9º CONGRESSO BRASILEIRO DE PESQUISA E DESENVOLVIMENTO EM PETRÓLEO E GÁS

TÍTULO DO TRABALHO:

DETERMINAÇÃO DE LITOLOGIA EM POÇOS DE PETRÓLEO ATRAVÉS DE DADOS DE REGISTROS: UMA ABORDAGEM BASEADA EM INTELIGÊNCIA COMPUTACIONAL

AUTORES:

Camila Martins Saporetti[1], Leonardo Goliatt da Fonseca[1], Leonardo da Costa Oliveira[2], Egberto Pereira[3]

INSTITUIÇÃO:

1 Universidade Federal de Juiz de Fora
2 Petrobras
3 Universidade Estadual do Rio de Janeiro

DETERMINAÇÃO DE LITOLOGIA EM POÇOS DE PETRÓLEO ATRAVÉS DE DADOS DE REGISTROS: UMA ABORDAGEM BASEADA EM INTELIGÊNCIA COMPUTACIONAL

Abstract

This work aims at the developing of a computational model for a lithology characterization that assists in the later stages of sedimentary basin analysis. The lithology will be derived from the recognition of patterns of petrophysical characteristics. To be addressed in the determination of the petrographic classes from tests under analysis, unsupervised computational intelligence techniques, and is not the condition of the pre-defined classes. For an visualization of the results was employed Principal Component Analysis. The data used for the validation of the methodology were collected from the Exploration and Production Database (BDEP) of the National Petroleum Agency (ANP). The proposed approach implements a procedure to evaluate the performance of computational intelligence techniques and appears as an alternative to assist the geologist / specialist in the determination and characterization of the lithology, contributing to a reduction of cost and increase in the accuracy of results.

Introdução

A definição de litologia em poços de petróleo por meio de múltiplos perfis de análise geofísica tem um papel importante no processo de caracterização do reservatório. A partir da litologia, pode-se gerar modelos que, por sua vez, serão a base a partir da qual alguns cálculos petrofísicos são feitos, e em seguida, podem ser utilizados em simuladores de fluxo para compreender e estudar o comportamento de um campo de petróleo. Através de sua determinação pode-se obter informações sobre as rochas petrolíferas, tais como: sua história deposicional e diagenética, estrutura do poro e mineralogia. A identificação de litologia pode ser feita por métodos diretos e indiretos. Métodos diretos são realizados através da obtenção de amostras petrofísicas do reservatório. Esta é a maneira mais segura de determinar inequivocamente o tipo de litologia e rocha, no entanto, obter essa amostra física nem sempre é uma tarefa fácil. Os *mud logs* são a primeira escolha em poços *wildcat*, mas às vezes a atribuição exata de um determinado fragmento de rocha em uma determinada profundidade pode não ser precisa. Métodos indiretos fazem uso de registros de poços que medem as propriedades físicas de formações geológicas e fluidos fornecendo a maioria dos dados subterrâneos disponíveis para um geólogo de exploração. Além de sua importância nas decisões finais, eles também são ferramentas inestimáveis para mapear e identificar litologias. No entanto, métodos indiretos não geram o mesmo desempenho que os métodos diretos. Consequentemente, torna-se necessário automatizar o procedimento de caracterização do reservatório, nesse contexto técnicas de inteligência computacional aparecem como uma alternativa para identificar litologia.

Existem na literatura trabalhos utilizando técnicas de inteligência computacional para resolução de problemas na geociência e na geofísica. SAGGAF & NEBRIJA (2000) criaram um método para automatizar a análise de dados de registros baseado em rede neural artificial para identificar litologias e fácies deposicionais. Recentemente METHE *et al*. (2017) empregaram análise de agrupamento para obter informações sobre litologia. Testaram vários algoritmos de agrupamento (agrupamento hierárquico de Ward, K-Means, Mean-Shift e DBSCAN) em dados geofísicos de uma perfuração. AMEUR ZAIMECHE *et al.* (2014) apresentaram um método baseado em clusterização para prever fácies. A técnica é baseada em um algoritmo de determinação de fácies seguido de codificação de eletrofácies usando dados petrofísicos, especialmente (GR, ROHB, THOR, POTH) e

descrições de núcleos detalhadas. SEBTOSHEIKH & SALEHI (2015) utilizaram o método SVM para prever a litologia a partir de dados de atributos sísmicos invertidos e registros de poços baseados em estudos petrográficos de núcleos litológicos em um reservatório carbonático heterogêneo no Irã. DONG *et al.* (2016) introduziram um estudo aplicando a Análise Discriminante com kernel Fisher e uma Análise Discriminante Linear melhorada para identificar litologia.

