

Previsão de Taxa de Perfuração em Poços de Petróleo Offshore Utilizando Aprendizado de Máquina

André Branco¹, Janaína Gomide¹

¹Instituto Politécnico - Universidade Federal do Rio de Janeiro (UFRJ)

abrancomrafaels@gmail.com, janainagomide@gmail.com

Abstract. *The rate of perforation (ROP) of an oil well is a very important metric to control, as it affects well productivity, bit wear, and well and operational security. This work evaluates the use of Machine Learning algorithms for predicting ROP in offshore oil wells as an alternative for other models traditionally used by the industry. Random forest models were evaluated for different hyperparameters for 4 different offshore oilwells in Santos Basin's pre-salt. In general, random forest models performed the best for each well, but there was no optimal combination of hyperparameters between all wells.*

Resumo. *A taxa de perfuração (ROP) de um poço de petróleo é uma métrica muito importante de se controlar, por influenciar na produtividade do poço, no desgaste da broca e na segurança do poço e operação. Este trabalho avalia a utilização do algoritmo de floresta aleatória para previsão da ROP em poços de petróleo offshore como alternativa para modelos tradicionalmente utilizados pela indústria. Modelos de floresta aleatória foram avaliados com diferentes hiper-parâmetros para 4 poços offshore no pré-sal da Bacia de Santos. De forma geral, os modelos de floresta aleatória apresentaram resultados superiores para cada poço, mas não houve uma combinação de hiper-parâmetros ótima comum entre os poços.*

1. Introdução

O mercado *upstream* de petróleo lida com as atividades de prospecção, perfuração e produção de petróleo. A prospecção pode ser definida como o conjunto de atividades para definir as áreas de uma bacia sedimentar em que é possível o acúmulo de petróleo, e em quais destas áreas há maior probabilidade de se encontrar petróleo, além de conferir informações importantes para as etapas seguintes e avaliar a viabilidade econômica de programas de exploração. A perfuração é uma etapa em que uma sonda de petróleo utiliza a rotação de uma broca para perfurar um poço que alcance o reservatório de petróleo, permitindo a próxima etapa. Na produção, o petróleo é extraído de seu reservatório para a superfície, onde ele é armazenado e transportado para etapas posteriores na cadeia de valor do petróleo [Thomas 2001].

A taxa de perfuração, ou *rate of perforation* (ROP) é uma variável muito importante no processo de perfuração, e representa a taxa de variação da profundidade do poço em relação ao tempo, calculada durante atividades de perfuração. O valor da ROP influencia diretamente no desgaste das brocas, na segurança do poço e da operação, bem como na eficiência geral do processo, e por isso deve ser controlada. Para controlar a ROP de forma efetiva, é necessário ser capaz de estimar seu valor para diferentes valores de entrada, e por isso o interesse em modelar seu comportamento [GANDELMAN 2012].

Modelos experimentais para a ROP existem e são largamente utilizados pela indústria, mas sua acurácia e capacidade de generalização para condições de perfuração diferentes são limitadas [GANDELMAN 2012]. Trabalhos acadêmicos recentes utilizam técnicas de aprendizado de máquina para obter modelos e reportam resultados significativamente melhores que os modelos tradicionais [Barbosa et al. 2019].

Uma revisão de literatura sobre a previsão de ROP foi realizada em [Barbosa et al. 2019]. Nesse artigo os autores apresentaram abordagens que utilizam técnicas de aprendizado de máquina e em como estes modelos podem ser usados para otimizar as atividades de perfuração. Nesse trabalho, os autores reportaram um aumento expressivo de artigos sobre uso destas técnicas para a tarefa de prever ROP. A maior parte dos trabalhos revisados pelos autores utilizam redes neurais artificiais de variadas arquiteturas para realizar a previsão da ROP, como em [Shi et al. 2016], mas uma outra abordagem bastante utilizada é a de métodos de *ensemble*, como a floresta aleatória (*random forest*). Em [Hegde et al. 2017] os autores utilizam o algoritmo de floresta aleatória para prever a ROP e os resultados obtidos foram posteriormente utilizados para escolha automática de parâmetros de perfuração ótimos em [Hegde and Gray 2017] e [Hegde et al. 2020].

