. GR, DRDN and RHOB are ranked higher than others.

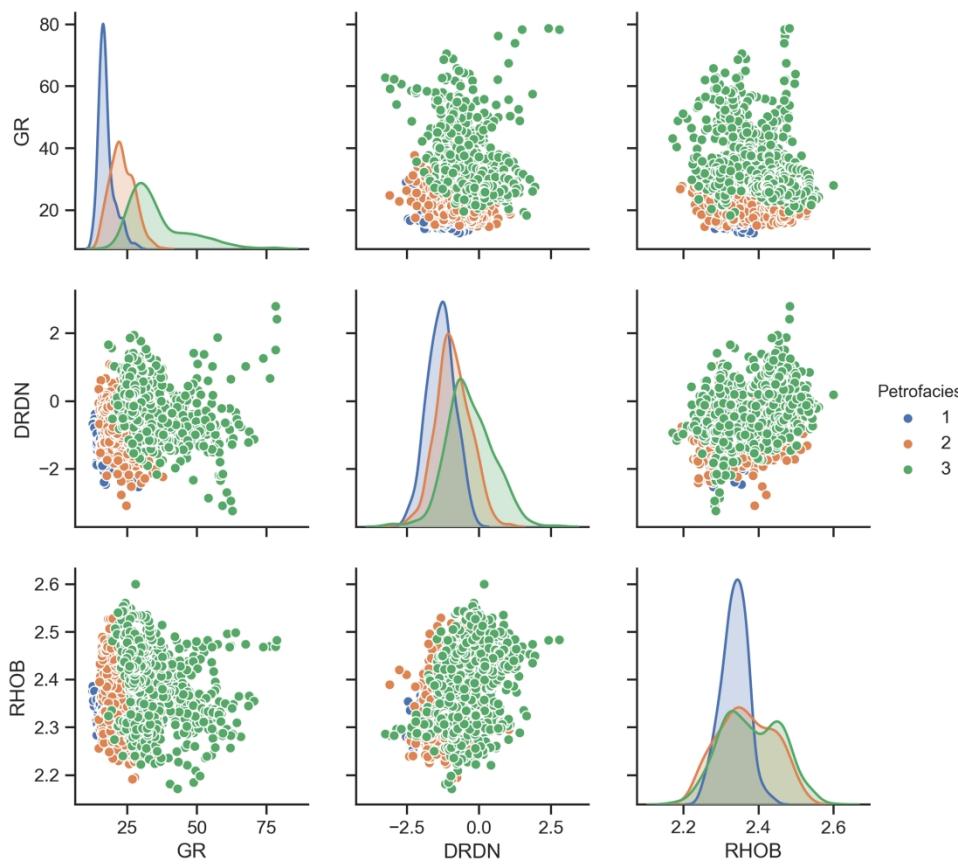


Figure 6: Scatter matrix plot of selected well logs, GR (API), RHOB (g/cc) and DRDN, color-coded according to petrofacies. The plot reveals overlapping pattern of petrofacies clusters.

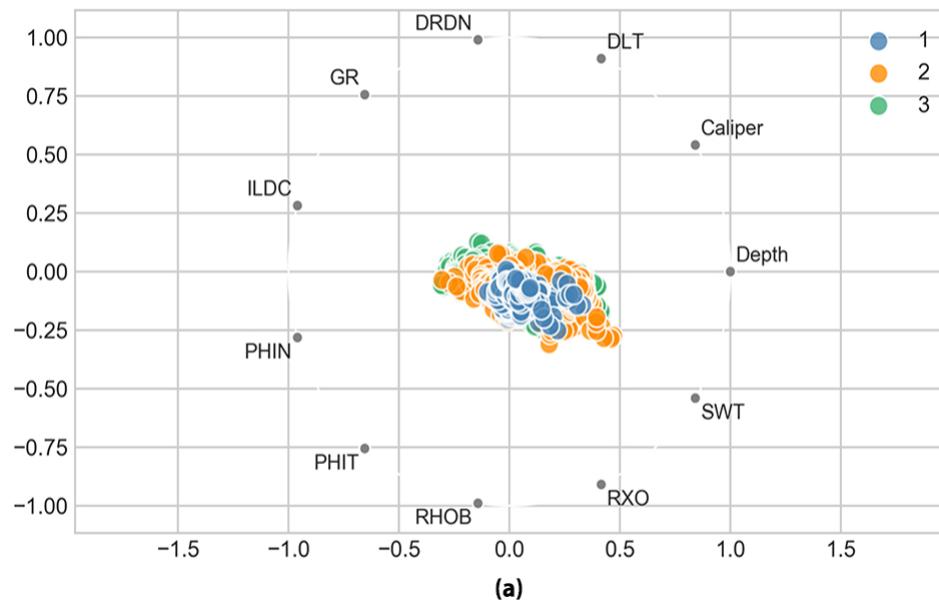


Figure 7(a): Radial coordinate visualization (RadViz) of P1, P2 and P3 in relation to (a) all 11 well-log measurements and to (b) GR, DRN and RHOB measurements. Each feature placed around the circle is attached to one data point through imaginary springs. The data point is plotted where the sum of all features' spring forces equals zero according to Hooke's law.

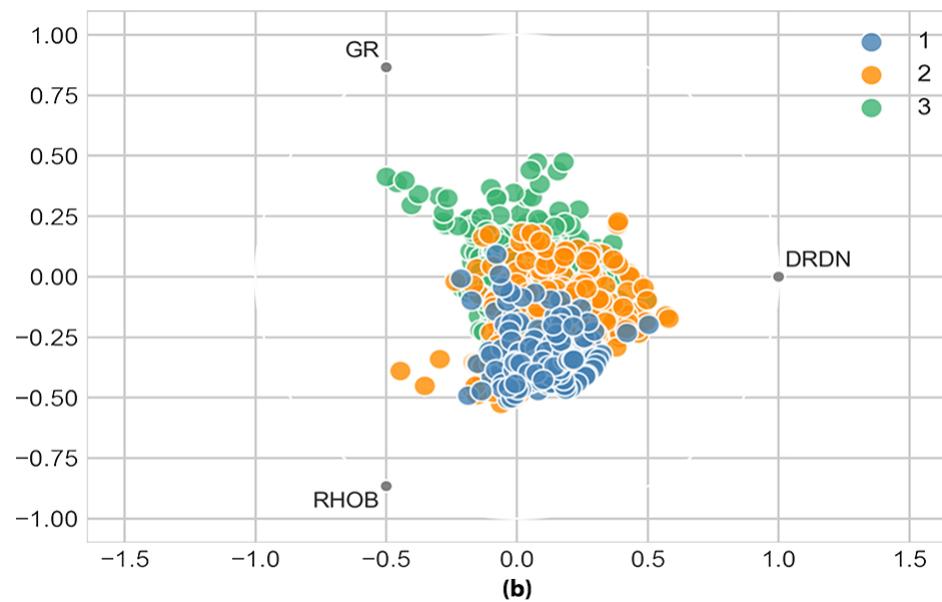


Figure 7(b): Radial coordinate visualization (RadViz) of P1, P2 and P3 in relation to (a) all 11 well-log measurements and to (b) GR, DRN and RHOB measurements. Each feature placed around the circle is attached to one data point through imaginary springs. The data point is plotted where the sum of all features' spring forces equals zero according to Hooke's law.