Tendo em vista o crescente interesse em estudar a litologia e o esforço gasto em identificá-la. Este trabalho objetiva o desenvolvimento de um modelo computacional para a caracterização de litologia que auxilie nas etapas posteriores da análise de bacias sedimentares. A litologia será derivada do reconhecimento de padrões de características petrofísicas, ou seja, características semelhantes serão agrupadas e formarão uma classe petrográfica. A descrição do procedimento adotado e das técnicas aplicadas podem ser vistos na Seção Materiais e Métodos.

Materiais e Métodos

Os dados de registros utilizados estão em arquivo com formato Las ASCII Standard (LAS), que podem ser importados e exportados para qualquer plataforma, estação de trabalho ou mainframe sem a necessidade de softwares proprietários. Desde sua criação o LAS se tornou o formato mais usado para transferência digital de dados de logs de poços e, de certa forma, um verdadeiro padrão da indústria (HESLOP *et al.*, 1999).

O arquivo possui 11 informações (a profundidade e 10 características petrofísicas) para 2297 entradas analisadas. As características e as descrições encontram-se na Tabela 1.

Tabela 1: Logname e descrições

Logname	Descrição
CALI	Diâmetro da perfuração
CNST	Espessura estratigráfica verdadeira
DRHO	Densidade Aparente Corrigida
FCNL	Registro de Neutron Compensado (detector longe)
FFDC	Registro de Formação Compensada (detector longe)
GR	Raio Gamma
NCNL	Registro de Neutron Compensado (detector próximo)
NFDC	Registro de Formação Compensada (detector próximo)
NPHI	Porosidade de Neutron
RHOB	Densidade Aparente

A metodologia usada neste trabalho é dividida em três procedimentos principais: (I) pré-processamento dos dados, (II) escolha dos parâmetros ótimos do método de agrupamento e aplicação do mesmo e (III) visualização dos grupos encontrados. Para os procedimentos realizados foi descartada a profundidade.

O pré-processamento dos dados foi através de uma normalização dos dados. Ao normalizar os dados a influência de escala é retirada. A fórmula utilizada é descrita abaixo:

$$X = \frac{X - media(X)}{std(X)} \quad (1)$$

em que X são os dados de cada característica petrofísica, media(X) é a média dos dados e std(X) é o desvio padrão.

O método de agrupamento utilizado foi o K-Means (VENDRAMIN *et al.*, 2010). Este é um dos algoritmos mais utilizados para realizar agrupamentos. A partir da escolha de centroides iniciais, de forma aleatória, cada amostra é atribuída ao centroide mais próximo. O próximo passo é atualizar os centroides tomando o valor médio de todas as amostras designadas para cada centroide anterior. Calcula-se a diferença entre os antigos e os novos centroides, repetindo o processo a partir da atribuição de amostras aos centroides, até que este valor seja inferior a um limiar. O processo se repete até que os centroides não se movam de forma significativa. O número de agrupamentos a ser gerados é passado como parâmetro da função. A medida de dissimilaridade utilizada foi a distância Euclidiana. A Figura 1 ilustra a solução final de um agrupamento para dois grupos.

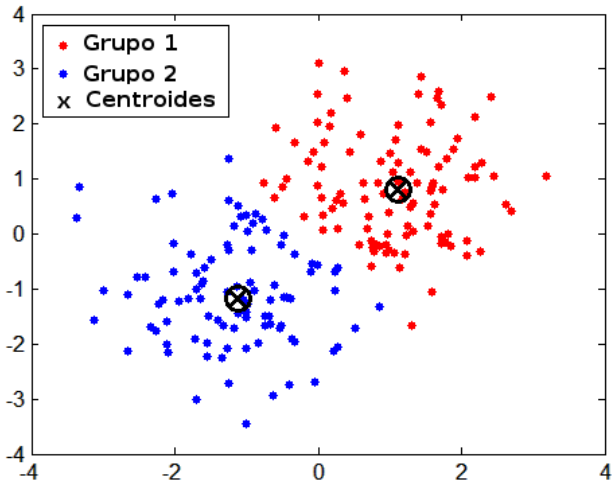

Figura 1: K-Means - K = 2.

