Rio Oil & Gas Expo and Conference 2020

ISSN 2525-7579

Conference Proceedings homepage: https://biblioteca.ibp.org.br/riooilegas/en/

Technical Paper

Plataforma de geociências para automação da predição de propriedades de reservatório usando IA: estudo de caso em um campo do pre-sal da bacia de Santos

Geoscience platform for reservoir property modeling automation using AI: a case study in a pre-salt field in the Santos basin

João Zambrini [1] | **Félix Gonçalves** [2] | **Francois Lafferriere** [3] | **Mathieu Ducros** [4] | **Vinicius Matins** [5].

1. UFF, G&G, . RIO DE JANEIRO - RJ - BRASIL, joaozambrini@gmail.com 2. KOGNITUS, G&G, . RIO DE JANEIRO - RJ - BRASIL, felix@kognitus.com.br 3. KOGNITUS, G&G, . RIO DE JANEIRO - RJ - BRASIL, francois@kognitus.com.br 4. KOGNITUS, G&G, . RIO DE JANEIRO - RJ - BRASIL, mathieu@kognitus.com.br 5. KOGNITUS, G&G, . RIO DE JANEIRO - RJ - BRASIL, vinicius@kognitus.com.br

Resumo

Apesar da maturidade das técnicas de modelagem geológica usando geoestatística, existem situações, como no pré-sal brasileiro, onde esses métodos tradicionais não produzem uma boa representação das propriedades do reservatório devido ao alto grau de heterogeneidade das rochas. Essas limitações, entretanto, podem ser superadas usando modernas técnicas de Machine Learning amplamente usadas em outras indústrias. Infelizmente, sendo uma tecnologia de desenvolvimento mais recente, poucos são os geocientistas com as competências informáticas e matemáticas necessárias, o que dificulta uma aplicação efetiva nas equipes de E&P. Para superar este desafio, desenvolvemos uma plataforma – MachLee - para facilitar o acesso às mais modernas tecnologias de machine learning e a avaliação da qualidade das predições. Para alcançar a solução mais adequada em cada caso de estudo, a plataforma permite carregar e preparar os dados de subsuperfície e realizar a modelagem com uma classificação automática dos algoritmos de machine learning e sua parametrização, gerando uma solução de predição automatizada e pronta para aplicar. O uso dessa plataforma éilustrado por meio de um exemplo em um campo do prá-sal da bacia de Santos onde propriedades de poços foram previstas com base em aprendizagem sobre os dados disponíveis.

Palavras-chave: Reservatório. Modelagem. Machine Learning. Predição. Pré-sal

Abstract

Despite the maturity of geological modeling techniques using geostatistics, there are situations, as in the Brazilian pre-salt, where these traditional methods do not produce a good representation of the reservoir properties due to the high degree of heterogeneity of the rocks. These limitations, however, can be overcome using modern Machine Learning techniques widely used in other industries. Unfortunately, due to the more recent development of such techniques, there are still few geoscientists with the necessary computer and mathematical skills to apply them in the E&P workflow. To overcome this challenge, we developed a platform - MachLee - to facilitate access to the most modern machine learning technologies and assess the quality of predictions. MachLee enables loading and preparing the subsurface data and automating the selection, classification, and parameterization of machine learning algorithms, generating a prediction solution ready to apply. The use of this platform is illustrated through an example in a Santos basin pre-salt field where well properties were predicted based on learning about the available data.

Keywords: Reservoir. Modeling. Machine Learning. Prediction. Pre-salt

Received: February 28, 2020 | **Accepted:** Jun 06, 2020 | **Available online:** Dec. 01, 2020.
Article nº: 406
Cite as: Rio Oil & Gas Expo and Conference, Rio de Janeiro, RJ, Brazil, 2020 (20)
DOI: https://doi.org/10.48072/2525-7579.rog.2020.406

1. Introduction

The geological modeling of the reservoir has undergone a significant evolution in the last decades with the development of software that facilitates the integration of 3D data and incorporates the most modern techniques of geostatistics. There are situations, however, in which the application of this traditional approach might be insufficient to obtain a good representation of the reservoir. This is the case when few well data are available to characterize the reservoir property distribution of rocks with a high degree of heterogeneity: geoscientists might then need to rely more heavily on secondary data such as seismic information for lateral extrapolation or interpolation of the data.

