



Petrophysical guided preconditioning applied to electrofacies classification by machine learning

Lucas Gabriel Silva de Aguiar¹, Maurício Laurindo de Matos¹, Jadson Muniz de Oliveira¹, Rodrigo da Silva Canário¹ and Fernando Sérgio de Moraes¹, ¹ Grupo de Inferência de Reservatório - GIR/LENEP

Copyright 2021, SBGf - Sociedade Brasileira de Geofísica

This paper was prepared for presentation during the 17th International Congress of the Brazilian Geophysical Society held in Rio de Janeiro, Brazil, 16-19 August 2021.

Contents of this paper were reviewed by the Technical Committee of the 17th International Congress of the Brazilian Geophysical Society and do not necessarily represent any position of the SBGf, its officers or members. Electronic reproduction or storage of any part of this paper for commercial purposes without the written consent of the Brazilian Geophysical Society is prohibited.

Abstract

Machine learning has been increasingly used to solve problems in the oil and gas industry as these methods help to automate tasks and optimize time. One application that has been standing out is electrofacies interpretation from well logs using classification algorithms. Recently, the FORCE 2020 Machine Learning Competition took place and aimed to classify rocks from 20 different wells from the North Sea. Our study shows that applying petrophysical interpretation concepts as a data preconditioning strategy helps to improve your model performance. We also highlight the importance of data imputation in zones that previously had no values and the process of creating attributes for better class separation. The results indicate a high performance, where an average accuracy of 80% was achieved and relative penalties (competition performance metrics) of -0.4792 and -0.5037 for the two test data. This result placed us in the second position in the final ranking.

Introduction

Serra & Abbot (1982) defined the term “electrofacies” as sediment characterization by its response to a set of well logs. The interpreter analyzes all the received data and then gives an opinion on the possible type of rock. As this conventional method can be repetitive and time-consuming, statistical principles of artificial intelligence have proven to be an alternative for the industry due to the high demand for work optimization and automation. The purpose of a machine learning application is to identify patterns in the data that can generalize well to new data (Russel et al., 2010).

Several works have been published with this theme. Some use algorithms with recursive partitioning techniques (tree methods) for classification (Chen & Guestrin, 2016; Zhang & Zhan, 2017); while others make the classification based on calculations of the Euclidean distance between samples (Hall, 2016; Puskarczyk, 2019). Among all these, the Extreme Gradient Boosting (XGBoost) algorithm has stood out in several researches and competitions with classification problems (Torlay et al., 2017; Zhang et al., 2018). This algorithm uses a method of boosted trees (ensemble technique), where it creates weaker models

(weak learners) and each model tries to correct the errors present in the previous one.

The lack of data in relevant regions can add uncertainty to your model. It is common to make an imputation on missing data through trend measures such as mean or median in signal studies. However, the imputation of a constant value to well logs would not effectively represent the heterogeneity of geological formations. Recent studies show the predictive power of machine learning algorithms for estimating these logs (Perez et al., 2003; Al-Mudhafar, 2020). The use of these techniques with the aid of proper processing can considerably help to reduce the uncertainty of missing data.

The oil and gas industry already uses attribute generation techniques to create more characteristic properties for a given problem. The creation of the clay volume log from others indicates the normalized estimate of this sediment along the well, for example. In recent years, studies show that mathematical transformations of data to increase features can help to improve the performance of machine learning models (Bestagini et al., 2017; Galli, 2020). These new attributes tend to better separate classes that are not originally separable.

Recently, Bormann et al. (2020) created the FORCE Machine Learning Competition to classify electrofacies with a dataset containing more than 100 wells in the North Sea region, offshore Norway, and there were more than 300 registered teams. The last open competition of the same type was organized by SEG with a much smaller database (Hall, 2016). After that, several researches have been developed using that data, which boosted the progress of studies in this area.