Critérios de validação são medidas adotadas para avaliar a qualidade de um agrupamento. A escolha do critério depende do problema a ser tratado. Existem algumas publicações tratando desse assunto e comparando critérios. A análise de silhueta (SC) é uma técnica proposta por (ROUSSEEUW, 1987) . Trata-se de um método geométrico baseado na compactação e separação de agrupamentos com o intuito de analisar a qualidade dos agrupamentos formados. O número ótimo de agrupamentos é definido pelo maior coeficiente de silhueta resultante da análise de silhueta. Para cada amostra *i* o valor s_i é definido pela seguinte fórmula:

$$s_i = \frac{b_i - a_i}{max(a_i, b_i)} \qquad (2)$$

Considerando que a amostra *i* pertença ao agrupamento A, a_i é descrito como a dissimilaridade média da amostra *i* em relação a todas as outras amostras do agrupamento A. Seja B um agrupamento diferente de A, b_i é a dissimilaridade média da amostra *i* em relação a todas as amostras de B. O coeficiente de silhueta de um conjunto de dados é dado pela média dos coeficientes individuais das amostras

$$SC = \frac{\sum_{i=1}^{N} s_i}{N} \qquad (3)$$

em que *N* é o número de amostras do conjunto de dados. A métrica utilizada para o cálculo da dissimilaridade é a distância Euclideana. O valor de SC varia de -1 a 1. Resultado próximo de 1, indica que os objetos estão bem agrupados.

Para visualizar os resultados foi utilizado a Análise de Componentes Principais (ACP). A ACP é um método que tem como propósito, a análise dos dados usados objetivando sua redução, eliminação de sobreposições e a escolha dos modos mais representativos de dados através de combinações lineares das variáveis originais. Esta técnica é uma forma de identificar relação entre atributos retirados de dados. É muito útil quando os vetores de atributos possuem muitas dimensões, uma vez que é impossível uma representação gráfica.

Resultados e Discussão

A implementação do método proposto foi realizada na linguagem Python. Esta linguagem conta com um módulo, *Scikit-learn (* PEDREGOSA *et al.*, 2011) que integra uma gama de algoritmos de aprendizado de máquina para problemas supervisionados e não-supervisionados.

Os dados foram normalizados e em seguida aplicou-se ACP. Na Tabela 2 encontram-se as componentes obtidas. Para a Componente 1 as características GR, NFDC e NPHI possuem influenciam mais, na Componente 2 são as características FFDC, NFDC e NPHI. A Figura 2 apresenta a distribuição das amostras, nota-se que há um grupo bem denso.

Tabela 2: Componentes Principais

	Componente 1	Componente 2
CNST	-6.74268787E-03	8.98905272E-04
DRHO	-1.60573376E-03	2.54982173E-04
FCNL	-2.40105011E-04	1.30685655E-03
FFDC	-1.77614567E-01	-2.54969427E-01*
GR	8.84238329E-01*	-8.03684520E-02
NCNL	-1.03019566E-01	7.06628846E-02
NFDC	-3.63797421E-01*	-5.43971276E-01*
NPHI	-2.07758007E-01*	7.92124287E-01*
RHOB	-1.97103441E-02	1.29965584E-02
DEPT	-3.74989694E-03	1.06468082E-03

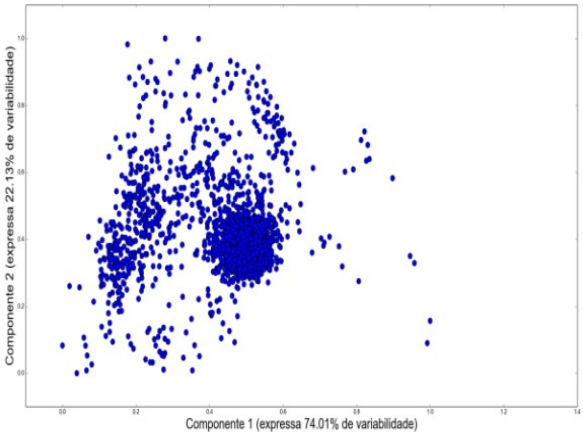

Figura 2: Distribuição dos Dados

O K-Means foi empregado para agrupar as entradas após o pré-processamento. O parâmetro ótimo foi encontrado através de uma busca exaustiva maximizando a silhueta, coeficiente de silhueta que gerou o melhor resultado foi 0.652. A Tabela 3 apresenta a média, o valor mínimo e o máximo para cada característica de acordo com os duas classes petrográficas geradas. As características que podem-se notar diferença entre os dois grupos são DRHO, NFDC e NPHI.