In recent years, the exponential advance of artificial intelligence technologies and the growing demand for increased efficiency in the E&P area have promoted the use of Machine Learning (ML) algorithms to accelerate and automate data analysis and support better business decisions. With great versatility, these algorithms have generated excellent results and are particularly performant when used in complementarity with traditional methods through the control of enabled domain experts. But its application in geosciences workflow like Reservoir Property Modeling raises some challenges such as data integration, data scaling, algorithms selection and parametrization, etc.

To overcome these difficulties, accelerate the learning curve of E&P teams and streamline the integration of new technologies in the day-to-day work of E&P teams, Kognitus jointly develops with RepsolSinopec a new ML platform for the integration of seismic and well data applied to reservoir geological modeling workflow. The platform integrated workflow was tested with a real pre-salt dataset from Santos Basin on a 3D reservoir grid property modeling example.

2. Machine Learning Integrated Geosciences workflow

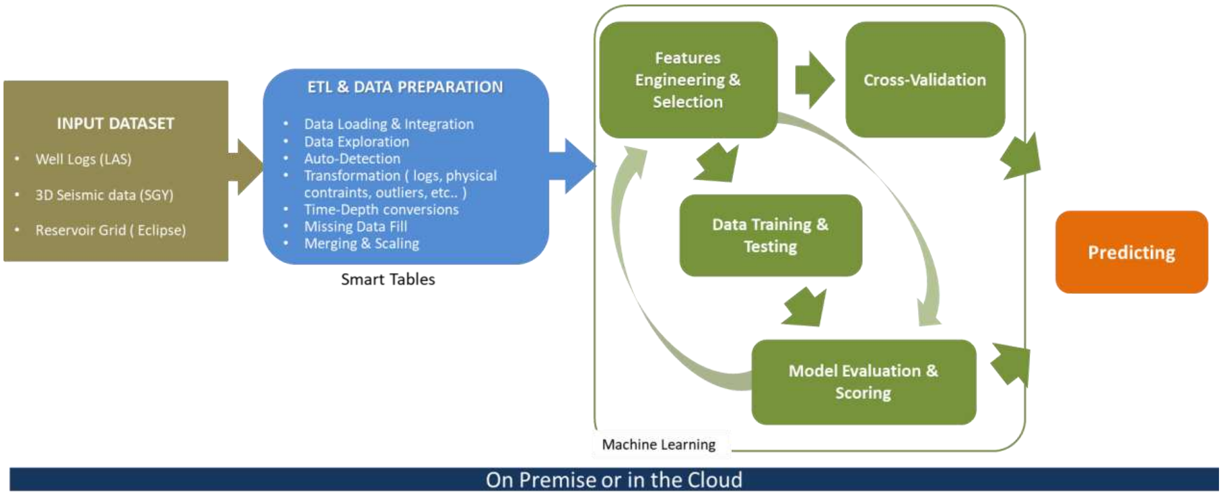
Traditional Reservoir Property Modeling workflows relies heavily on geostatistic to propagate reservoir properties in the 3D space based on well data (hard dataset). Through variography of an input dataset, geostatistic captures the main spatial characteristics of the input data (both vertically and laterally) and reproduces these characteristics to simulate or to estimate a property at reservoir scale. To better control lateral distribution of the property, it is often used a soft-data derived from seismic information through a direct correlation with the estimated property.

While geostatistic-centric workflows are very powerful to quickly provide many realistic images of the Reservoir with their associated uncertainties, it does not allow to incorporate directly the multiple factors that are involved in complex geological deposition. In comparison, Machine Leaning techniques have proven to be very powerful at dealing with extensive list of input variables and ranking their respective relative influence. It allows to easily include several seismic attributes and inversions that can be derived from recent high-quality seismic surveys and modern processing. Each seismic attribute – and more generally any secondary data- provide a specific insight on the rock property through a direct or indirect relationship.