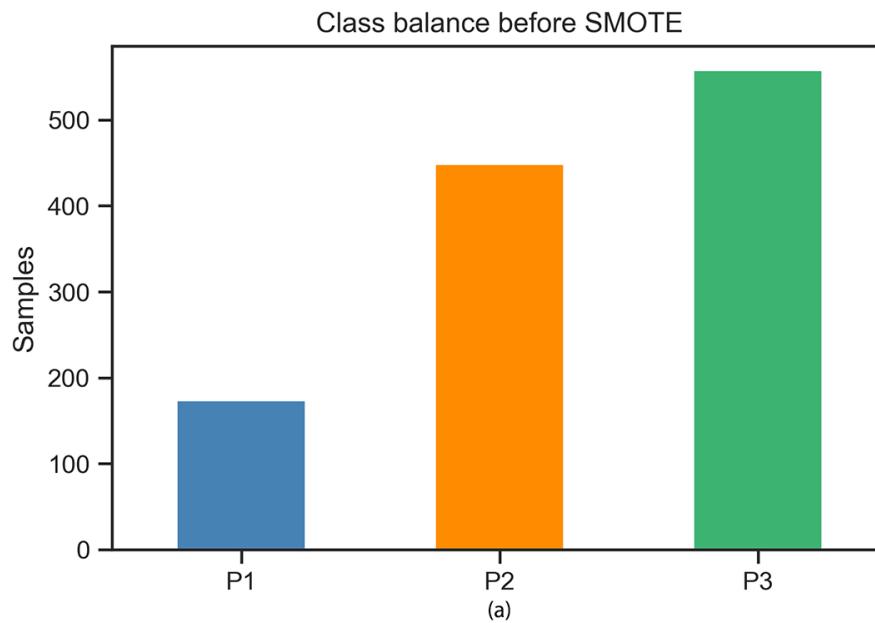


Figure 8(a): (a) Class distribution before SMOTE; (b) Class distribution after SMOTE; (c) P1 and P2 samples before SMOTE; (d) P1 and P2 samples after SMOTE.

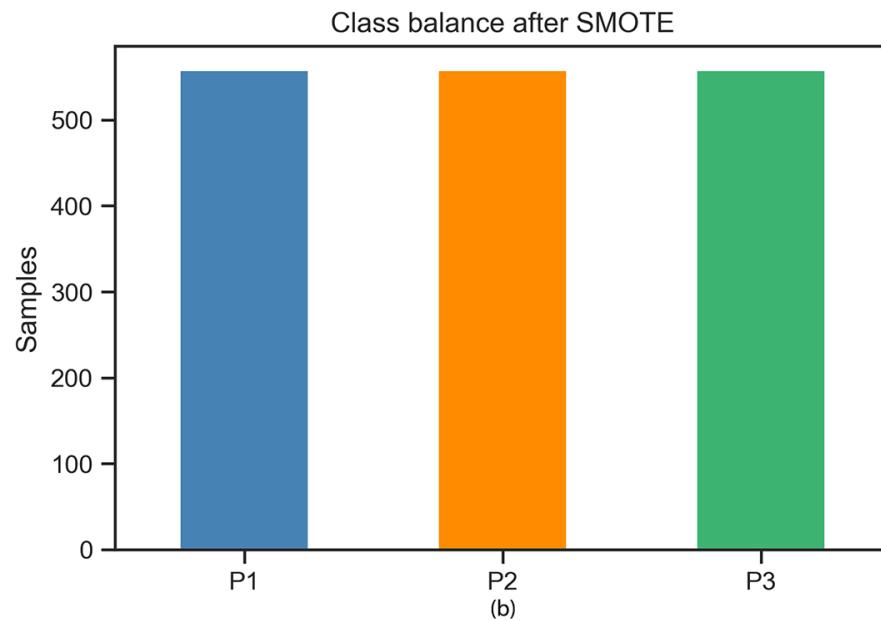


Figure 8(b): (a) Class distribution before SMOTE; (b) Class distribution after SMOTE; (c) P1 and P2 samples before SMOTE; (d) P1 and P2 samples after SMOTE.

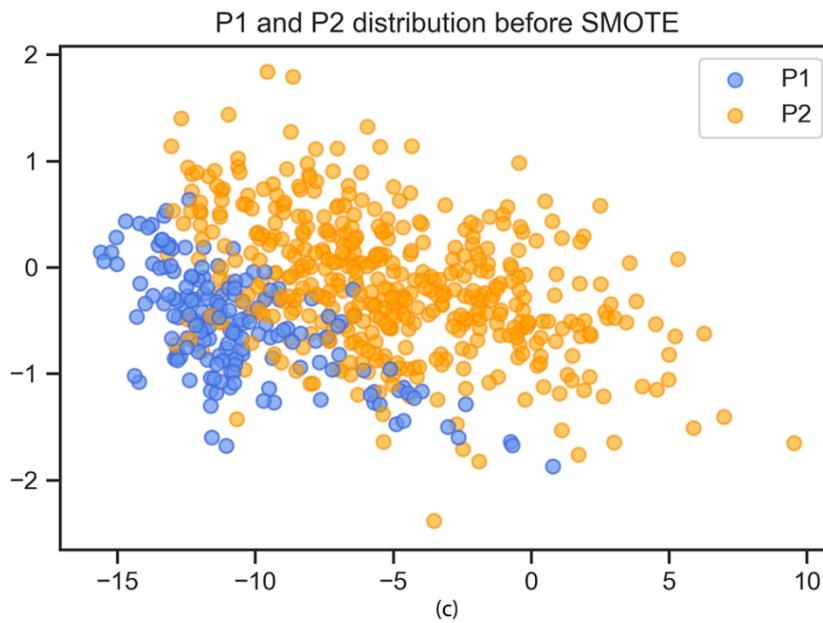


Figure 8(c): (a) Class distribution before SMOTE; (b) Class distribution after SMOTE; (c) P1 and P2 samples before SMOTE; (d) P1 and P2 samples after SMOTE.

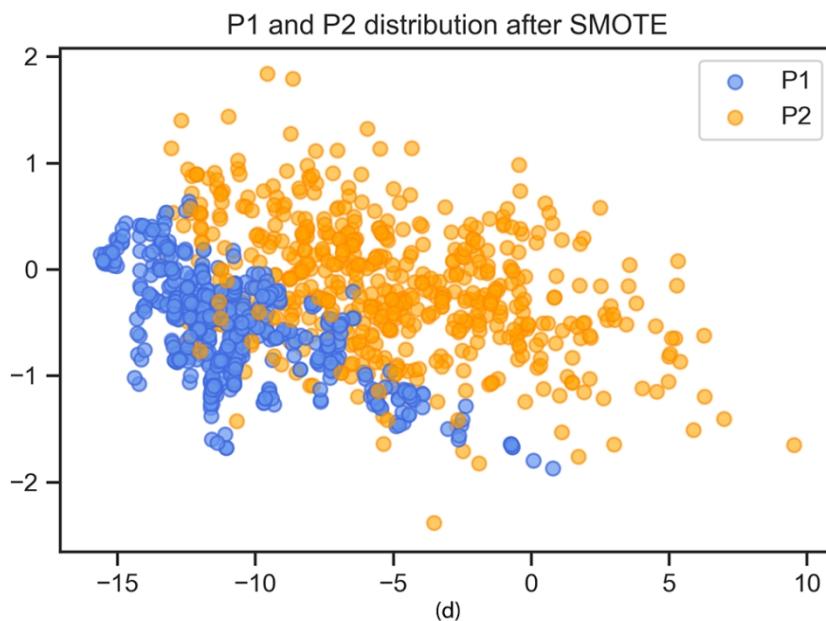


Figure 8(d): (a) Class distribution before SMOTE; (b) Class distribution after SMOTE; (c) P1 and P2 samples before SMOTE; (d) P1 and P2 samples after SMOTE.