A Figura 3 mostra DRHO, NFDC e NPHI em relação a profundidade para cada classe. A primeira coluna são as características referente a primeira classe petrográfica, G0, e a segunda, G1. Pode-se visualizar o comportamento das características e como elas se diferem entre os grupos, definindo as classes petrográficas. De acordo com a Tabela 3, para G0 DRHO tem valor normalizado entre aproximadamente -0.0228 e 0.0063, NFDC entre 0.0003 e 0.0005 e NPHI entre 0.0001 e 0.0015. Para G1 DRHO tem valor normalizado entre aproximadamente 0.0064 e 0.0252, NFDC entre 0.0004 e 0.0007 e NPHI entre 0.0001 e 0.0009. A Figura 4 exibe os grupos formados em relação as profundidades das entradas. Pode-se observar que a maioria das entradas de G0 estão entre 2740 e 2800 metros e de G1 entre 2790 e 2810 metros.

A Figura 5 exibe a distribuição das entradas nas duas primeiras componentes principais. A primeira expressa 74.01% de variabilidade dos dados, a segunda expressa 22.13% da variabilidade. As entradas que estão na cor vermelha são pertencentes a classe G0, e na cor azul são G1.

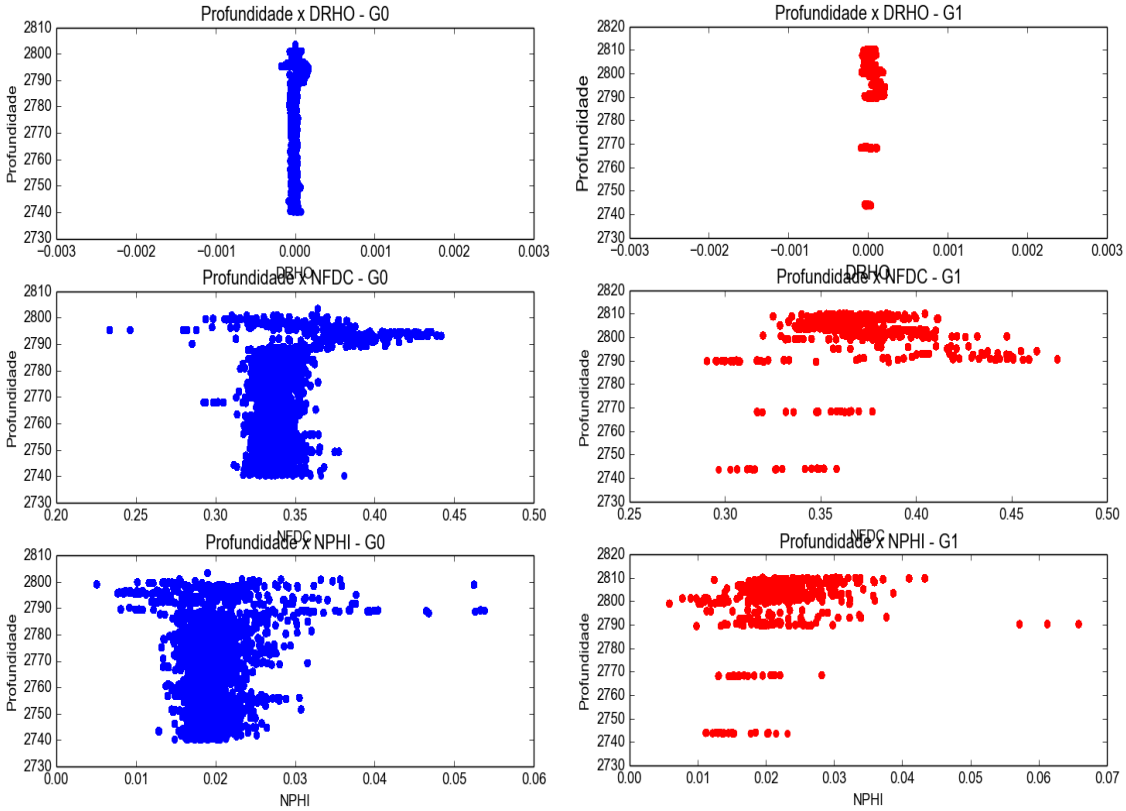

Figura 3: DRHO,NFDC e NPHI para cada grupo respectivamente.