The Machine-Learning platform - MachLee - seeks to guide the geoscientist through the data loading and preparation workflow, accessing state-of-the-art supervised ML algorithms, predictive modeling, and results analysis. The objective is to allow the practical and rapid application of several ML algorithms for the propagation of rock properties in the well or reservoir scale and to rank the results to select the best methods in each specific case.

Because the ML algorithms require a training and a testing set (different from geostatistics) and an extensive data preparation step, it is needed to adapt our traditional Geosciences workflow, as described in the Figure 1.

Figure 1 – MachLee workflow



Source – From the authors

Several critical aspects of classical Machine Learning workflow were addressed for the implementation of this workflow in the MachLee platform:

- Agnostic ML environment. Because ML algorithms are in constant evolution with heavy international research programs -not limited to O&G-, generic wrappers were been developed to easily plugin any public or private ML algorithm library onto the geological object loaded into the platform (Well, Seismic, Reservoir Grid) .

- Repeatable. To be able to share the workflow among geoscientists teams and to continuously improve the workflow, any individual step of the workflow as well as its entire orchestration are stored so they can be easily repeated.

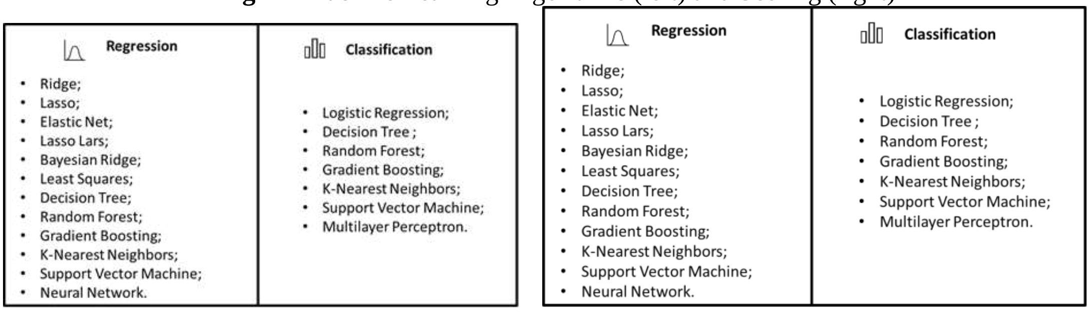
- Automated. To simplify the Data Preparation step, a series of automated detections are implemented, including: property type (categorical vs continuous) , property distribution type (logarithmic, normal, etc..) and outliers. An automated strategy has also been implemented to propose to the user to automatically selected the best features for a given property prediction.

All input data are automatically transformed into SmartTables, that are tabular container storing geosciences data in optimized format. SmartTables keeps track of all data transformations applied to any property (column of the table) such as log-transformation, mathematical operation, normalization, etc... With this 'memory' capability, ETL workflows can be fully automatized at larger scale, and to treat new incoming data (new well for instance).

The algorithms implemented include both Regression for continuous properties (porosity, permeability, etc.) and Classification for discrete properties (Rock Type). At that point only Pointwise algorithms were considered, working on individual point or cell basis, meaning they do not consider the neighboring value for the prediction of a given point of cell. To mitigate this strong limitation when dealing with spatially correlated variable, we used a windowing strategy as described later.

For Algorithm evaluation, scoring metrics have also been implemented to gauge the performance of each algorithm on the testing or cross-validation dataset.

Fig 2 – Machine Learning Algorithms (left) and Scoring (right)