Este trabalho tem como objetivo aplicar técnicas de aprendizado de máquina para a previsão da taxa de perfuração (ROP) durante a perfuração de 4 poços de petróleo *offshore* de pré-sal da Bacia de Santos. O algoritmo utilizado foi a floresta aleatória, e o desempenho é avaliado pelas métricas MAE normalizado pela média e R^2 .

2. Perfuração de Poços de Petróleo

A perfuração de um poço de petróleo é realizada através de uma sonda de perfuração, que no caso de poços *offshore* de águas profundas ou ultraprofundas são plataformas semissubmersíveis ou navios-sonda [Morais 2013]. As sondas possuem vários sistemas: conjuntos de equipamentos responsáveis pela correta operação das sondas para a perfuração, que podem estar direta ou indiretamente relacionados à atividade de perfuração [Thomas 2001].

Durante a perfuração, as rochas são perfuradas para a formação do poço por meio da aplicação de peso e rotação sobre a broca enquanto esta encontra-se no fundo do poço, realizando a ruptura e desagregação das rochas. Simultaneamente, fluido de perfuração é injetado na coluna de perfuração, atravessando seu interior tubular e saindo pela broca, o que auxilia na perfuração e limpa o fundo do poço, transportando o cascalho gerado na perfuração para a superfície. Este processo causa o aumento progressivo da profundidade, até que seja finalizada uma sessão do poço. A perfuração é feita em sessões progressivas, perfuradas com brocas de diâmetro progressivamente menores, até que a profundidade final seja alcançada [Thomas 2001].

A taxa de perfuração é a velocidade com que a broca perfura as rochas em um poço de petróleo, e é um valor muito importante de se controlar. [GANDELMAN 2012] discute, em sua dissertação, os parâmetros que mais influenciam ROP. São eles: o peso sobre broca (WOB), o diâmetro e tipo da broca, o diferencial de pressão poço-formação (*overbalance*), a rotação da broca (RPM), a vazão do fluido de perfuração, a profundidade vertical do poço, a litologia da formação, o tempo de trânsito (de onda sonora) e a composição da coluna de perfuração. Dentre esses, o WOB e o RPM são os parâmetros

Tabela 1. Descrição dos conjuntos de dados

	Nº de Linhas	Profundidade
Poço 1	761467	Lâmina d'água 1855 m e dados 3160–6016 m
Poço 2	710753	Lâmina d'água 2001 m e dados 3040–6063 m
Poço 3	365857	Lâmina d'água 2139 m e dados 5323–5600 m
Poço 4	252955	Lâmina d'água 2120 m e dados 5376–5557 m

mais facilmente ajustáveis para controlar a ROP, enquanto os outros parâmetros são escolhidos baseados em outros requisitos do projeto do poço, como no caso da vazão e diferencial de pressão (que são escolhidos baseado na limpeza e estabilidade de poço), ou estão completamente fora do controle dos operadores, como no caso da litologia.

Há na literatura modelos tradicionais utilizados para a previsão de ROP: equações matemáticas que buscam modelar o comportamento físico das variáveis de perfuração. Uma revisão de diferentes métodos podem ser vistos em [Barbosa et al. 2019]. Entretanto, como afirma [GANDELMAN 2012], a eficiência destes modelos é consideravelmente limitada e não consegue generalizar bem para regiões e condições diferentes daquelas para que foram desenvolvidos. Nesse trabalho é apresentada uma metodologia baseada em aprendizado de máquina para abordar esse problema como uma alternativa aos métodos tradicionais.

3. Metodologia

A função a ser estimada pelo algoritmo de aprendizado é a taxa de perfuração de poços de petróleo *offshore* (ROP). As etapas da metodologia utilizada são descritas a seguir.