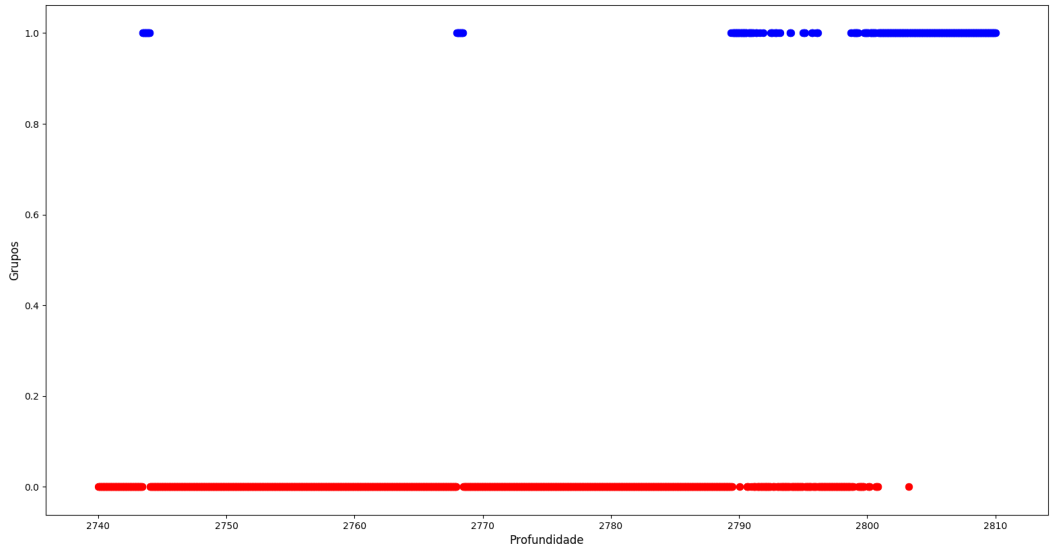

Figura 4: Grupos x Profundidade

Tabela 3: Grupos encontrados pelo K-Means. Para cada grupo tem-se a média de cada carcterística petrofísica e da profundidade. O número entre parênteses é o número de "amostras" atribuídas a cada grupo. O * indica as características que se diferenciaram entre os grupos.

Grupo	LogName	Média	Min	Max
G0(2063)	CALI	0.00043367	0.00041877	0.00045750
	CNST	0.00043529	0.00022435	0.00071220
	DRHO*	-0.00108014	-0.02284787	0.00628386
	FCNL	0.00043817	0.00022727	0.00089642
	FFDC	0.00043305	0.00022296	0.00084710
	GR	0.00042257	0.00024591	0.00080673
	NCNL	0.00043902	0.00030009	0.00073205
	NFDC*	0.00042334	0.00033336	0.00050843
	NPHI*	0.00043408	0.00010887	0.00148757
	RHOB	0.00043182	0.00030253	0.00051735
	DEPT	2772.77483	2740.01806	2809.99994
G1(234)	CALI	0.00045010	0.00042947	0.00048342
	CNST	0.00043580	0.00024015	0.00063496
	DRHO	0.01379635	0.00637737	0.02521027
	FCNL	0.00022727	0.00025388	0.00076064
	FFDC	0.00045560	0.00025347	0.00086969
	GR	0.00054799	0.00029905	0.00094482
	NCNL	0.00040296	0.00026328	0.00069490
	NFDC	0.00054123	0.00042412	0.00066522
	NPHI	0.00044649	0.00013159	0.00097717
	RHOB	0.00046646	0.00037439	0.00053589
	DEPT	2794.70565	2740.04848	2809.96946