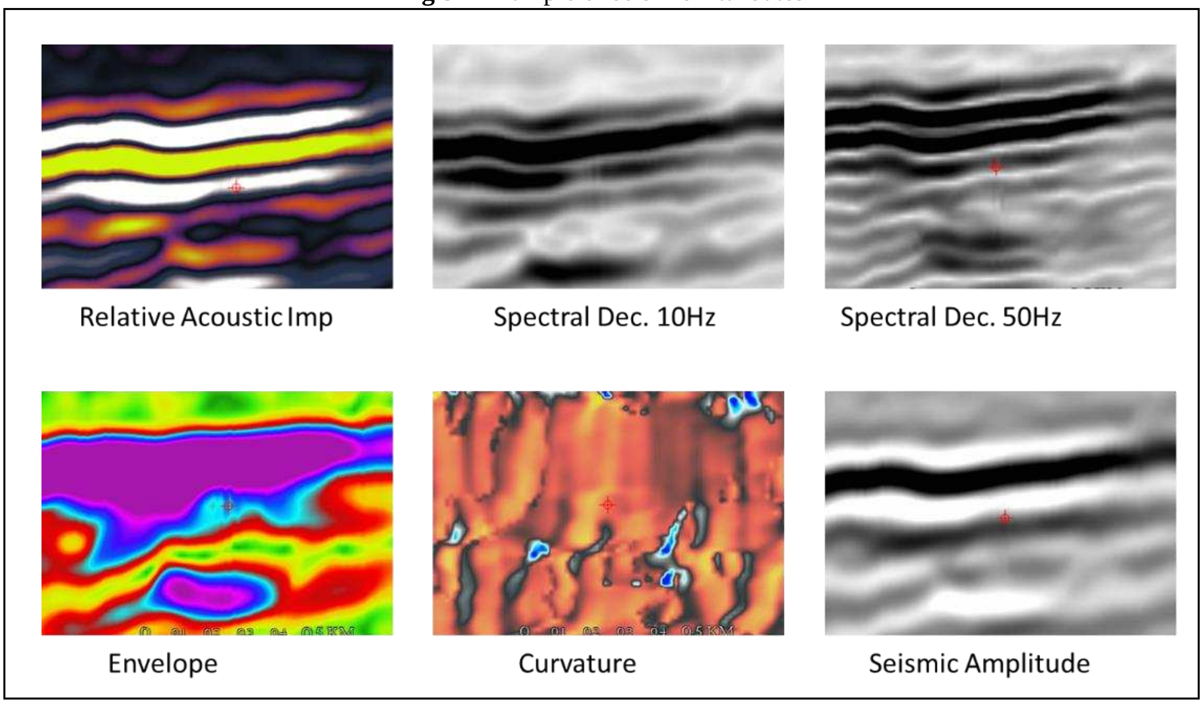
Source – From the authors

3. Application to Offshore Brazil Dataset

The above workflow was applied on an offshore field dataset from the Santos Basin, composed of seven wells profiles together with a 3D cube seismic and a 3D reservoir grid previously built at reservoir level into a Modeling software. The main objective was to evaluate the predictive capacity of each algorithm for the distribution of PHIE based on seismic derived features, to identify the best algorithms to be used for Reservoir Grid prediction.

Several Seismic Attributes have been extracted from the original post-stack Seismic Dataset, of which 13 were kept, representative of the different attribute families between Instantaneous, Geometric, Amplitude Accentuating and Spectral attributes. On purpose, the number of input attributes was not restricted too much to let the workflow automatically select the best features through the auto-selection functionality. In addition, we used an acoustic inversion result as an additional input feature and the K value of the ReservoirGrid to account for the relative depth, totalizing 15input features as input for the ML algorithms.