3.1. Coleta de dados

Os dados utilizados nos experimentos numéricos foram cedidos por uma empresa do setor de óleo e gás *upstream offshore* atuante no Brasil e exterior. Os dados foram disponibilizados para 4 poços diferentes da Bacia de pré-sal de Santos. Informações sobre os poços podem ser encontrados na Tabela 1. Os poços 1 e 2 são ambos poços do campo de Búzios e os poços 3 e 4 são do campo de Sépia.

As entradas selecionadas para o estudo foram escolhidas baseado em características utilizadas em trabalhos relacionados, na opinião de especialistas da área para características que afetariam os dados, e por fim baseado na disponibilidade de cada uma delas. As características selecionadas foram: profundidade do poço (m), profundidade da broca (m), peso sobre broca (klbf), torque do *top drive* (klbf-ft), velocidade de rotação do *top drive* (rpm), vazão de entrada de fluido de perfuração (gal/min), vazão de saída de fluido de perfuração (gal/min), pressão de *standpipe* (psi) e inclinação do poço (°).

As oito primeiras características acima são dados em tempo real coletados por sensores em equipamentos na superfície da sonda e armazenados com uma frequência de 1 dado por segundo. Já a inclinação do poço foi disponibilizada para uma quantidade discreta e comparativamente baixa de valores de profundidade; estes valores são medidos após a perfuração da determinada profundidade e disponibilizados a intervalos de tempo variáveis.

Tabela 2. Amplitude de dados em tempo real

Característica	Amplitude			
	Poço 1	Poço 2	Poço 3	Poço 4
Profundidade de poço (m)	3160,3–6016,6	3039,7–6063,4	5323,4–5600,2	5376,8–5556,8
Profundidade da broca (m)	3159,3–6016,6	3038,3–6063,4	5322,5–5600,2	5373,0–5556,7
Peso sobre broca (klbf)	-79,7–930,7	-151,4–456,6	-55,9–519,3	-82,1–476,4
Torque (klbf · ft)	-22,1–55,4	-15,7–67,4	0,0–50,0	0,0–15603,4
Velocidade de rotação (rpm)	0,0–239,1	0,0–241,3	0,0–80,0	0,0–81,1
Vazão de entrada de fluido (gal/min)	0,0–5483,4	0,0–9298,6	0,0–558,7	0,0–675,4
Vazão de saída de fluido (gal/min)	-0,3–828,4	0,0–601,2	0,0–292,8	0,0–365,7
Pressão de <i>standpipe</i> (psi)	18,8–4438,2	0,0–5859,5	15,0–4127,2	15,0–5260,9

Os dados considerados não incluem informações sobre as propriedades físicas das formações perfuradas, do fluido de perfuração utilizado ou de geometria e tipo de broca utilizados. Estes dados, apesar de recomendados por especialistas e serem utilizados em outros trabalhos por influenciarem consideravelmente a ROP, não estavam disponíveis: alguns por não serem coletados pela empresa e outros por não serem cedidos por ela.

Como a função alvo é a ROP, os dados desejados limitam-se àqueles durante a atividade de perfuração, desconsiderando, por exemplo, atividades de revestimento e repasse, que, embora vitais para o processo de perfuração, são desconsideradas para a estimativa da ROP. Para isso, durante a coleta de dados, utilizou-se os relatórios diários de perfuração, que indicam, entre outras informações, os períodos em que ocorreu perfuração, com precisão máxima de meia hora.

3.2. Preparação dos Dados

Para realizar a preparação dos dados, bem como para implementar as etapas seguintes, utilizou-se a linguagem de programação Python, versão 3.8.6 com as bibliotecas `numpy`¹ e `pandas`² para leitura, análise e tratamento dos dados, e as bibliotecas `matplotlib`³ e `plotly`⁴ para geração de gráficos de diferentes tipos.

3.2.1. Análise Exploratória dos Dados

Uma análise inicial de estatísticas dos dados em tempo real revelou o número de linhas por poço, e também que, para cada uma das linhas, não havia colunas com dados faltantes. As amplitudes de cada característica divididas por poço estão resumidas na Tabela 2.