We, from the GIR team, finished the FORCE competition in second place and this study will show how efficient data imputation and attribute generation helped us to achieve high performance, in addition to other types of processing. All contest data and winners' code are available online*.

The study area is part of the Norwegian continental shelf, mainly in the North Sea region (Figure 1). Rifting phases during the Upper Jurassic/Lower Cretaceous, partially controlled by older structures, helped to form the structural scene of the area (Halland et al., 2014). It presents highly variable geology, containing predominantly shales and shaly variations. Reservoirs are characterized by sandstones from a marginal marine environment and deltaic systems, originating mainly during the Upper Jurassic (Halland et al., 2014).

* <https://github.com/bolgebrygg/Force-2020-Machine-Learning-competition>



Figure 1 – Location of the studied area. The dashed red polygon represents the Norwegian North Sea. Source: adapted from Google Earth.

Method

The competition made 98 wells available as *training data*, ten wells as test data for the formation of a partial ranking, here called *test data 1*, and then another ten wells for the final evaluation of the model, called *test data 2*. This work will present the classification on the *test data 1* and 2, using the *training data* to fit the model.

The data has 29 well logs available: well name (WELL), resistivity logs (RDEP, RSHA, RMED, RXO and RMIC), spontaneous potential (SP), neutron porosity (NPHI), photoelectric factor (PEF), sonic logs (DTC and DTS), gamma ray (GR), bulk density (RHOB), density correction (DRHO), caliper (CALI), borehole size (BS), differential caliper (DCAL), rate of penetration logs (ROP and ROPA), spectra gamma ray (SGR), drilling mud weight (MUDWEIGHT), measured depth (DEPTH_MD), X and Y coordinates (X_LOC and Y_LOC), true vertical depth (Z_LOC), lithostratigraphic units (GROUP and FORMATION), true lithology (FACIES) and qualitative measure of interpretation confidence (CONFIDENCE). The last one is available only on the training data. And there are 12 types of facies: Sandstone, Sandstone/Shale, Shale, Marl, Dolomite, Limestone, Chalk, Halite, Anhydrite, Tuff, Coal and Basement.

Quality Control

First, we evaluated the facies count in the available data to better understand their distributions (Figure 2). Then the study assessed the number of missing data to identify which logs are most reliable (Figure 3).

The approach removed anomalous values from the training data that could add uncertainty to the model, as in the case of washout areas, for example. Highly positive or highly negative values in DRHO and DCAL logs suggest that these areas could be affected by some properties of the drilling mud after the wellbore enlargement. Therefore, it was decided to carefully remove these samples with extreme values.

Samples with infinite or meaningless values, such as negative values for resistivity logs, were converted into null values to later be estimated through imputation.

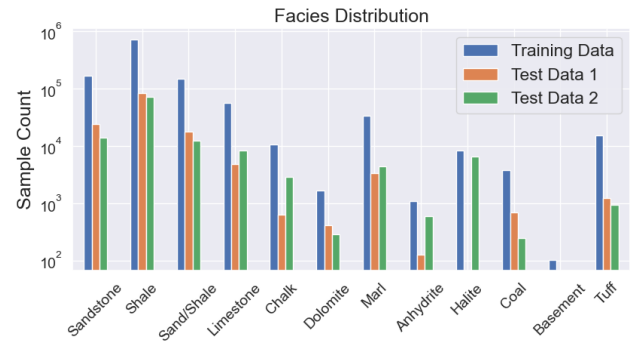


Figure 2 - Sample count showing the facies relation for each available data.

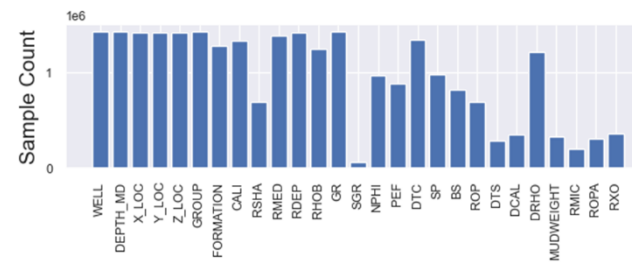


Figure 3 - Sample count with real values (not null) for training, test 1 and test 2 data altogether.