Fig 3 – Example of seismic Attributes



Source – From the authors

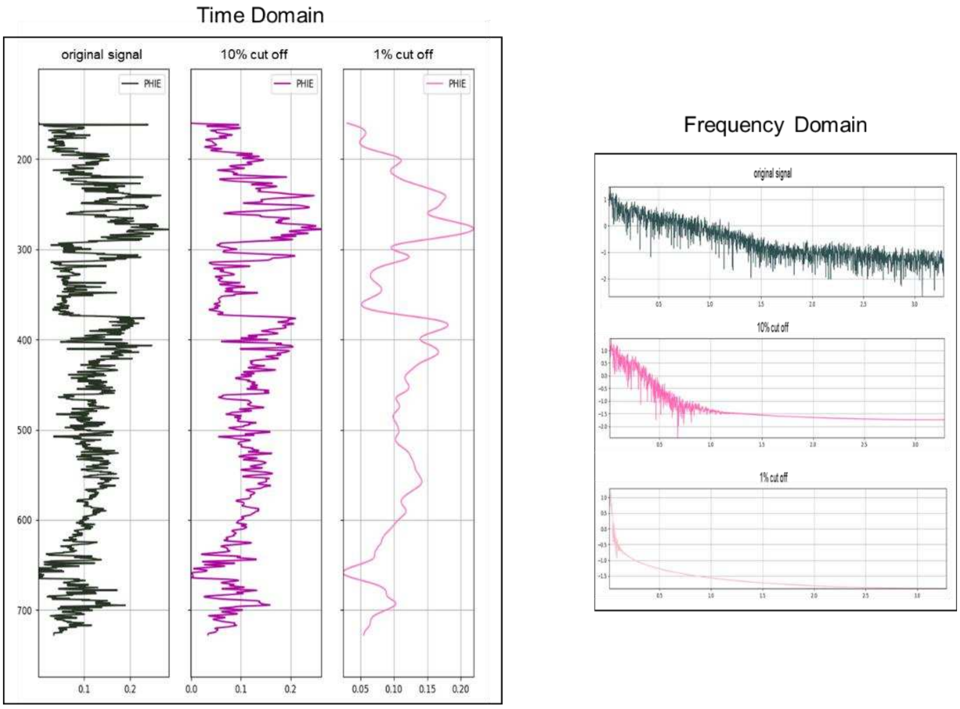
3.1 Data Preparation, Feature Engineering &Scaling

The platform allows fast and efficient loading, quality control and cleaning of all input subsurface data. A series of tools are available to guide the geoscientist in the exploration and correlation of data, elimination of null values, statistical treatments, and data transformations, among other editions. In particular, it automatically identified Continuous vs Categorial variables, as well as outliers based on physical range pre-defined template. It also diagnoses the shape of the property distribution (Logarithmic vs Normal) to alert the user on non-gaussian distribution that may bias the prediction.

These automated editions and corrections are fundamental for the workflow, as ML algorithms are very sensitive to quality of features quality. Geoscientists (as much as Data Scientists) knows this data preparation is also the most time consuming of the workflow. We aim at reducing this time pushing further the automated process while still unable geoscientists to QC and correct manually the data if necessary.

The different input dataset not only have different sampling (0.2m for well logs / 5m for the seismic-derived attribute), as well as resolution and frequency content. The ML algorithm would not be able to reasonably predict any variable at higher frequency content than its input features. To reconcile the different scale of the input data, we used a Fourier Transform Frequency filtering in the time-domain to filter the target variable (PHIE) to match the frequency content of the input features.

Fig 4. Frequency Filtering of PHIE log with low pass filter at different cut-off values



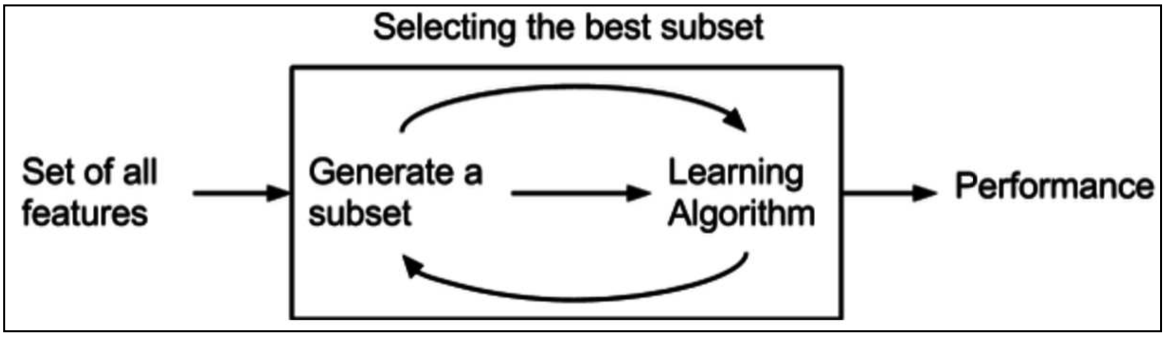
Source – From the authors

Finally, to account for the spatial continuity of Reservoir Properties, we used a Windowing strategy: we translated each original input features of 3 vertical samples the in both direction (+ 3 / -3 samples). The number of samples to be shifted has been chosen arbitrarily after several testing. This windowing considerably increases the number of input features, which will require a tool to select only the most relevant features as input for the algorithm of Machine Learning.