Observando inicialmente os dados medidos a cada segundo, notou-se que, para alguns poços, ocorria falhas ocasionais em que o tempo entre dois dados subsequentes era maior que 1 segundo, sem considerar os casos em que a perfuração foi interrompida para realizar outras atividades. Este é o primeiro problema identificado nos dados (**P1**).

Analisando estes dados em gráficos de formato de série temporal, notou-se alguns outros problemas nos dados:

P2 - algumas porções dos dados eram de momentos em que não se estava perfurando;

¹<https://numpy.org/>

²<https://pandas.pydata.org/>

³<http://matplotlib.org/>

⁴<https://plotly.com/python/>

- P3** - Erros de medição ou transmissão em alguns dos valores para vazão de entrada de fluido de perfuração, identificados por estarem muito acima de valores nominais para os equipamentos e consistirem em menos de 0,1% dos dados;
- P4** - Valores de profundidade que permaneceram constantes quando deviam aumentar;
- P5** - Instâncias de dados em que houve redução na profundidade do poço, que se consideradas poderiam ter grande efeito na ROP prevista pelos modelos;
- P6** - Lacunas de múltiplos metros perfurados em alguns dos poços.

A solução destes problemas será apresentada nas seções seguintes.

Quanto aos dados de inclinação que é medida graus em determinadas profundidades, pela quantidade baixa de dados, foi feita observando diretamente a tabela de valores. Não havia dados faltantes nas colunas, mas nem sempre os dados de inclinação estavam disponíveis para toda a amplitude de profundidade dos poços. Adicionalmente, no poço 3 foi identificado um caso em que a inclinação estava incorreta, porém ocorreu em profundidades abaixo de onde havia dados em tempo real para o poço, e portanto não afetaria o resultado.

3.2.2. Correção de problemas nos dados

Para resolver o problema **P2**, foi feito, inicialmente, uma identificação e limpeza manual dos dados; em seguida, filtrou-se os dados, eliminando as linhas em que a diferença entre a profundidade da broca e do poço fossem maiores que um metro (para considerar o efeito de *bit bounce* [Al Sairafi et al. 2016]), e em que a vazão de fluido de perfuração fosse menor que 120 gal/min (para eliminar operações de conexão).

Para resolver o problema **P3** aplicou-se um segundo filtro para os dados, excluindo aqueles em que o vazão de entrada do fluido de perfuração é maior que 2000 gal/min e o torque do *top drive* é maior que 100 klbf-ft. Os valores foram selecionados de forma a englobar com boa margem de segurança a amplitude de operação destas variáveis.

Para resolver o problema **P4**, a porção em que o sinal de profundidade parou de ser atualizado, primeiro identificou-se o intervalo de tempo em que a profundidade não se atualizava, e posteriormente preencheu-se este intervalo com a regressão linear feito com os valores antes e depois do congelamento.

Para resolver o problema **P5**, os dados em que a profundidade do poço reduziu foram eliminados após o cálculo da "ROP instantânea" de cada exemplo, calculada a partir da diferença entre as profundidades atual e do exemplo anterior: caso a "ROP instantânea" seja negativa, o exemplo é desconsiderado.

Por fim, os problemas das lacunas dos dados **P1 e P6** foram resolvidos de maneira automática ao agregar os dados na profundidade, como será explicado na Subseção 3.2.4.

3.2.3. Engenharia de Características

Na engenharia de características, foram criadas duas novas características: proximidade da broca e potência. A proximidade da broca foi obtida calculando a diferença entre a profundidade do poço e a profundidade da broca, e foi criada porque é a relação entre