Data Imputation

This study correlated the features of the training data through Pearson's coefficient to analyze the statistical relationship between them and use this information in favor of the model (Figure 4). This kind of analysis is necessary to identify some type of important trend in your data.

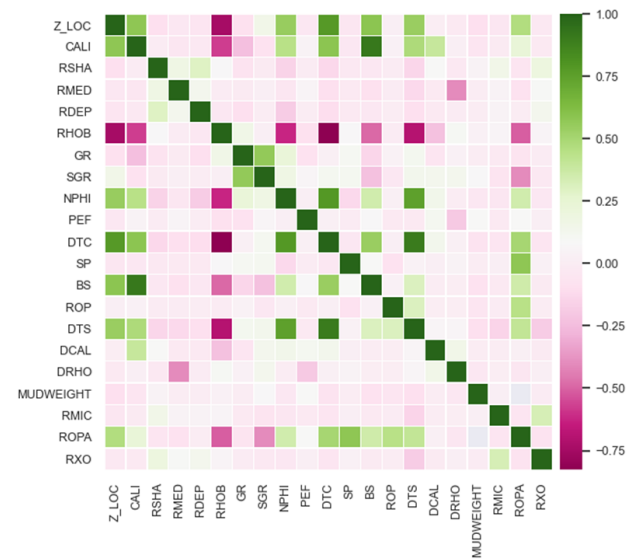


Figure 4 - Features correlation from training data with Pearson coefficient. Darker blocks represent high positive correlations and lighter ones high negative correlations.

This work decided to estimate values in place of missing data using a regression algorithm (*ExtraTrees*). The imputation is done in an automated way, taking into account the statistical and petrophysical character of the logs. After analyzing the largest absolute statistical correlations between them, groups of different logs were chosen manually to estimate each target property (Figure 5). In other words, this method created 20 regression models for each data set to estimate missing regions of 20 logs, each with its specific training data. Not all of them needed to be estimated because they were complete or did not have their own character for regression, as in the case of lithostratigraphy logs (GROUP and FORMATION).

Target	Training Data	Target	Training Data
Z_LOC	DEPTH_MD	DTC	DTS, RHOB, NPHI, Z_LOC
CALI	BS, DTC, RHOB, RDEP, RSHA	DTS	DTC, RHOB, NPHI, Z_LOC
RSHA	RXO, RDEP, RMED, RMIC	BS	CALI, Z_LOC, DTC
RMED	RSHA, RDEP, DRHO, DTS	ROP	ROPA, DTS
RDEP	RSHA, RMED, NPHI, RXO	DCAL	RHOB, CALI
RHOB	DTC, Z_LOC, DTS, NPHI, CALI	DRHO	DCAL, RMED, SGR
SGR	ROPA, GR	MUDWEIGHT	CALI, PEF
NPHI	DTS, DTC, RHOB, Z_LOC	RMIC	RSHA, DTS, RXO
PEF	DCAL, MUDWEIGHT, GR	ROPA	Z_LOC, RHOB, DTC, SP, ROP
SP	ROPA, NPHI, GR	RXO	RMIC, DTS, RDEP, RSHA

Figure 5 – Targets and their respective training data for missing data imputation.

Wavelet Transform

We decided to use the continuous wavelet transform (CWT) to generate attributes that would better separate classes that are difficult to differentiate, such as Shale and Shale/Sandstone. These facies with similar properties can be confused during the prediction, so a more qualitative analysis is needed. Perez-Muñoz et al. (2013) use wavelet transforms in different logs to correlate wavelet coefficients with sedimentary sequences to better identify facies.

