Predição de propriedades petrofísicas em rochas reservatório da Bacia de Santos

Palavras-Chave: PETROFÍSICA, MACHINE LEARNING, PRÉ SAL

Lucca Munhoz dos Santos, UNICAMP

Prof. Dr. Gelvam André Hartmann, UNICAMP

INTRODUÇÃO

As reservas de hidrocarbonetos em rochas do Pré-sal brasileiro, descoberto em 2006, ocupam uma área de 149.000 km2 e com potencial para 100 bilhões de barris de óleo equivalente (Riccomini et al., 2012). Há, contudo, uma série de desafios a serem superados. As adversidades começam na localização do Pré-sal, que além de estar em águas profundas com lâminas d'água de 2.000 metros, ainda está sob 5.000 metros de rochas a serem perfuradas (Riccomini et al., 2012). Os desafios relevantes para este trabalho residem no fato de que os reservatórios visados não são “clássicos” como um arenito; a formação do Pré-sal é de um carbonato heterogêneo. Reservas carbonáticas tem uma estratigrafia complexa, pois são formadas por diferentes processos físicos, químicos e biológicos. Além disso, deformações tectônicas posteriores à formação dos reservatórios produzem um desafio adicional na identificação e caracterização desses reservatórios (Johann et al., 2012). Com essas dificuldades, as propriedades das rochas apresentam pouco contraste nos perfis de poços e nas técnicas de sísmica, dificultando a identificação das fácies geológicas e a extrapolação dessas informações para escalas de reservatório, não sendo possível manter a confiabilidade das escalas de poço (Silva et al., 2020).

A avaliação das propriedades petrofísicas dos reservatórios carbonáticos como porosidade e permeabilidade, de forma eficiente, é essencial para uma produção econômica e sustentável. Isso porque essas propriedades regulam as direções dos fluidos através dos poros. Assim, a determinação “confiável” dessas propriedades é fundamental para a exploração de reservatórios carbonáticos. No caso do Pré-sal é muito desafiador desenvolver uma fórmula generalizada de porosidade / permeabilidade, uma vez que se trata de um ambiente muito heterogêneo. Mesmo em laboratório é difícil desenvolver uma correlação entre uma variável física e um dado de poço, uma vez que parâmetros como fluidos de reservatório, mineralogia e invasão de fluido de perfuração, se mostram bastante distorcidos (Ahmadi et al., 2018). Abordagens empíricas de acordo com as estatísticas da relação entre variáveis eficazes e permeabilidade, são uma possibilidade simples e rápida de contornar esses desafios de determinação e predição de dados no Pré-sal.

A predição de propriedades petrofísicas como porosidade, saturação de água e litofácies são essenciais para a exploração e desenvolvimento de reservatórios de hidrocarbonetos (e.g. Jennings Jr. e Lucia, 2001). Essas estimativas são rotineiramente determinadas pela construção de modelos de física de rocha que ligam parâmetros elásticos obtidos a partir da inversão de dados sísmicos com as propriedades petrofísicas extraídas de perfis de poços ou de medições de laboratório de amostras de testemunho. No entanto, os dados sísmicos são geralmente a única fonte de informação disponível fora dos poços que podem ser usados para prever a distribuição 3D de propriedades físicas com resolução espacial apropriada. Além disso, as predições em condições mais profundas e não convencionais são difíceis de se determinar. A caracterização petrofísica de algumas rochas carbonáticas do pré-sal (como nos reservatórios do pré-sal brasileiro) se mostra muito difícil devido às suas heterogeneidades em várias escalas. Por exemplo, as propriedades petrofísicas em sistemas de poros de coquina podem resultar em uma ampla gama de texturas, tecidos e processos diagenéticos (Camara et al., 2014).

O objetivo principal deste trabalho é aplicar algoritmos preditivos que permitiram inferir propriedades petrofísicas, como porosidade e permeabilidade, em rochas carbonáticas da seção do Pré-sal, mesmo em situações em que só se tenha dados de perfis petrofísicos básicos, visando melhorar as determinações destas propriedades em escalas de reservatório. Com os dados de 50 poços de exploração petrolífera dos campos de Gato do Mato e de Sapinhoá, da Bacia de Santos, ensaiou-se diferentes técnicas de aprendizado de máquina para avaliar e predizer a distribuição de propriedades petrofísicas de forma intra e inter poços. Pretende-se com os resultados contribuir para minimizar custos e aumentar a confiabilidade na exploração de hidrocarbonetos no Pré-sal brasileiro.