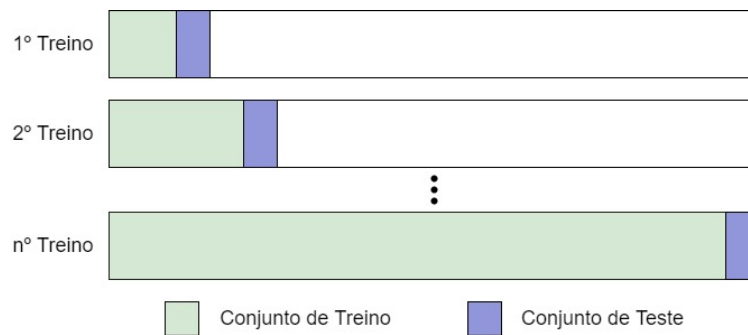


Figura 1. Divisão de Treino e Teste

a profundidade do poço e a profundidade da broca que afeta o processo da ROP, e não o valor isolado da profundidade da broca. Já a potência foi obtida pelo produto entre a velocidade de rotação e o torque; este valor foi criado com o objetivo de modelar a energia utilizada para perfurar a rocha. Após a criação das novas características, eliminou-se a profundidade da broca por considerar-se que seria um valor de pouca relevância para o modelo.

3.2.4. Transformação dos Dados

Como a inclinação é medida para apenas um pequeno conjunto de profundidades específicas, ela foi estendida para as profundidades em que não há valor especificado. Essa extensão foi feita propagando a inclinação para as profundidades superiores, i.e., utilizando o valor da última inclinação medida para valores de profundidades maiores. Caso não haja inclinações anteriores, considerou-se que seu valor é 0.

Por fim, os dados foram agregados pela profundidade. Para isso, definiu-se faixas de 0,08 metros de profundidade, agregando todos os dados em cada faixa de profundidade pela média. O valor de 0,08 metros foi selecionado por ser aproximadamente igual a 0,25 pés, valor utilizado no trabalho de [Hegde et al. 2017]. Neste processo de agregação de dados, o problema **P6** das lacunas de dados é resolvido.

3.2.5. Divisão de Treino e Teste

Como a perfuração é um processo sequencial, em que não é possível perfurar uma determinada profundidade antes de perfurar as profundidades anteriores, e as propriedades das rochas perfuradas variam com a profundidade, é necessário tomar cuidado para não treinar com dados posteriores aos dados de teste, pois estes dados não estariam disponíveis para treino antes de serem utilizados para previsão.

Assim, utilizou-se o esquema de treino e teste representado na Figura 1, em que define-se uma porção dos dados para treino, e testa com os dados que os sucedem. Em seguida, aumenta-se o conjunto de treino, e testa com os dados seguintes, repetindo este processo até não haver mais dados. Considerou-se inicialmente o conjunto dos 360 primeiros dados do poço para treino, e os 180 seguintes para teste do modelo treinado. Para cada novo treino, o conjunto de treino foi aumentado em 360.

Tabela 3. Número de treinos realizados por poço

	Poço 1	Poço 2	Poço 3	Poço 4
Quantidade de Treinos	89	95	9	5

Os dados dos diferentes poços estão disponíveis para amplitudes de profundidade diferentes, conforme Tabela 1, então o número de treinos e testes realizados irá variar dependendo do poço. Esta informação está resumida na Tabela 3.

Como o processo de perfuração é altamente não linear [Barbosa et al. 2019] e o valor da ROP por profundidade é errático, realizou-se uma remoção de *outliers* do conjunto de treino. De todos os dados de treino, foram removidos os 0,5% de maior valor de ROP do conjunto de treinos. Esta remoção precisou ser feita após a divisão dos conjuntos de treino e teste pois ela é baseada em uma estatística calculada sobre o conjunto.

3.3. Experimentos para cada poço

Os experimentos numéricos a seguir foram feitos utilizando a biblioteca *scikit-learn*⁵ versão 0.24.1 do Python para aplicação do algoritmo de aprendizado da floresta aleatória. As estatísticas foram calculadas utilizando funções da biblioteca *pandas* versão 1.1.4.

Após a divisão dos conjuntos de treino e teste, definida na Subseção 3.2.5, realizou-se os experimentos numéricos, em que treinou-se modelos de floresta aleatória com diferentes hiper-parâmetros em cada conjunto de treino de cada poço, e em seguida testou-se os modelos em cada conjunto de teste de cada poço. A quantidade de pares de conjunto de treino e teste foi mostrada na Tabela 3.