3.2 Automated Features Selection with SFS

‘To limit the number of features, to avoid the redundant information and to mitigate the associated issues such as overfitting, we used a Sequential Forward Selection (SFS) wrapper method that automatically test k features among the total n features (n=14 in our case) for a particular algorithm. The algorithm is iteratively adding new features and measuring the incremental additional (or loss) of performance on a given metric.

Fig 5 – Wrapper best features selection

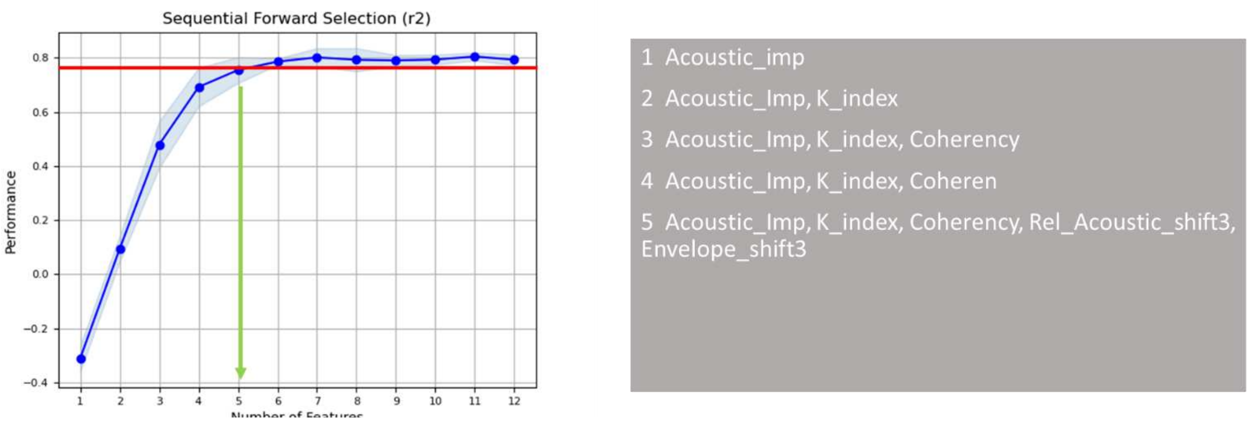


Source – Analytics Vidhya

Some features might be more or less relevant for a given algorithm, and consecutively each algorithm will have a unique best k-feature subset. This technique is computer intensive, but considering the limited size of the training dataset in geoscience (number of wells) , it is not perceived as a limitation for the automated workflow.

As we can see in the graph below, the performance typically reaches a plateau (at 5 features in this case) , to which point the marginal improvement of adding more input features is not significant, and could even be detrimental to the explicability of the model.

Fig 6 – Performance of subset of k-feature and associated features for the Gradient Boosting Algorithm