The idea here is to use CWT as an auxiliary feature to identify facies transitions by the signal pattern, which would help to differentiate lithologies. The analysis of the signal spectrum creates qualitative models to identify the physical properties of the well. This study used the Morlet wavelet transform on the GR log to generate power and phase spectra (Figure 6). Three equidistant power and phase spectra, covering the whole CWT range, were added to the model for each data set as new data attributes.

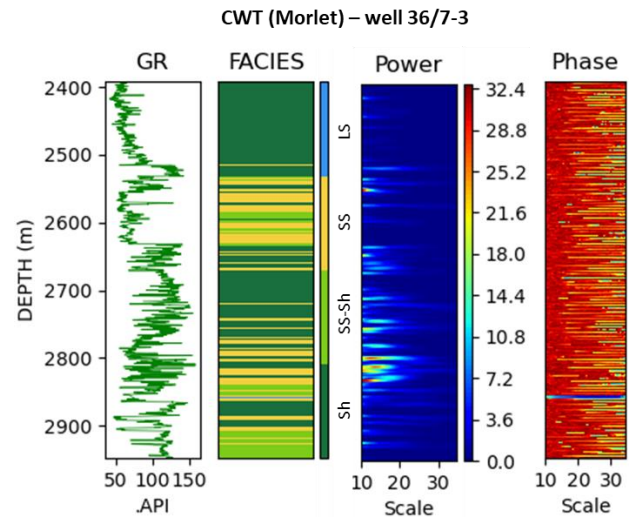


Figure 6 - GR log, facies and scalograms using Morlet wavelet transform for well 36/7-3 as example.

Feature Selection

This study selected the features that an interpreter would probably choose to identify the facies, as these logs have a greater petrophysical and spatial character. They are X_LOC, Y_LOC, Z_LOC, RDEP, GROUP, CALI, GR, RHOB, NPHI, PEF, DTC, SP and DRHO. Logs that originally had many missing values were not selected as they could add some uncertainty to the model.

Feature augmentation

Feature engineering is the process of creating additional features from the data to improve the model performance. This work applies two techniques: augment feature by gradient and polynomial expansion.

It is assumed that facies in neighboring layers have strong correlation, considering the physical nature of rock strata. Bestagini et al. (2017) proposes to employ the forward difference of a feature f in relation to depth, expecting that features in layers $d+1$ can help to classify a sample in position d :

$$f_{d,w}^{new} = \frac{f_{d+1,w} - f_{d,w}}{depth_{d+1,w} - depth_{d,w}}$$

where all operations are applied elementwise at depth d for each well w . The process creates a new feature for each existing one and then adds to the data.

The new combined set of features is then applied to an automatically polynomial expansion. This method returns polynomial combinations of features according to a degree. For example, a second-degree combination between three features returns the following new ones:

$$[a, b, c]^2 = a, b, c, ab, ac, bc, a^2, b^2, c^2, abc$$

These new attributes add information to the data and can capture relationships that did not exist before, thus increasing the predictive capacity of the model (Galli, 2020). Even if some of the augmented features might have no physical meaning, this is not a problem. Gradient boosting classifier will take care of non-informative features

by avoiding selecting them for tree creation at training time (Bestagini et al., 2017).

Shoulder Effect Removal

Commonly, facies interpreted from logs do not truly represent the lithology transitions due to the sampling rate of the well (about 15 cm). Therefore, two samples were removed at each facies limit to reduce the uncertainty that could be added to the model.

Data Split

The training data already contains more than 1 million samples and after all the steps of features generation, it becomes difficult to use this big data to make test runs and evaluate the model performance. This study uses the learning curve technique to identify an ideal reduced amount of samples so that your model does not lose important information during training.

The process evaluates the model accuracy with predictions about the training data itself and in cross-validation as the number of samples increases (Figure 7). The ideal value is determined from the point at which the curves are almost converging. Beyond that, unlikely more samples would significantly increase the accuracy of your model.