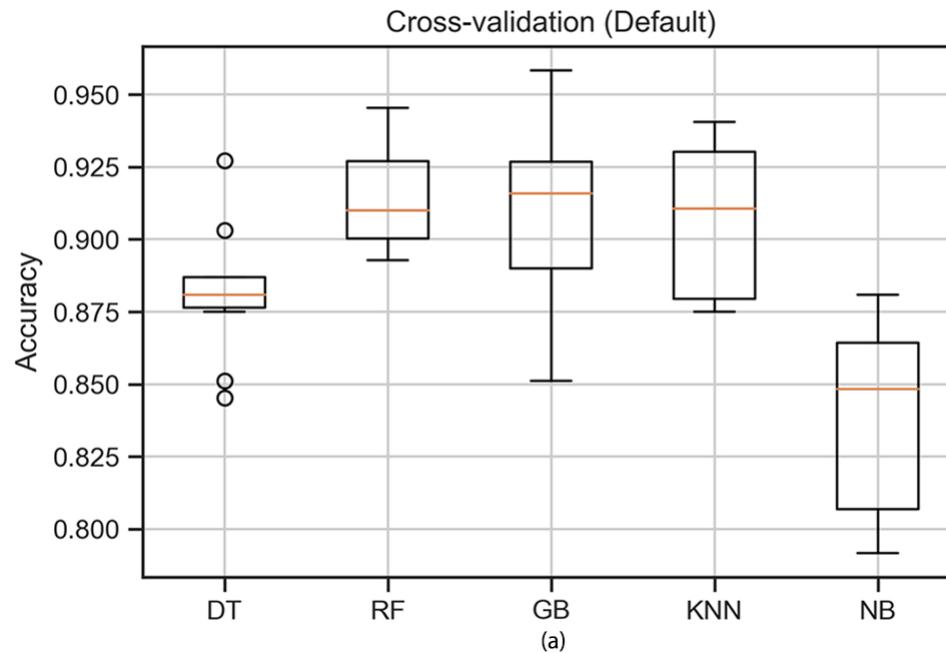


Figure 9(a): Boxplots of ten-fold cross-validation accuracies using the default parameters (a) and the optimized parameters (b). The orange lines represent the median accuracy achieved, the width is related to range of accuracies and the open circles represent the supposed outliers.

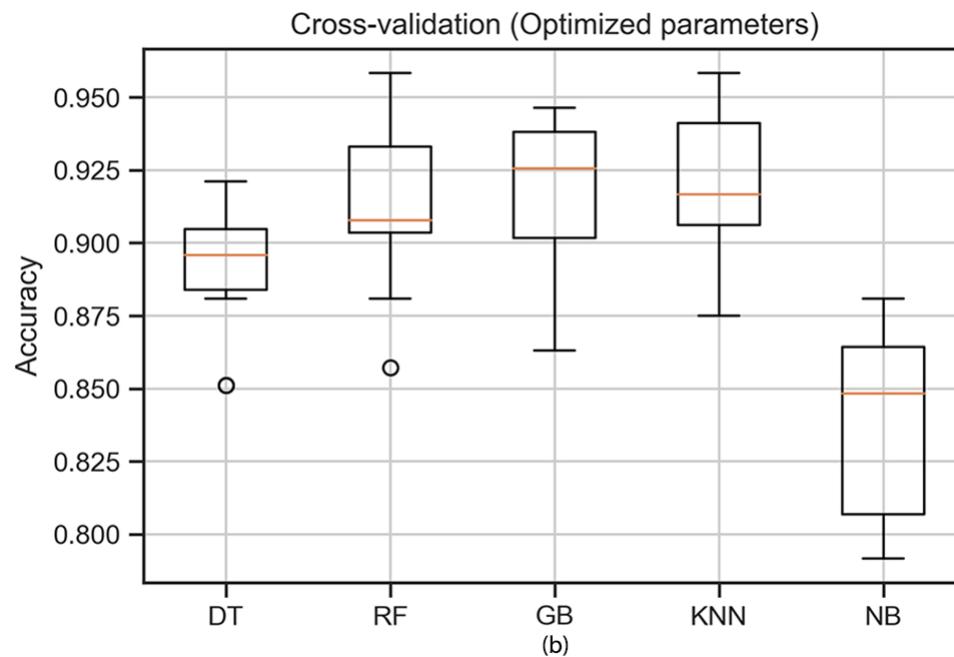


Figure 9(b): Boxplots of ten-fold cross-validation accuracies using the default parameters (a) and the optimized parameters (b). The orange lines represent the median accuracy achieved, the width is related to the range of accuracies and the open circles represent the supposed outliers.

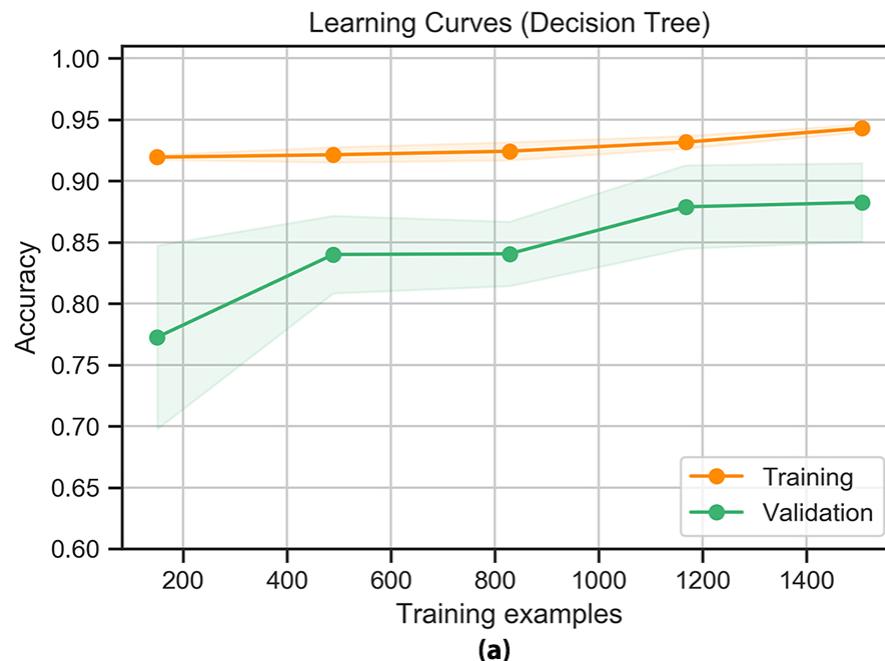


Figure 10(a): Influence of training dataset size. Learning curves for (a) DT, (b) RF, (c) GB, (d) KNN and (e) NB. The orange lines show the training accuracy and the green solid lines show the cross-validation scores which evaluates the model on the validation set. The shaded areas represent the standard deviation of the accuracies after running the model multiple times with the same number of training data.