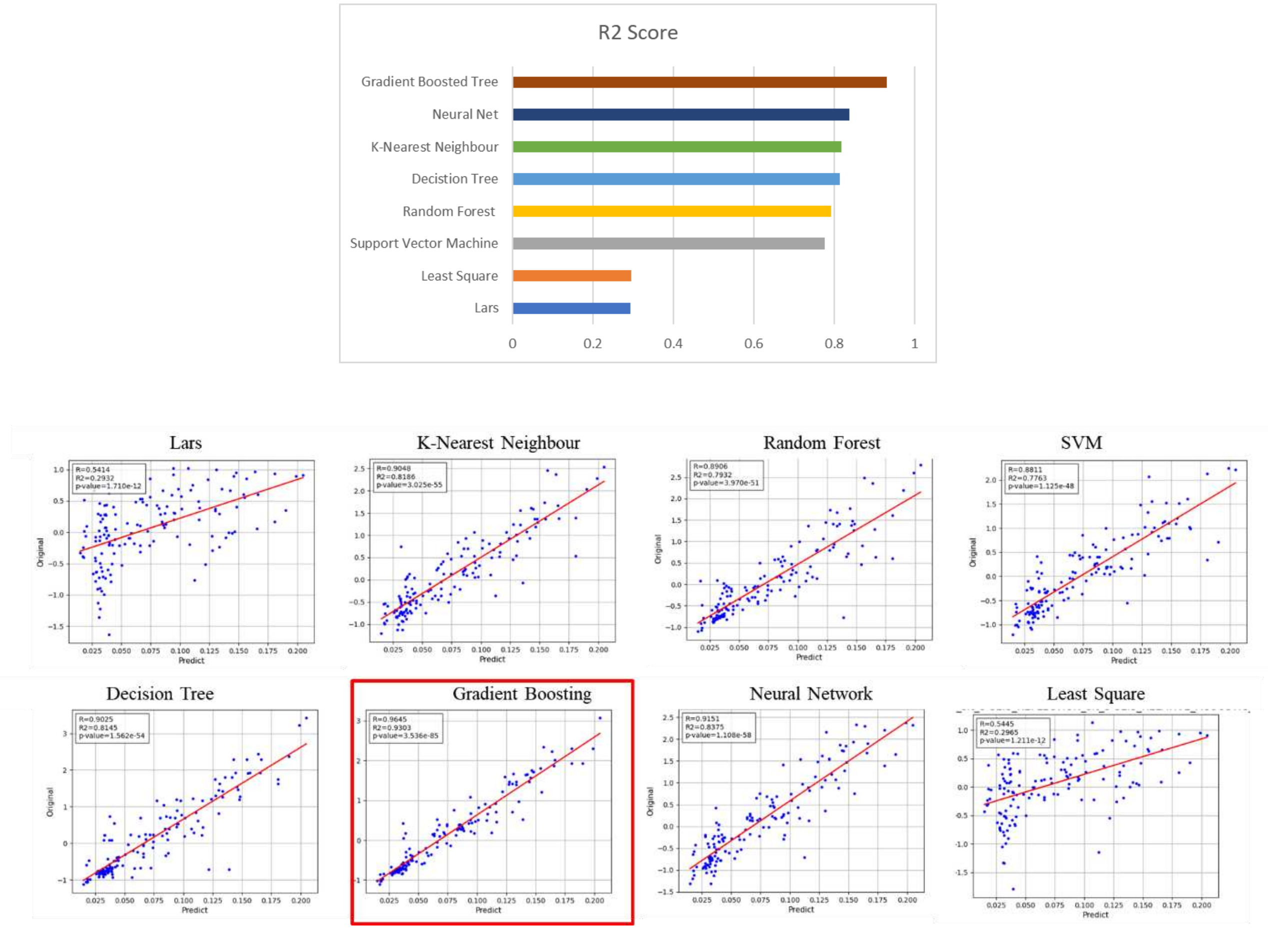
Source – From the authors

3.3 Scoring & Ranking

ML relies entirely on testing strategy to be evaluated and to provide a trusted output prediction. The input dataset was randomly split between Training (80% of the data) and Testing (20% of the data). Because of the reduced size of the training dataset available with only 7 wells, we could not perform a cross-validation test, that would have still reduced the number of data points from the training.

Each of the 8 algorithms pre-selected was individually trained using the optimal number of features selected from the SFS strategy run previously and ranked according to its R2 results. Most of the algorithms performed well, with a highest R2 Score for the Gradient Boosted Tree.