This technique indicated an ideal number of about 100 thousand samples, which were selected as 9 random wells from the *training data*. Another 5 wells of the same data were chosen for validation. Several processing steps were tested on this reduced data and, when maximum performance was achieved, these steps were applied on the original *training data*.

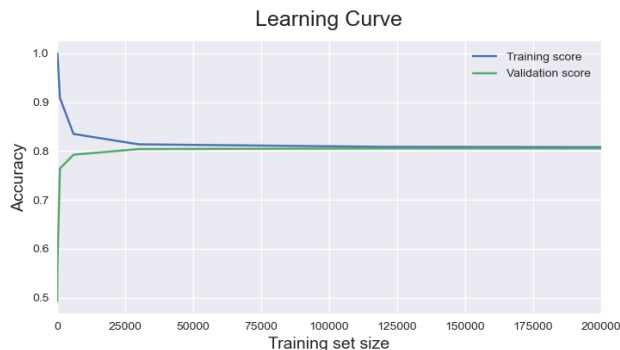


Figure 7 - Learning curve to indicate the ideal number of samples to be used in the training data for validation steps.

Extreme Gradient Boosting (XGBoost)

This work uses the XGBoost algorithm for the electrofacies classification. It uses ensembles of decision trees to boost the algorithm performance (boosted trees). It creates weak models and each model tries to correct the errors in the previous one (Dietterich, 2000; Chen & Guestrin, 2016). This process continues until the prediction is made correctly or up to a maximum addition of models.

XGBoost implements gradient boosting that is important to avoid overfitting. The process updates the new model with the latest prediction, minimizing the objective function by gradient descent (Friedman, 2001; Chen & Guestrin, 2016).

Facies Refinement

After prediction, if there is only one sample with facies different from neighboring, it may be that this sample represents an outlier and should be treated as a classification error (Bestagini et al., 2017). To solve this problem, it was decided to replace one or two isolated samples between similar strata of facies with the lithology of the predominant neighboring rock.

Performance metric

This work will use a confusion matrix to estimate the classification performance and the precision metric to evaluate how accurate is the model out of the predicted facies (Figure 8). Also, the competition suggested using a relative penalty matrix to assess the model performance. It works as a confusion matrix with a geological character that penalizes the errors of the interpreter (Figure 9). The sample penalty ranges from 1.375 for lighter errors to 4.000 for more serious ones. For example, inferring a region like dolomite, which was coal, would add a maximum penalty to your model. The predictions are then evaluated using the following function S:

$$S = -\frac{1}{N} \sum_{i=0}^N A_{\hat{y}_i y_i}$$

where *A* represents the matrix scores, *N* is the number of samples, \hat{y}_i is the real facies and y_i is the predicted facies.

		Predicted Class	
		Positive	Negative
True Class	Positive	True Positive	False Negative
	Negative	False Positive	True Negative

$$Precision = \frac{TruePositive}{TruePositive + FalsePositive}$$

Figure 8 - Representation of how a confusion matrix works and its relation to the precision metric.

Sandstone	0.000	2.000	3.500	3.000	3.750	3.500	3.500	4.000	4.000	2.500	3.875	3.250
Sand/Shale	2.000	0.000	2.375	2.750	4.000	3.750	3.750	3.875	4.000	3.000	3.750	3.000
Shale	3.500	2.375	0.000	2.000	3.500	3.500	3.750	4.000	4.000	2.750	3.250	3.000
Marl	3.000	2.750	2.000	0.000	2.500	2.000	2.250	4.000	4.000	3.375	3.750	3.250
Dolomite	3.750	4.000	3.500	2.500	0.000	2.625	2.875	3.750	3.250	3.000	4.000	3.625
Limestone	3.500	3.750	3.500	2.000	2.625	0.000	1.375	4.000	3.750	3.500	4.000	3.625
Chalk	3.500	3.750	3.750	2.250	2.875	1.375	0.000	4.000	3.750	3.125	4.000	3.750
Halite	4.000	3.875	4.000	4.000	3.750	4.000	4.000	0.000	2.750	3.750	3.750	4.000
Anhydrite	4.000	4.000	4.000	4.000	3.250	3.750	3.750	2.750	0.000	4.000	4.000	3.875
Tuff	2.500	3.000	2.750	3.375	3.000	3.500	3.125	3.750	4.000	0.000	2.500	3.250
Coal	3.875	3.750	3.250	3.750	4.000	4.000	4.000	3.750	4.000	2.500	0.000	4.000
Basement	3.250	3.000	3.000	3.250	3.625	3.625	3.750	4.000	3.875	3.250	4.000	0.000