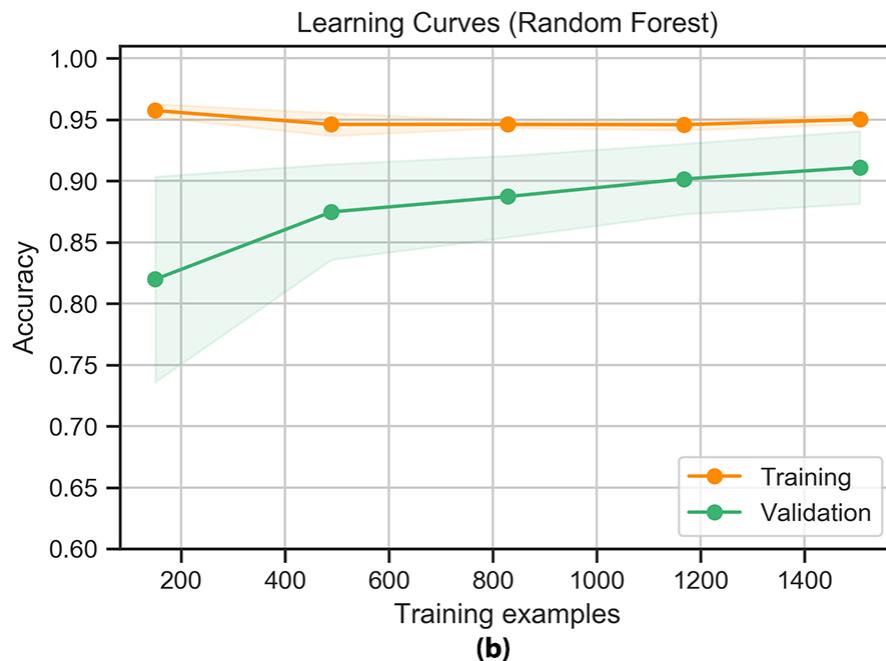


Figure 10(b): Influence of training dataset size. Learning curves for (a) DT, (b) RF, (c) GB, (d) KNN and (e) NB. The orange lines show the training accuracy and the green solid lines show the cross-validation scores which evaluates the model on the validation set. The shaded areas represent the standard deviation of the accuracies after running the model multiple times with the same number of training data.

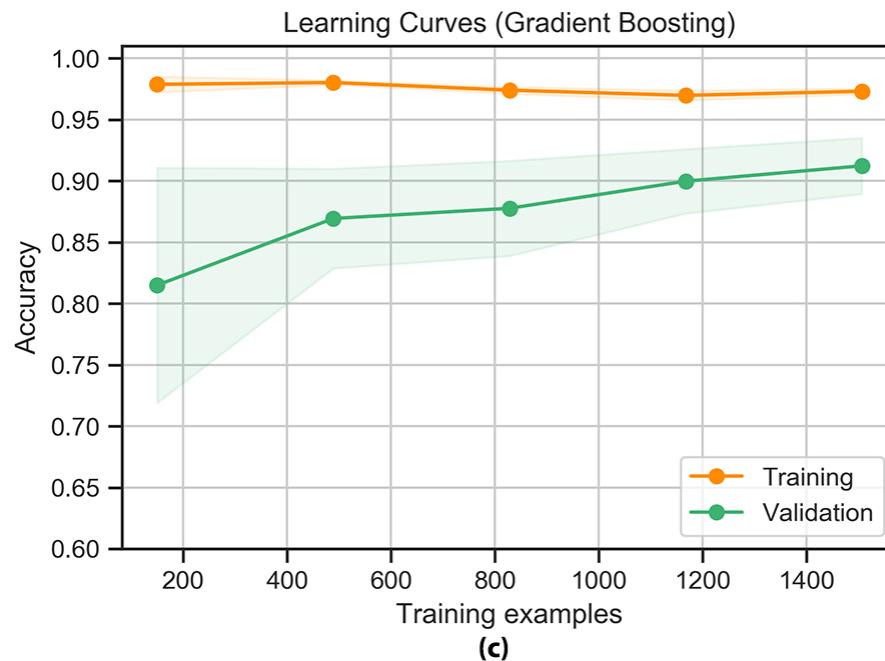


Figure 10(c): Influence of training dataset size. Learning curves for (a) DT, (b) RF, (c) GB, (d) KNN and (e) NB. The orange lines show the training accuracy and the green solid lines show the cross-validation scores which evaluates the model on the validation set. The shaded areas represent the standard deviation of the accuracies after running the model multiple times with the same number of training data.

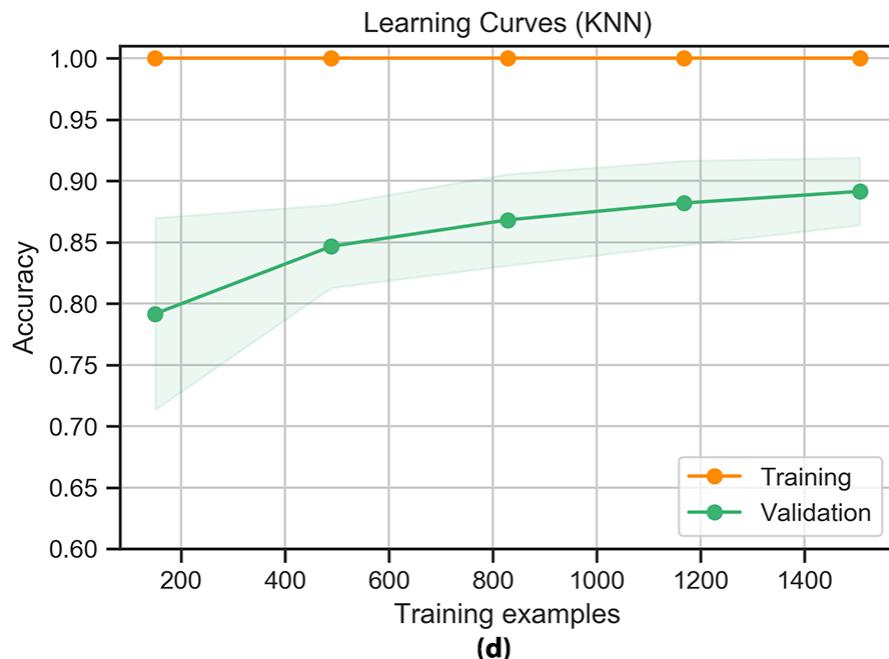


Figure 10(d): Influence of training dataset size. Learning curves for (a) DT, (b) RF, (c) GB, (d) KNN and (e) NB. The orange lines show the training accuracy and the green solid lines show the cross-validation scores which evaluates the model on the validation set. The shaded areas represent the standard deviation of the accuracies after running the model multiple times with the same number of training data.

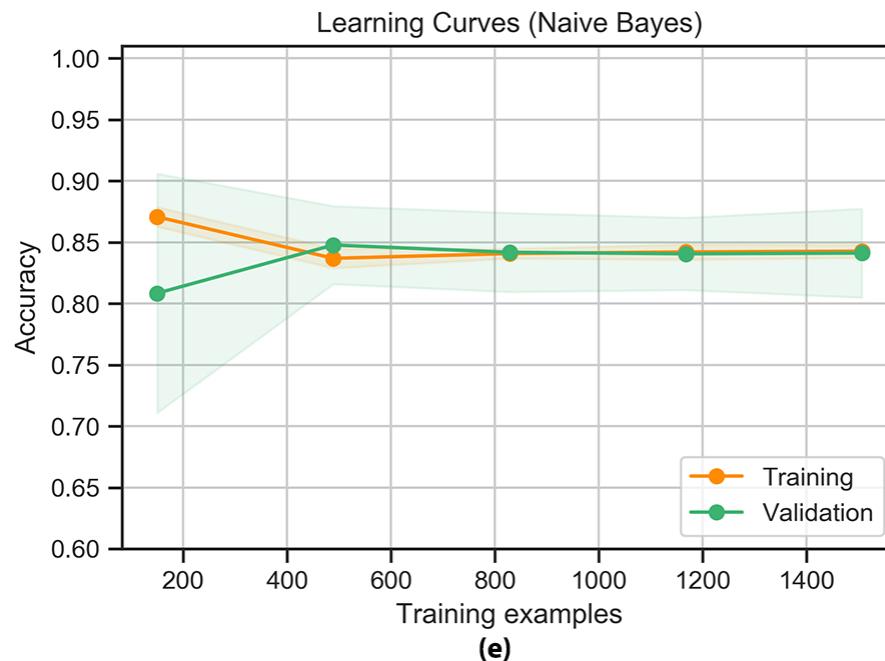


Figure 10(e): Influence of training dataset size. Learning curves for (a) DT, (b) RF, (c) GB, (d) KNN and (e) NB. The orange lines show the training accuracy and the green solid lines show the cross-validation scores which evaluates the model on the validation set. The shaded areas represent the standard deviation of the accuracies after running the model multiple times with the same number of training data.