Fig 7 – Ranking & Prediction on the Testing Dataset

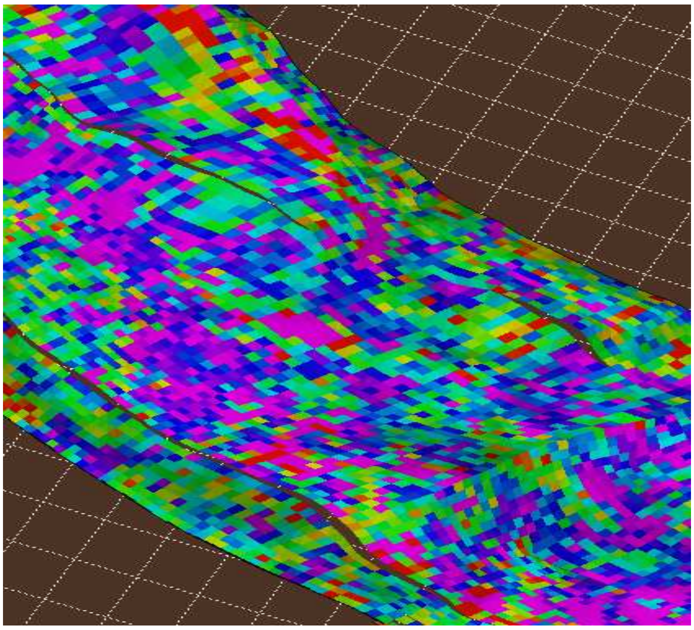


Source – From the authors

3.4 Predicting

The best model (according to R2 metric) from the 'Gradient Boosted Tree' was selected with its associated features and was simply applied to the entire ReservoirGrid (about 3 million cells). PHIE result is displayed below, where we noticde the mix of influence of several of the seismic attributes, as well as the lateral continuity inherited from the input features.

Fig 8 – 3D top view of the predicted PHIE distribution. Scale omitted on purpose for confidentiality



Source – From the authors

4. Final Considerations

The use of an integrated platform (MachLee) with automated functionalities that assist he user in the choice of ML variables and algorithms, provides an environment to test alternative workflow from traditional methodology for predicting reservoir properties. The key advantage of the Machine Learning algorithms is to be able to incorporate several input features into the prediction, and to validate a posteriori the performance of the prediction on a testing dataset. Spatial continuity of the geological rock properties is controlled by the input 3d features derived from the seismic, and by a windowing strategy that use the neighbors of each cell to estimate the output property. This workflow can also incorporate geostatistical simulation as additional input features if necessary. Additionally, it allows real-time updates of the model when more data is added to the project, as every Data Preparation step is fully registered.

In the case study described above, good results were obtained (based on R2 score on testing dataset), even though only seven wells were available. Future work includes the incorporation of latest generation of ML algorithm accounting for spatial information (Convolutional Neural Network) that potential can improve further the predictive capability.

The development of the platform used in this study meets one of the main challenges faced by the O&G industry with regard to the incorporation of AI technologies into the E&P decision-making process, which is the lack of domain-specific AI platforms that can allow asset teams to collect and prepare data, build models, automate workflows and deploy ML/DL solutions without necessarily being able to code.

5. Thanks

The authours would like to thank RepsolSinopec for the partnership to develop MachLee and the authorization to give this oral presentation.

Referências

Al-Mudhafar, W, & Autor 1. (2017). *Integrating well log interpretations for lithofacies classification and permeability modeling through advanced machine learning algorithms*. Journal of Petroleum Exploration and Production Technology.

Anifowose, F, & Autor 1. (2013). *Artificial intelligence application in reservoir characterization and modeling: Whitening the black box*. Society of Petroleum Engineers - Saudi Arabia Section Young Professionals Technical Symposium.

Chaki, S, Routray, A., & Mohanty, W. (2018, March 1). Well-Log and Seismic Data Integration for Reservoir Characterization: A Signal Processing and Machine-Learning Perspective. *IEEE Signal Processing Magazine*.

Pudil, P. N., & Autor 1. (1994). *Floating search methods in feature selection*. Pattern recognition letters 15.11(: 1119-1125.

Saikia, P., Baruah, R, Singh, S., & Chaudhuri, P. (2020). *Artificial Neural Networks in the domain of reservoir characterization: A review from shallow to deep models.* Computers and Geosciences: Computers and Geosciences.

Soumi Chaki, & Routray William Mohanty. (2015). *RESERVOIR CHARACTERIZATION: A MACHINE LEARNING APPROACH*. Master of Science: Master of Science.