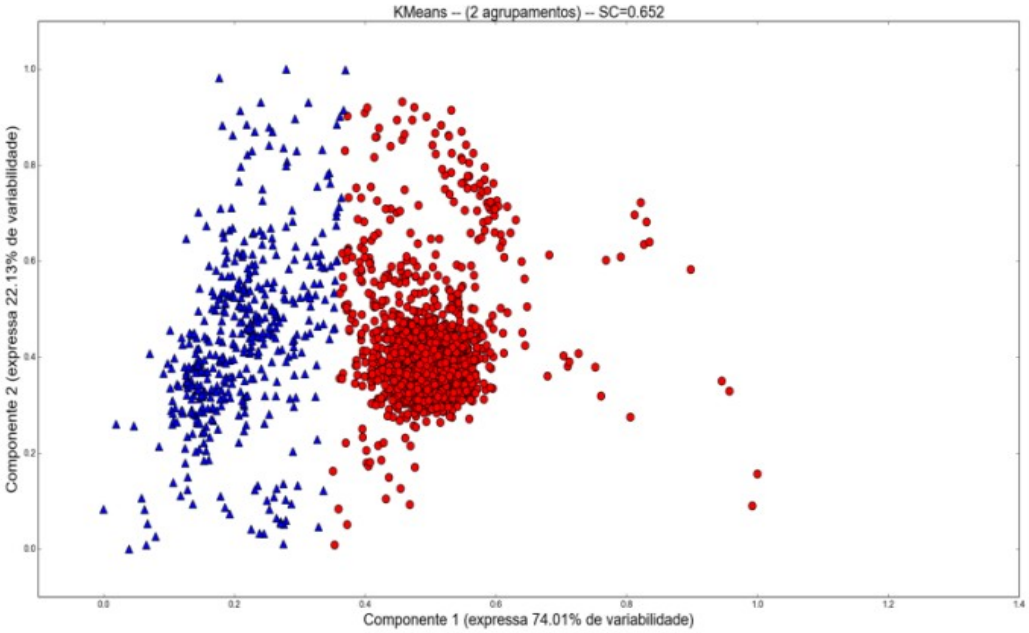

Figura 5: Resultado K-Means.

Conclusões

No presente trabalho foi apresentado um modelo computacional para a caracterização de litologia. A abordagem utilizada na determinação das possíveis classes petrográficas foi baseada em Análise de Agrupamento e para a visualização dos resultados foi empregada Análise de Componentes Principais. Duas classes petrográficas foram geradas onde foram diferenciadas pelas características DRHO, NFDC e NPHI, duas destas se mostraram influentes nas componentes principais. A abordagem proposta permitiu avaliar o desempenho da análise de agrupamentos na identificação de classes

petrográficas que representa a litologia do reservatório e dessa forma, surge como uma alternativa para assistir o geólogo/especialista na determinação e caracterização da litologia, contribuindo para a redução de custo e aumento na precisão dos resultados.

Agradecimentos

Os autores agradecem à Universidade Federal de Juiz de Fora (UFJF), ao Programa de Pós-Graduação em Modelagem Computacional (PPGMC), à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (Capes) pelo suporte dado e ANP por conceder acesso aos dados utilizados no estudo.

Referências Bibliográficas

AMEUR ZAIMECHE, O.; ZEDDOURI, A.; KOUADRIA, T.; KECHICHED, R.; Belaksier, M. S.. Use of Cluster Analysis method in log's data processing: prediction and rebuilding of lithologic facies. In: **International Conference on Environmental Science and Geoscience (ESG'14) Venice, Italy**. 2014. p. 98-101.

DONG, S.; WANG, Z.; ZENG, L.. Lithology identification using kernel Fisher discriminant analysis with well logs. **Journal of Petroleum Science and Engineering**, v. 143, p. 95-102, 2016.

HESLOP, K.; KARST, J.; PRENSKY, S.; SCHMITT, D.. Log Ascii Standard (LAS) Version 3.0. **The Log Analyst**, v. 40, n. 06, 1999.

METHE, P.; GOEPEL, A.; KUKOWSKI, N.. Testing the results of estimating lithological stratigraphy through cluster analysis on geophysical borehole logging data through Multi Sensor Core Logging data. In: **EGU General Assembly Conference Abstracts**. 2017. p. 9642.

PEDREGOSA, F.; VAROQUAUX, G.; GRAMFORT, A.; MICHEL, V.; THIRION, B.; GRISEL, O.; ... ; VANDERPLAS, J.. Scikit-learn: Machine learning in Python. **Journal of Machine Learning Research**, v. 12, n. Oct, p. 2825-2830, 2011.

ROUSSEEUW, P. J. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. **Journal of computational and applied mathematics**, v. 20, p. 53-65, 1987.

SAGGAF, M. M.; NEBRIJA, L. Estimation of lithologies and depositional facies from wire-line logs. **AAPG bulletin**, v. 84, n. 10, p. 1633-1646, 2000.

SEBTOSHEIKH, M. A.; SALEHI, A.. 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**, v. 134, p. 143-149, 2015.

VENDRAMIN, L.; CAMPELLO, R. J.; HRUSCHKA, E. R. Relative clustering validity criteria: A comparative overview. **Statistical analysis and data mining: the ASA data science journal**, v. 3, n. 4, p. 209-235, 2010.