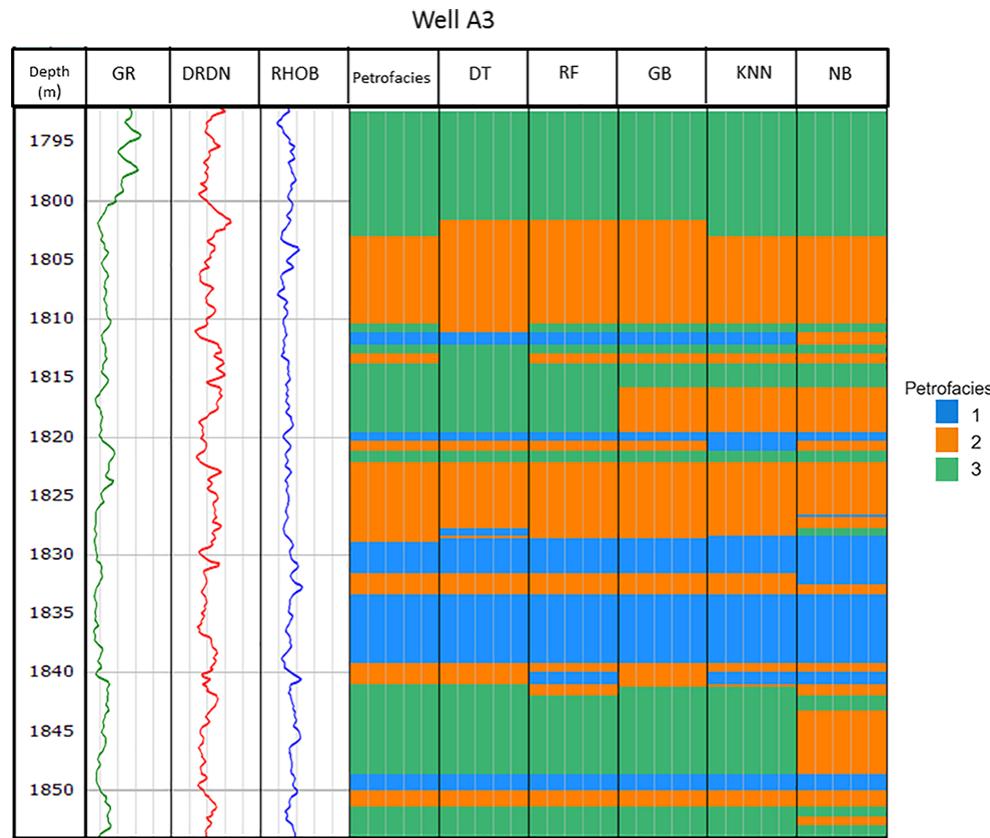


Figure 11: Well logs and petrofacies classification results from Well A3.

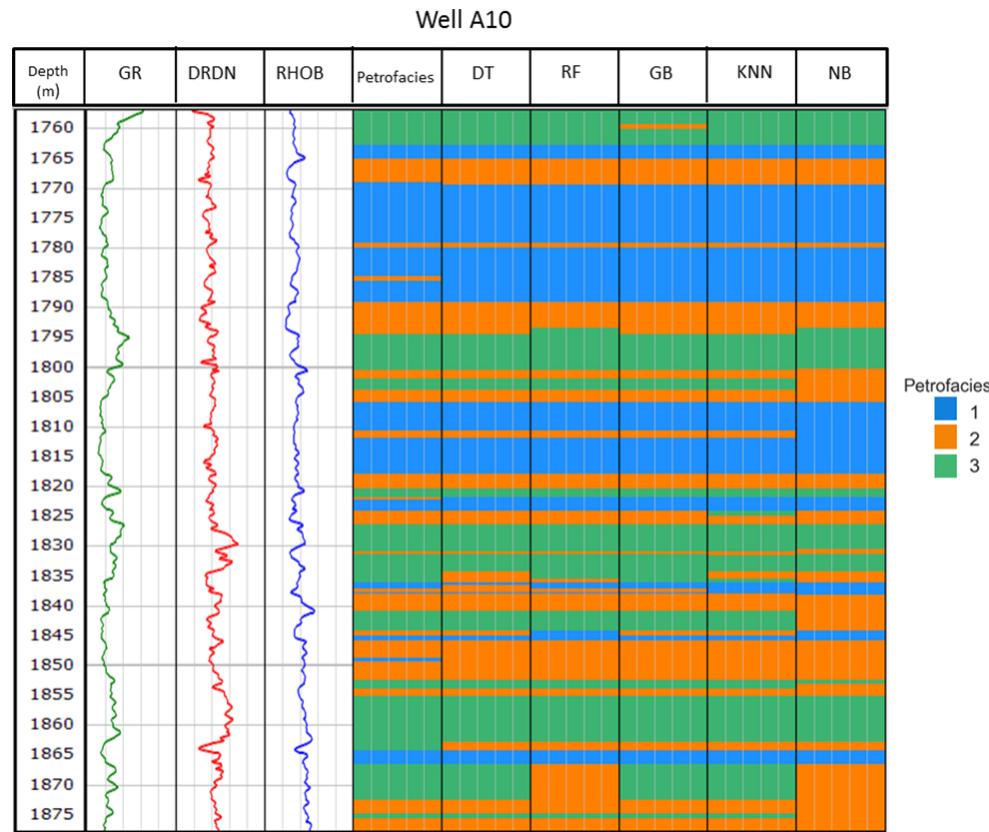


Figure 12: Well logs and petrofacies classification results from Well A10.