Figure 9 - Penalty matrix used by the competition to penalize errors according to the real and estimated facies.

Results

First, the model performance on the ten wells of *test data 1* achieved an accuracy of 80% and a relative penalty of -0.5037. The classification on *test data 2* also reached an accuracy of 80% and a relative penalty of -0.4792. These results indicate a relevant performance for the model given the large number of samples and facies (Figure 10).

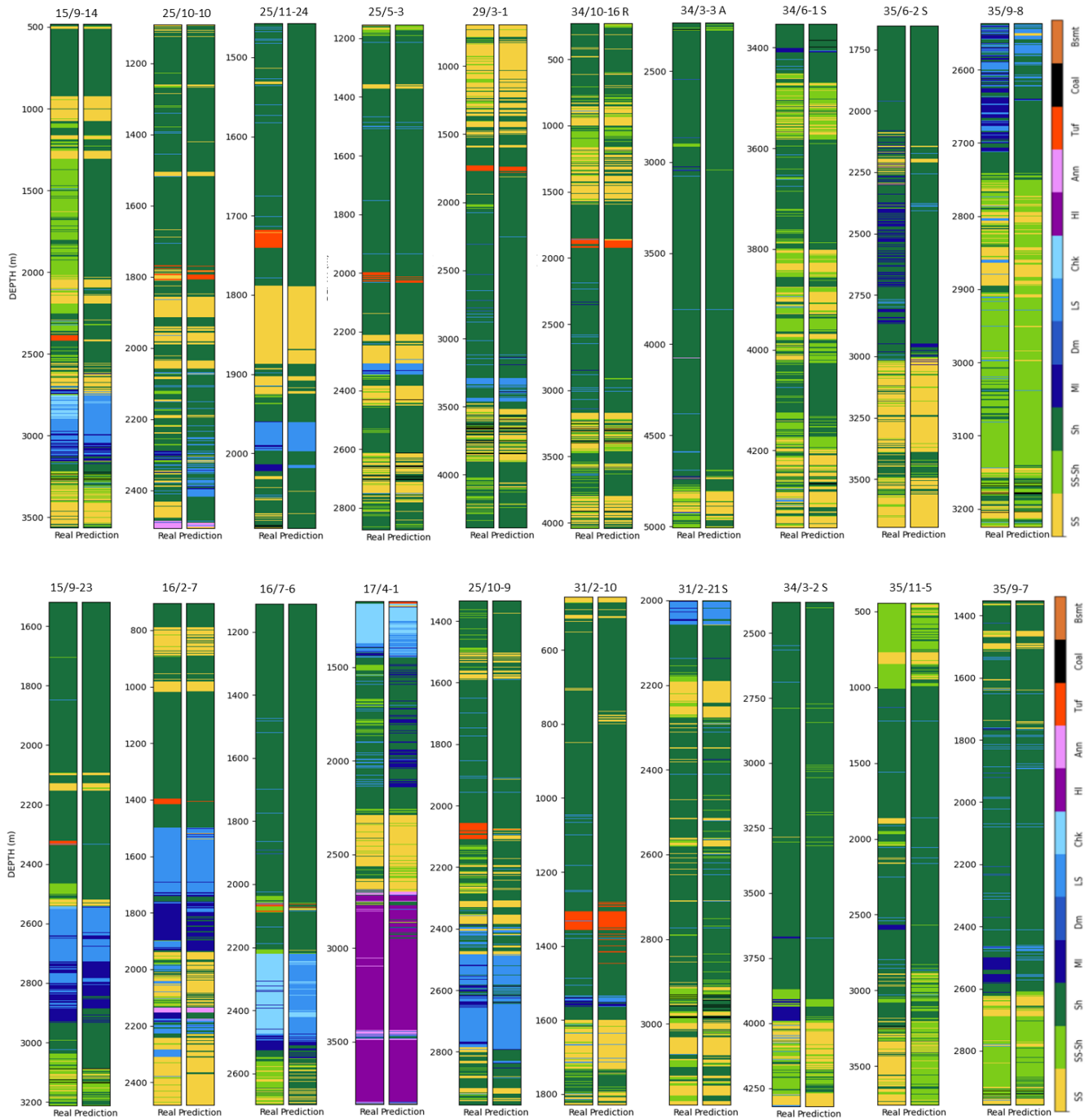


Figure 10 – Electrofacies classification of each well from *test data 1* and *2*, respectively. The facies on the left represent the real lithologies and the ones on the right represent the prediction.

When analyzing the precision metric in the confusion matrices, it can be seen that the model presented greater difficulty in correctly identifying samples from Sandstone/Shale, Limestone, Dolomite, Marl and Tuff (Figure 11). These facies are mostly confused with Shale. This can be explained by the large volume of Shale samples in the training data, which adds more weight to the model to estimate these facies. On the positive side, the classification achieves high precision for the most prevalent facies (Shale and Sandstone) and less common facies such as Chalk, Anhydrite, Halite and Coal.