BASE DE DADOS E MÉTODOS

Os algoritmos de Machine Learning (ML) podem ser compreendidos em três grandes grupos. A aprendizagem supervisionada usa dados com rótulos e o desempenho é medido pela capacidade de um modelo de prever rótulos com precisão. A aprendizagem não supervisionada usa dados sem rótulos e, portanto, o desempenho deve ser medido de forma diferente, como o agrupamento de rótulos. A última técnica de aprendizagem, principalmente associado ao aprendizado profundo (deep learning), é frequentemente referido como semi-supervisionado. O aprendizado semi-supervisionado usa dados com rótulos para produzir um modelo, mas limita a interação do usuário com alguns processos que operam em camadas ocultas, onde os recursos são comumente selecionados e manipulados por métodos não supervisionados (Dahl, 2008).

Neste trabalho são usadas técnicas de aprendizagem supervisionada. O princípio básico dessas técnicas é prever valores reais numéricos da variável dependente de uma ou mais variáveis independentes em um conjunto de dados. A forma mais básica de regressão é a regressão linear, que assume a forma: $Ax = b – c$, onde A é uma matriz de variáveis independentes, b é um vetor, x é um vetor de parâmetros e c é uma constante que satisfaz a equação quando $Ax = 0$ (Bhattacharya e Mishra, 2020). Para este caso é importante determinar essas funções em sua forma matricial, pois a multiplicação de matrizes reduz o tempo e o custo computacional das operações. Com relação à estatística, a regressão linear é simplesmente uma equação que estima o valor de uma variável y dados os valores de outras variáveis x, de forma tal que possa ser representada como uma função linear.

Os algoritmos de ML foram escritos no ambiente do Google Colab com apoio de bibliotecas em linguagem Python, open-source e de uso gratuito, como “Pandas” que fornece ferramentas para análise e manipulação de dados, “Pycaret” que permite que se faça todo o ciclo da criação de um modelo de Machine Learning, “Numpy” que suporta grandes cálculos multidimensionais, arranjos, matrizes e grande coleção de funções matemáticas e “Seaborn” que é uma biblioteca de visualização de dados do Python baseado em Matplotlib.

Foram selecionados 5 dos 50 poços iniciais para compor a base de dados dos algoritmos. Os 5 poços são do campo de produção Sapinhoá, localizado na porção central da Bacia de Santos, a aproximadamente 360 km da costa do estado de São Paulo, em lâmina d'água de 2.140 metros e com área de desenvolvimento de 233 km². Por conta da confidencialidade dos dados vamos chamar os poços de: “A”, “B”, “C”, “D” e “E”. Estes poços foram escolhidos porque apresentaram um conjunto promissor de perfis convencionais que são ideais para este tipo de aplicação e terem disponíveis seus perfis compostos para relacionar as curvas petrofísicas com a litologia descrita.

Para compor os dados foram selecionadas as seguintes curvas básicas para as variáveis independentes:

- Caliper, mede o diâmetro do poço de 6 a 16 polegadas.
- Raios Gama, mede a radioatividade natural da formação de 0 a 150 API .
- Resistividade, mede a resistência elétrica da formação à eletricidade de 0.2 a 2000 ohm.m.
- Fator Fotoelétrico, mede a ejeção de elétrons da camada que é exposta a radiação de 0 a 10 b/e.

As seguintes curvas alvo, que serão as variáveis dependentes, foram selecionadas:

● Porosidade Efetiva, mede a relação entre o volume total de espaços vazios interconectados e o volume total da rocha de 0 a 0.3 m³.
● Permeabilidade, pelo método Timur-Coates, mede a capacidade de um material transmitir fluido de 0.01 a 10000 mD.