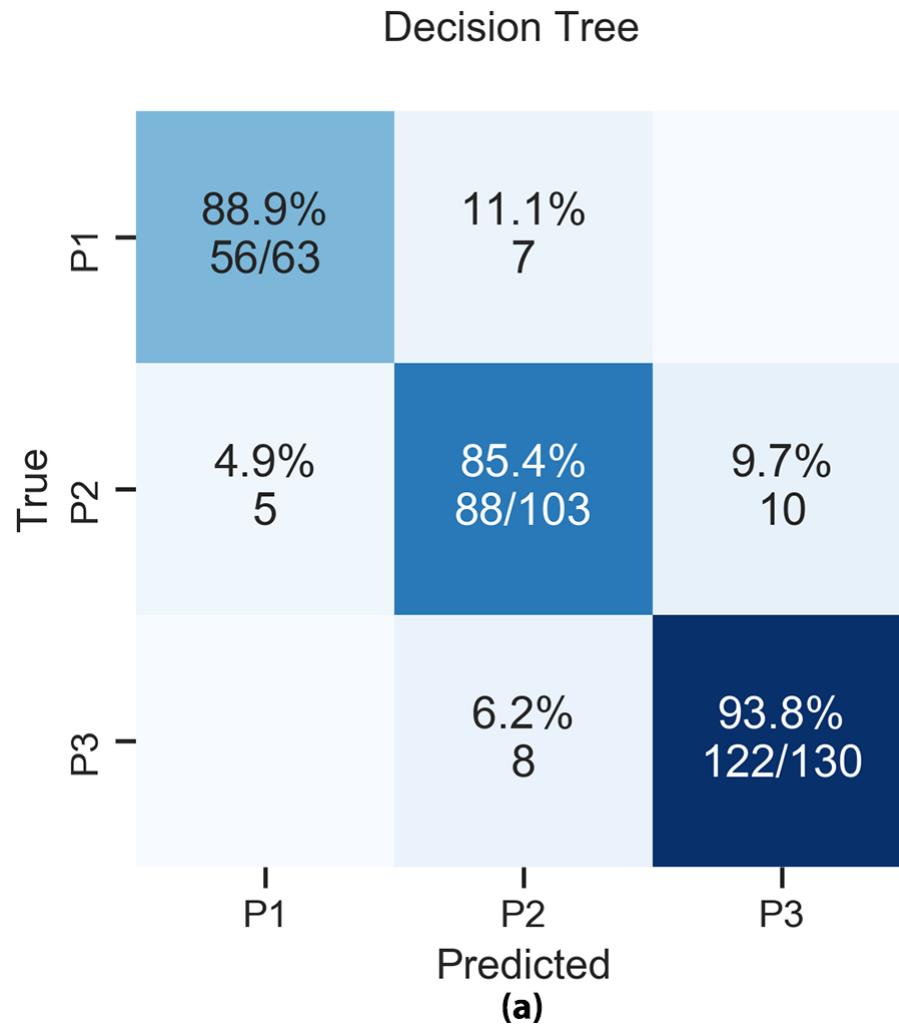


Figure 13(a): Normalized confusion matrix of (a) DT, (b) RF, (c) GB, (d) KNN and (e) NB. The rows represent the actual petrofacies, whereas the columns represent the predicted petrofacies. The values of the main diagonal represent the proportion of correctly predicted petrofacies. Mistakenly are represented by the off-diagonal elements.

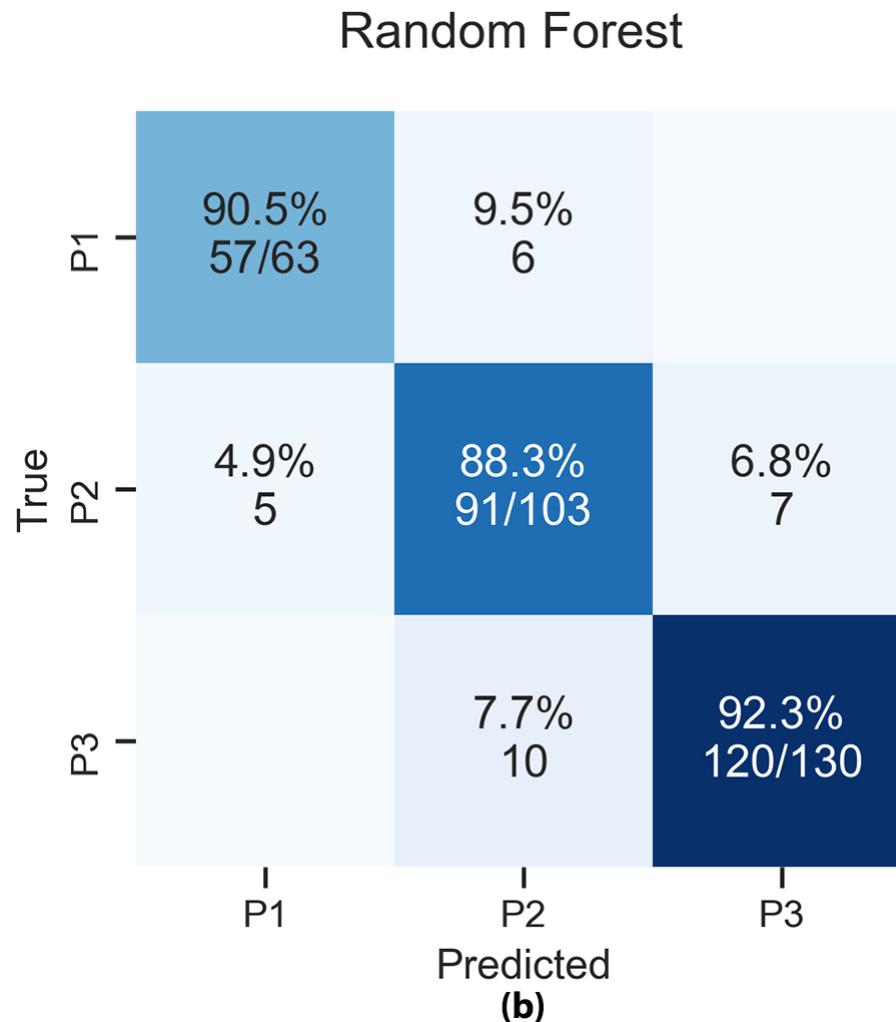


Figure 13(b): Normalized confusion matrix of (a) DT, (b) RF, (c) GB, (d) KNN and (e) NB. The rows represent the actual petrofacies, whereas the columns represent the predicted petrofacies. The values of the main diagonal represent the proportion of correctly predicted petrofacies. Mistakenly are represented by the off-diagonal elements.