In general, common errors, such as confusion between Shale and Shale/Sandstone, and Limestone with Chalk and Marl, are understandable since these rocks have similar properties. These errors are common even in non-automated interpretations through the reading of well logs. We realized that applying wavelet transform and feature augmentation were more effective in separate classes than oversampling methods.

These strategies with efficient imputation helped to obtain a better model performance. The algorithm would achieve relative penalties of -0.5578 and -0.5020 for *test data 1* and

2, respectively, without these three preconditioning steps. This proves that the estimation of values by regression instead of regions with missing values helps to create more reliable information; and the spatial data augmentation seeks a more detailed class differentiation.

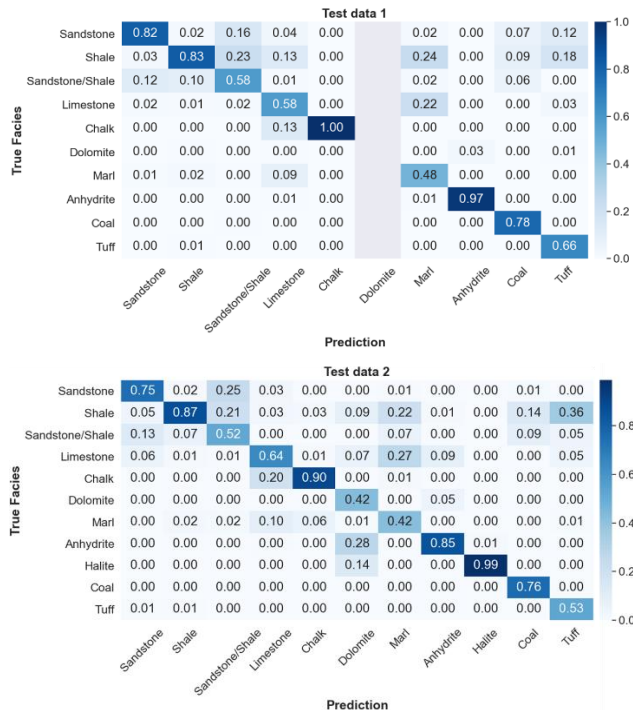


Figure 11 – Confusion matrices of test data 1 and 2 with precision indication for each facies.