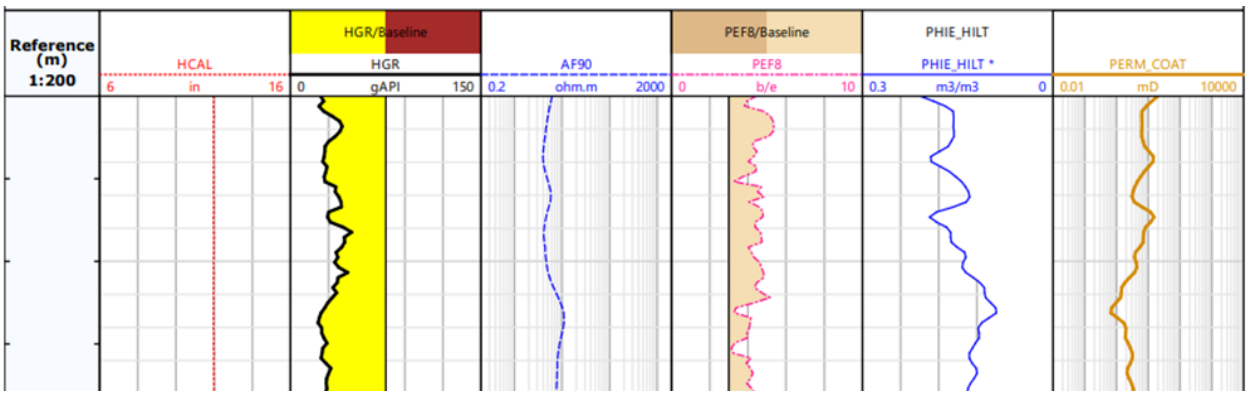

Figura 1: Curvas das variáveis dependentes e independentes no Software Techlog 2015.3

Esses dados junto com a profundidade do aparelho perfilador em relação à superfície, formam a base de dados das variáveis independentes que vão prever as curvas de porosidade efetiva e permeabilidade, nossas variáveis alvo, dependentes. Foram gerados quatro cenários e resultados principais, os dois primeiros são regressões intra poço e os dois últimos predições de propriedades interpoços:

● 1° cenário: Predição da porosidade usando os dados do poço A
● 2° cenário: Predição da permeabilidade usando os dados do poço A
● 3° cenário: Predição da porosidade usando os dados dos cinco poços
● 4° cenário: Predição da permeabilidade usando os dados dos cinco poços

Nos dois primeiros cenários os algoritmos serão treinados de forma aleatória com as variáveis independentes do poço A, e vão predizer as variáveis dependentes (porosidade e permeabilidade) do próprio poço A, de uma porção para teste cego que não foi usada para treinamento. Já nos dois últimos cenários isso se repete porém com os algoritmos de ML expostos aos dados de 5 poços, prevendo de forma aleatória propriedades alvo desses poços, simulando situações reais quando é necessário extrapolar informações de um poço para outro, somente com a disposição de curvas básicas de perfilagem.

A análise e tratamento das curvas petrofísicas foi feita dentro do Software Techlog 2015.3, e por fim foram exportadas como arquivo “CSV”, para o ambiente do Google Colab onde foi realizada um breve processamento de dados, além do já feito no Software Techlog 2015.3, para não deixar na tabela nenhum outlier ou valor em branco. Em seguida o banco de dados foi separado em “X” para as variáveis independentes e “Y” para as variáveis dependentes.

Ainda que se chama “Inteligência Artificial” essas técnicas de estatística aplicada, às máquinas ainda não são tão inteligentes assim; portanto, se colocarmos uma variável em escala diferente da outra, a máquina vai entender que uma variável é mais importante do que a outra, então nessa última etapa antes da aplicação dos algoritmos regressores foi necessário fazer o escalonamento da base de dados “X” e ”Y”.

RESULTADOS E DISCUSSÃO

A maior correlação entre as variáveis dependentes e independentes foi de 0.35, como pode ser observado na figura 2, no primeiro cenário, intra poço, mesmo assim o algoritmo conseguiu resultados expressivos em todos os cenários. A divisão das bases de dados intra poço ou interpoços foi feita 20% para teste e 80% para treinamento.