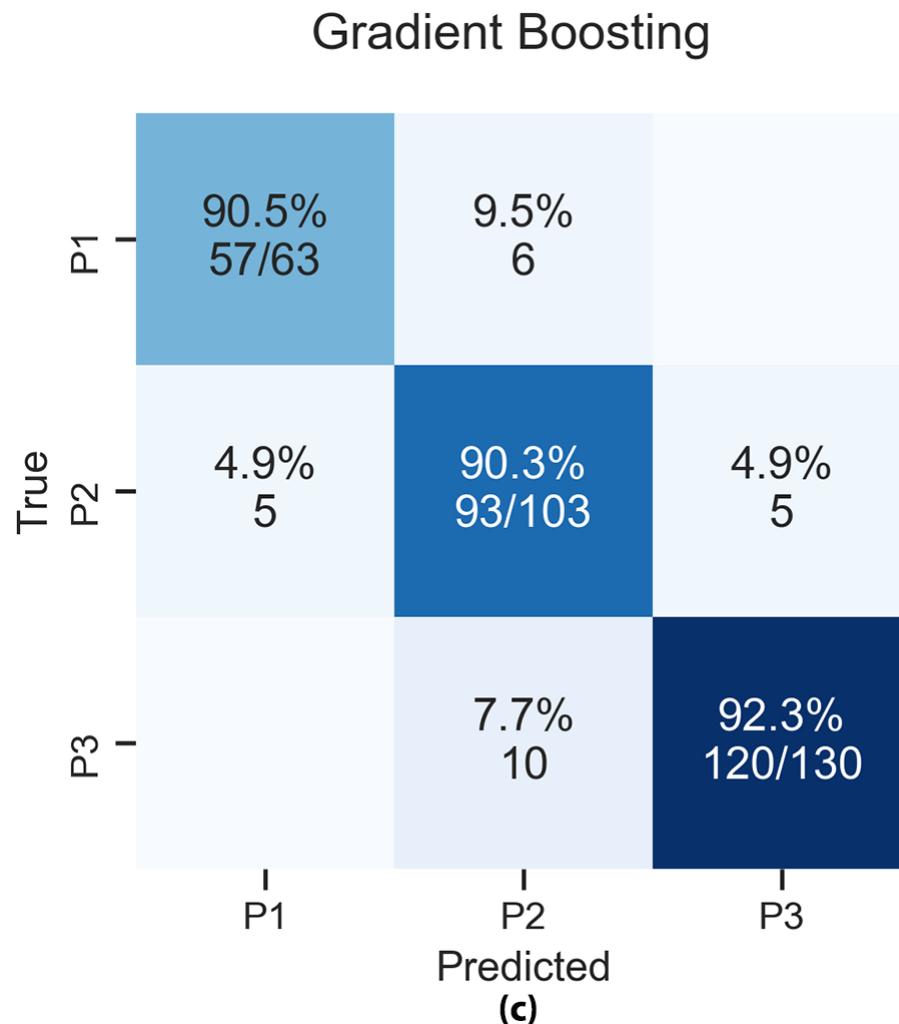


Figure 13(c): Normalized confusion matrix of (a) DT, (b) RF, (c) GB, (d) KNN and (e) NB. The rows represent the actual petrofacies, whereas the columns represent the predicted petrofacies. The values of the main diagonal represent the proportion of correctly predicted petrofacies. Mistakenly are represented by the off-diagonal elements.

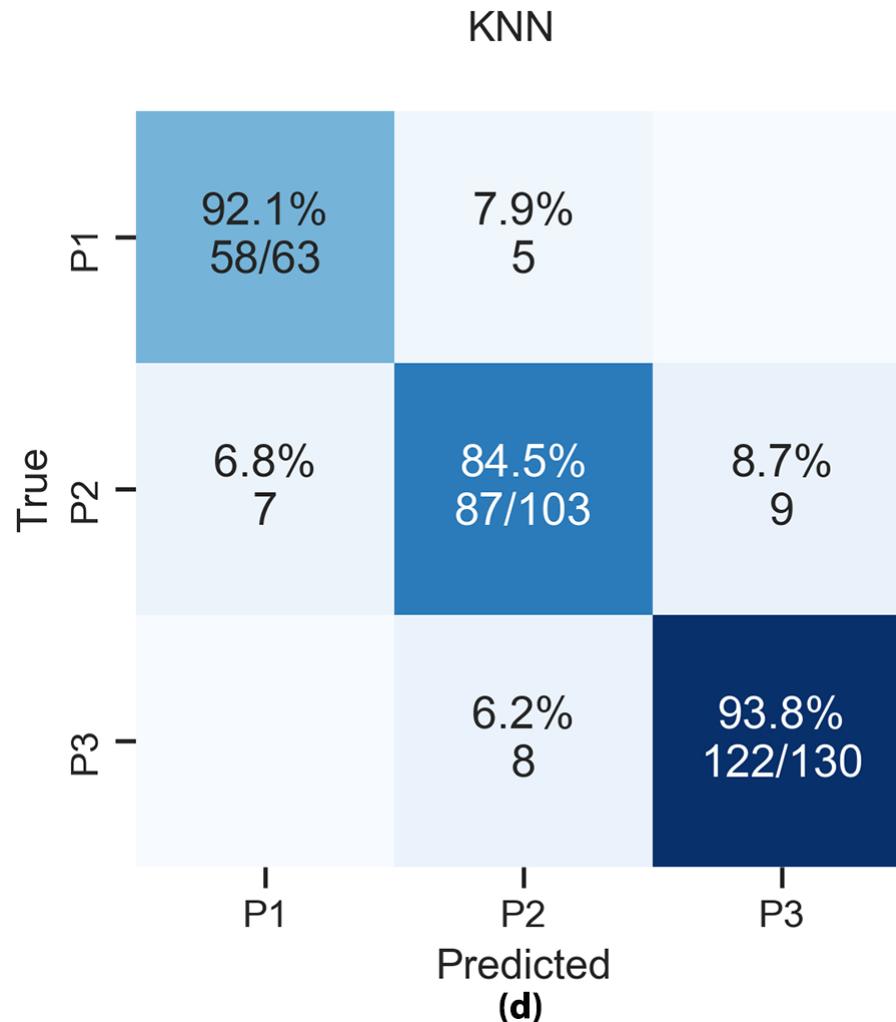


Figure 13(d): Normalized confusion matrix of (a) DT, (b) RF, (c) GB, (d) KNN and (e) NB. The rows represent the actual petrofacies, whereas the columns represent the predicted petrofacies. The values of the main diagonal represent the proportion of correctly predicted petrofacies. Mistakenly are represented by the off-diagonal elements.

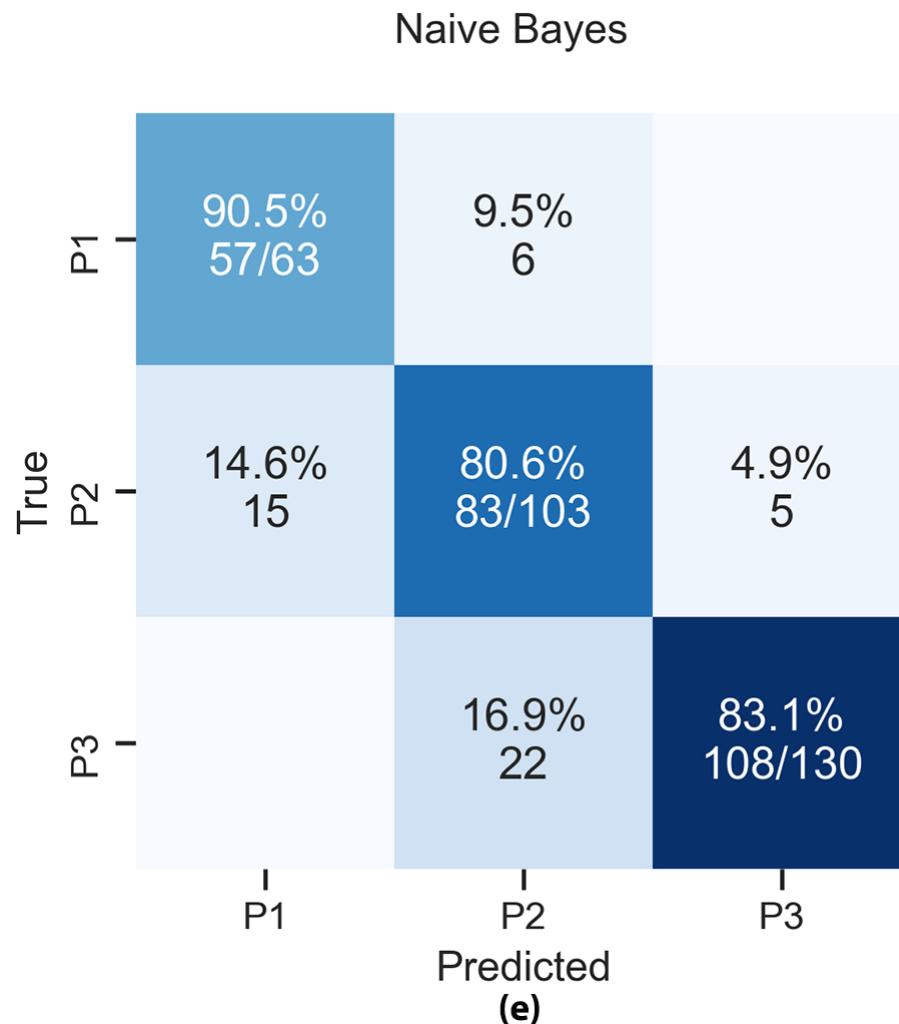


Figure 13(e): Normalized confusion matrix of (a) DT, (b) RF, (c) GB, (d) KNN and (e) NB. The rows represent the actual petrofacies, whereas the columns represent the predicted petrofacies. The values of the main diagonal represent the proportion of correctly predicted petrofacies. Mistakenly are represented by the off-diagonal elements.