Conclusions

We show in this study the possibility of obtaining high performance from an automated interpretation and optimization of the interpreter's time. This work adopted petrophysical perspective to specialize preprocessing to perform imputation, feature selection and engineering using wavelet transform to help differentiate specific lithologies. All the process concentrated on validation strategies in a reduced data to save time and gain information to define the ideal model.

Acknowledgments

The authors would like to thank Petrobras for funding the R&D project registered as “Desenvolvimento de fluxos de trabalho e metodologias para processamento e interpretação quantitativa de dados sísmicos 3D e 4D” and would like to thank the competition organizers for the opportunity and for the excellent experience we had.

References

Al-Mudhafar, W. J. (2020). Integrating Electrofacies and Well Logging Data Into Regression and Machine Learning Approaches for Improved Permeability Estimation in a Carbonate Reservoir In a Giant Southern Iraqi Oil Field. Offshore Technology Conference, Houston, Texas, USA, May 2020.

Bestagini, P., Lipari, V., & Tubaro, S. (2017). A machine learning approach to facies classification using well logs. SEG Technical Program Expanded Abstracts 2017.

Bormann, P., Aursand, P., Dilib, F., Dischington, P., Manral, S. (2020). FORCE Machine Learning Competition.

Chen, T., & Guestrin, C. (2016). XGBoost. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD '16.

Dietterich, T. G. (2000). An Experimental Comparison of Three Methods for Constructing Ensembles of Decision Trees: Bagging, Boosting, and Randomization. Machine Learning, 40(2), 139–157.

Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. The Annals of Statistics, 29(5), 1189–1232.

Galli, S. (2020). Python Feature Engineering Cookbook: Over 70 recipes for creating, engineering, and transforming features to build machine learning models. Packt Publishing Ltd.

Hall, B. (2016). Facies classification using machine learning. The Leading Edge, 35(10), 906–909.

Halland, E. K. Mujezinovic, J., Riis, F. (2014). CO2 Storage Atlas: Norwegian Continental Shelf, Norwegian Petroleum Directorate.

Perez, H. H., Datta-Gupta, A., & Mishra, S. (2003). The Role of Electrofacies, Lithofacies, and Hydraulic Flow Units in Permeability Predictions from Well Logs: A Comparative Analysis Using Classification Trees. SPE Annual Technical Conference and Exhibition.

Perez-Muñoz, T., Velasco-Hernandez, J., & Hernandez-Martinez, E. (2013). Wavelet transform analysis for lithological characteristics identification in siliciclastic oil fields. Journal of Applied Geophysics, 98, 298–308.

Puskarczyk, E. (2019). Artificial neural networks as a tool for pattern recognition and electrofacies analysis in Polish palaeozoic shale gas formations. Acta Geophysica. 67, 1991-2003.

Russell, S. J., Norvig, P., & Davis, E. (2010). Artificial intelligence: a modern approach. 3rd ed. Upper Saddle River, NJ: Prentice Hall.

Serra, O., & Abbott, H. T. (1982). The Contribution of Logging Data to Sedimentology and Stratigraphy. Society of Petroleum Engineers Journal, 22(01), 117–131.

Torlay, L., Perrone-Bertolotti, M., Thomas, E., Baciú, M., (2017). Machine learning - XGBoost analysis of language networks to classify patients with epilepsy. Brain Inform. 4 (3), 159–169.

Zhang, L., & Zhan, C. (2017). Machine Learning in Rock Facies Classification: An Application of XGBoost. International Geophysical Conference, Qingdao, China, 17-20 April 2017.

Zhang, D., Qian, L., Mao, B., Huang, C., Huang, B., Si, Y., (2018). A data-driven design for fault detection of wind turbines using random forests and XGboost. IEEE Access 6, 21020–21031.