	PROFUNDIDADE	HCAL	HGR	PEF8	AF90	PHIE_HILT
PROFUNDIDADE	1.000000	0.599193	-0.525322	0.223691	-0.075314	0.127208
HCAL	0.599193	1.000000	-0.269650	0.177389	-0.091631	-0.001616
HGR	-0.525322	-0.269650	1.000000	-0.129085	0.106146	-0.254900
PEF8	0.223691	0.177389	-0.129085	1.000000	-0.118368	0.350840
AF90	-0.075314	-0.091631	0.106146	-0.118368	1.000000	-0.272587
PHIE_HILT	0.127208	-0.001616	-0.254900	0.350840	-0.272587	1.000000

Figura 2: Correlação Linear de Pearson das variáveis dependentes com a independente, no 1° cenário.

Os algoritmos foram importados da biblioteca do Python "Scikit-Learn" desenvolvida especificamente para aplicação prática de machine learning. Duas métricas foram usadas para avaliar o desempenho de cinco algoritmos de regressão, aplicados nos 4 cenários. A primeira métrica usada foi o "R^2" uma medida estatística do quão próximos os dados estão da linha de regressão ajustada. O "R^2" está sempre entre 0 e 1. 0 indica que o modelo não explica nada da variabilidade dos dados de resposta ao redor de sua média e 1 indica que o modelo explica toda a variabilidade dos dados de resposta ao redor de sua média. A segunda métrica usada foi o "MAE" (Mean Absolute Error), essa métrica calcula o valor da diferença entre o valor real e o previsto (resíduo) para cada um dos pontos e depois é tirada a média de todos esses resíduos. Essa medida apresenta valor mínimo 0 e não apresenta valor máximo. A figura 3 mostra uma tabela que compila todos esses dados para os quatro cenários.

	Predição da Porosidade			
	1° Cenário		3° Cenário	
	Intrapoço		Interpoços	
	R^2	MAE	R^2	MAE
Linear Regression	0,16	0,67	0,19	0,72
Decision Tree	0,77	0,31	0,75	0,34
Random Forest	0,86	0,25	0,87	0,25
Support Vectors	0,56	0,50	0,64	0,44
Neural Networks	0,75	0,38	0,76	0,37

	Predição da Permeabilidade			
	2° Cenário		4° Cenário	
	Intrapoço		Interpoços	
	R^2	MAE	R^2	MAE
Linear Regression	0,10	0,57	0,097	0.69
Decision Tree	0,68	0,24	0,74	0,30
Random Forest	0,87	0,20	0,87	0,23
Support Vectors	0,36	0,42	0,46	0,45
Neural Networks	0,63	0,34	0,66	0,38

Figura 3: Comparação entre as métricas dos cinco regressores nos quatro cenários.

Podemos observar sem a necessidade de ferramentas estatísticas que o algoritmo Random Forest (RF), tem superioridade tanto na métrica "R^2" quanto na "MAE" em relação a todos os outros algoritmos, chegando a 87% de precisão, o que corrobora com os resultados do trabalho "The linear random forest algorithm and its advantages in machine learning assisted logging regression modeling" publicado no Journal of Petroleum Science and Engineering, em 2019, por Yile Ao, Hongqi Li, Liping Zhu, Sikandar Ali e Zhongguo Yang.

Ainda que na pesquisa eles não tenham usado somente curvas básicas nem feito análises interpoços como no presente trabalho, e o algoritmo RF tenha algumas diferenças na construção das árvores, também obtiveram resultados superiores com o algoritmo RF quando fizeram regressões para curvas de perfilagem de poços, superando até mesmo redes neurais complexas.

RF é um algoritmo que aplica a técnica geral de agregação de bootstrap, entre suas Decision Tree (DT). Dado um conjunto de treinamento (X = x1,....,xn) com respostas (Y = y1,....,yn), é feito o ensacamento repetidamente (B vezes) e selecionada uma amostra aleatória com substituição do conjunto de treinamento e as

DT são ajustadas a estas amostras, para (b = 1,...,B). Após o treinamento, as previsões podem ser feitas calculando a média das previsões de todas as DT de regressão individuais ou tomando a maioria dos votos no caso de DT de classificação.