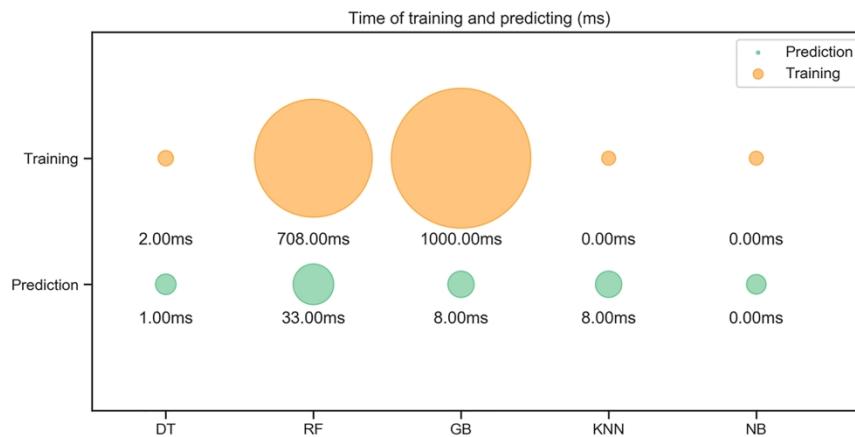


Figure 14: Computation time of machine learning methods.

1  
2  
3 **Table 1 - Multilinear regression coefficients of Winland (1972), Pittman (1992) and Tavares (2018).**

| 5<br>6<br><b>Coefficients</b>  | 7<br><b>Winland (1972)</b> | 8<br><b>Pittman (1992)</b> | 9<br><b>Tavares (2018)</b> |
|--------------------------------|----------------------------|----------------------------|----------------------------|
| 10<br>Permeability (c1)        | 11<br>0.59                 | 12<br>0.57                 | 13<br>0.56                 |
| 14<br>Porosity (c2)            | 15<br>-0.86                | 16<br>-0.52                | 17<br>0.73                 |
| 18<br>Y-axis intersection (c3) | 19<br>0.73                 | 20<br>0.26                 | 21<br>-1.68                |

Table 2 - Petrofacies established for Well A3 and A10 (Tavares, 2018).

|                      | Class   | Permeability                     | Texture/Lithology  |
|----------------------|---------|----------------------------------|--|
| <i>Petrofacies 1</i> | Good    | > 100mD                          | Grainstones and porous packstones  |
| <i>Petrofacies 2</i> | Regular | $10\text{mD} < k < 100\text{mD}$ | Grainstones and oncolytic, peloidal and microlytic/microoolytocous packstones. |
| <i>Petrofacies 3</i> | Bad     | < 10mD                           | Very fine packstones, mudstones, wackestones                                   |

Table 3 - Well logs measurements used in this study. DRDN normalizes the RHOB (g/cc) and PHIN (%) logs, highlighting sandstone or carbonate from other reservoir lithologies.

| Parameter | Nomenclature                             | Unit       |
|-----------|--|------------|
| GR        | Gamma Ray log                            | API        |
| RHOB      | Density log                              | g/cc       |
| Caliper   | Caliper log                              | m          |
| DLT       | Sonic log                                | ms/ft      |
| PHIN      | Neutron Log                              | %          |
| SWT       | Total Water Saturation                   | fractional |
| RXO       | Resistivity of flushed zone              | Ohm*m      |
| DRDN      | Difference RHOB-PHIN                     | -          |
| PHIT      | total porosity derived from the RHOB log | %          |
| Depth     | Depth                                    | m          |
| ILDC      | Deep Resistivity Log                     | Ohm*m      |

1  
2  
3      **Table 4 - Total of samples for each petrofacies.**

|                         | <b>Samples</b> |
|-------------------------|----------------|
| <i>Petrofacies 1</i>    | 237            |
| <i>Petrofacies 2</i>    | 552            |
| <i>Petrofacies 3</i>    | 688            |
| <b>Total of samples</b> | <b>1477</b>    |

1  
2  
3 **Table 5 - Total of samples of each petrofacies for training/validation and blind dataset.**  
4  
5

|                      | <b>Training and Validation</b> | <b>Blind Test</b> |
|----------------------|--------------------------------|-------------------|
| <i>Petrofacies 1</i> | 174                            | 63                |
| <i>Petrofacies 2</i> | 449                            | 103               |
| <i>Petrofacies 3</i> | 558                            | 130               |

Table 6 - Hyper-parameters optimized for each machine learning technique.

| Algorithm                        | Parameters                  |   |
|----------------------------------|-----------------------------|---|
| <b>Decision Tree (DT)</b>        | <b>min_samples_leaf</b> = 3 | The minimum number of samples for a node to be considered a leaf. Controls the depth and complexity of decision tree. |
|                                  | <b>max_features</b> = 3     | Number of features to consider when computing the best node split.  |
|                                  | <b>criterion</b> = ‘gini’   | Function used to measure the quality of a split.  |
|                                  | <b>max_depth</b> = 8        | The maximum depth of the tree.  |
| <b>Random Forest (RF)</b>        | <b>n_estimators</b> = 300   | Number of decision trees in the ensemble.   |
|                                  | <b>min_samples_leaf</b> = 3 | The minimum number of samples for a node to be considered a leaf. Controls the depth and complexity of decision tree. |
|                                  | <b>max_features</b> = 2     | Number of features to consider when computing the best node split.  |
|                                  | <b>criterion</b> = ‘gini’   | Function used to measure the quality of a split.  |
| <b>Gradient Boosting (GB)</b>    | <b>learning_rate</b> = 0.01 | The maximum depth of the tree.  |
|                                  | <b>n_estimators</b> = 300   | Number of decision trees in the ensemble.   |
|                                  | <b>max_features</b> = 3     | Number of features to consider when computing the best node split.  |
| <b>K-Nearest Neighbors (KNN)</b> | <b>learning_rate</b> = 0.01 | Shrinks the contribution of each successive decision tree in the ensemble.  |
|                                  | <b>max_depth</b> = 7        | Maximum depth of the decisions trees. Controls the complexity of the decision trees.                                  |
| <b>K-Nearest Neighbors (KNN)</b> | <b>n_neighbors</b> = 10     | Number of neighbors to use.   |
|                                  | <b>weights</b> = ‘distance’ | Function to weight the neighbors’ votes.  |
| <b>Naïve Bayes (NB)</b>          | <b>No parameters.</b>       |   |

1  
2  
3 **Table 7 - Performance metrics for the test dataset.**  
4  
5

|                          | <b>Accuracy</b> | <b>Precision</b> | <b>Recall</b> | <b>F1-score</b> |
|--------------------------|-----------------|------------------|---------------|-----------------|
| <b>Decision Tree</b>     | 89.86%          | 91.80%           | 88.89%        | 90.32%          |
| <b>Random Forest</b>     | 90.54%          | 90.66%           | 90.54%        | 90.58%          |
| <b>Gradient Boosting</b> | 91.22%          | 91.09%           | 91.03%        | 91.02%          |
| <b>KNN</b>               | 90.20%          | 89.23%           | 92.06%        | 90.62%          |
| <b>Naïve Bayes</b>       | 83.78%          | 79.17%           | 90.48%        | 84.44%          |

## DATA AND MATERIALS AVAILABILITY

Data associated with this research are confidential and cannot be released.