Para ter resultados estatisticamente robustos, foi aplicada a técnica de validação cruzada no algoritmo RF, nos 4 cenários. Foram feitos 30 testes de validação cruzada, com 10 "k" separações para cada validação, gerando no total 300 testes com diferentes bases de treinamento e testes para cada um dos cenários, como mostra a figura 4.

	Predição de Porosidade	
	Intrapoço	Interpoços
	1° Cenário	3° Cenário
N° de testes	300	300
Média	0,8705	0,8731
Desvio Padrão	0,0285	0,0145
Mínimo	0,7623	0,8326
25%	0,8534	0,8646
50%	0,8733	0,8733
75%	0,8905	0,8840
Máximo	0,9438	0,9068

	Predição de Permeabilidade	
	Intrapoço	Interpoços
	2° Cenário	4° Cenário
N° de testes	300	300
Média	0,8661	0,8601
Desvio Padrão	0,0507	0,0240
Mínimo	0,6560	0,7805
25%	0,8356	0,8462
50%	0,8733	0,8631
75%	0,9045	0,8775
Máximo	0,9489	0,9076

Figura 4: Comparação das métricas das validações cruzadas, com o algoritmo RF nos quatro cenários.

Ainda que o valor máximo de precisão dentre os 1200 testes seja no cenário 2, a maior média de precisão para o algoritmo RF é no 3° cenário, onde é feita a predição da curva de porosidade efetiva tendo como base de dados 5 poços diferentes. Além disso, é no 3° cenário que temos o menor desvio padrão mostrando consistência nas predições. Analisando essas métricas podemos afirmar que quando exposto a mais dados e dados interpoços o algoritmo aprende melhor que na análise intra poço. Pretende-se fazer ainda testes de hipóteses para provar a superioridade estatística da predição interpoços, em relação a predição intra poço, além de novas configurações de base de teste, treino e parâmetros, para comprovar a robustez do algoritmo se utilizado para prever curvas alvo quando somente se tem curvas básicas nos dados, contribuindo para uma exploração mais eficaz das reservas brasileiras de hidrocarboneto.

BIBLIOGRAFIA

HONGQI, L.; LIPING ZHU.; SIKANDAR, A.; YILE, A.; ZHONGGUO, Y. **The linear random forest algorithm and its advantages in machine learning assisted logging regression modeling**. Journal of Petroleum Science and Engineering, 2019. https://doi.org/10.1016/j.petrol.2018.11.067.

AHMADI, M, A.; CHEN, Z. **Comparison of machine learning methods for estimating permeability and porosity of oil reservoirs via petro-physical logs.** Petroleum 2018. https://doi.org/10.1016/j.petlm.2018.06.002.

BHATTACHARYA, S.; MISHRA, S. 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, doi: 10.1016/j.petrol.2018.06.075.

JOHANN, P.; MARTINI, A.; MAUL, A.; NUNES, J, P. **Reservoir Geophysics in Brazilian Pre-Salt Oilfields.** Offshore Technology Conference. Houston, Texas, USA, OTC 23681. 30 April--3 May 2012.

CAMARA, R.; CORBETT, P, W, M.; TAVARES, C.; MACHADO, A.; JESUS, E.; BORGHI, L. (2014). **Carbonate coquina reservoirs: New insights for petrophysical reservoir characterisation.** p. 76th EAGE Conference and Exhibition (June), 5 p.

DAHL, M. **A Gentle Introduction to Machine Learning.** Focus Article. Nanaimo, B.C. VOL 43, NO 01. 2018.

RICCOMINI, C., SANT´ANNA, L. G., TASSINARI, C. C. G. (2012). **Pré-sal: geologia e exploração.** Revista USP, (95), 33-42. https://doi.org/10.11606/issn.2316-9036.v0i95p33-42

SILVA, A, A.; TAVARES, W, M.; CARRASQUILLA, A.; MISSÁGIA, R.; CEIA, M. **Petrofacies classification using machine learning algorithms.** Society of Exploration Geophysicists, São Paulo, vol. 85, N°. 4 (julho-agosto de 2020); P.101-113.