No treino da floresta aleatória, variou-se os hiper-parâmetros de profundidade máxima de árvore e de número de árvores (estimadores) do comitê. Para a profundidade foram utilizados os valores: 3, 6, 8 e 10, e para o número de árvores foram utilizados os valores 100, 200 e 400.

3.4. Importância das Características

Para medir quanto cada característica influencia na previsão final da taxa de perfuração para cada poço, utilizou-se a medida de importância das características do algoritmo de floresta aleatória. As combinações de algoritmo e hiper-parâmetros utilizados para medir a importância das características foram aquelas que tiveram melhores resultados para cada poço. Estas combinações de algoritmo e hiper-parâmetros foram usadas para treinar um modelo usando todos os dados do poço, para cada poço, e este modelo foi usado na medida da importância das características desse poço.

3.5. Aprendizado por Transferência

O melhor modelo obtido nos experimentos anteriores é escolhido para testar a possibilidade de aprendizado por transferência. Para este experimento numérico, foram escolhidos os poços 1 e 2 por ambos serem do campo de Búzios, e portanto possuem características litológicas semelhantes, e por possuírem dados para uma maior extensão de poço que os poços 3 e 4.

⁵<https://scikit-learn.org/>

Os experimentos numéricos de aprendizado por transferência foram feitos tanto treinando no poço 1 e testando no poço 2 quanto treinando no poço 2 e testando no poço 1. Foram testadas duas situações: treinar um modelo todos os dados do primeiro poço e usar somente este modelo para prever no segundo poço, o que será chamado de modelo simples; e treinar um modelo com todos os dados do primeiro poço e um segundo modelo com os dados do segundo poço, utilizando o esquema de treino-teste da Figura 1, e usando a média de suas previsões como previsão final, o que será chamado de modelo composto. Para fins de comparação, mesmo que o experimento numérico com o modelo simples pudesse ser usado para prever para todos os exemplos do segundo poço, o teste será feito com os conjuntos de teste utilizado para os outros experimentos numéricos, de modo a coincidir com o experimento com o modelo composto.

4. Resultados e Discussão

Nessa seção são apresentados os resultados e discussões sobre os experimentos numéricos nos 4 poços *offshore*. Por fim, avaliou-se quais são as características mais relevantes para a previsão da ROP e a possibilidade de aprendizado por transferência entre poços usando os poços 1 e 2.

4.1. Experimentos Numéricos do Poço 1

A Tabela 4 indica que os modelos com profundidade máxima igual 8 ou 10 têm melhor resultado, sem grandes diferenças entre eles ou para diferentes números de estimadores com uma pequena melhora com o aumento de estimadores e da profundidade máxima. Para o R^2 , o resultado é análogo. Um ponto importante é o valor mínimo de R^2 para cada um dos conjuntos de hiper-parâmetros, que é negativo; este valor negativo ocorre somente em um dos pares de conjunto treino-teste, perto do final do poço, e todos os outros valores de R^2 são positivos. Como o valor de R^2 é limitado superiormente por 1, mas não possui limite inferior, valores negativos afetam desproporcionalmente a média. Entretanto, como é somente um valor entre 89, seu efeito na média é relativamente baixo.

Tabela 4. Resultados da floresta aleatória — poço 1

Prof.	# Arv.	Erro Normalizado				R^2			
		méd	dev	mín	máx	méd	dev	mín	máx
3	100	0.587	0.342	0.359	3.142	0.494	0.244	-1.279	0.781
	200	0.587	0.342	0.358	3.138	0.493	0.248	-1.316	0.781
	400	0.587	0.342	0.358	3.135	0.493	0.247	-1.295	0.781
6	100	0.518	0.203	0.324	1.975	0.539	0.227	-1.066	0.833
	200	0.517	0.201	0.326	1.948	0.539	0.226	-1.042	0.830
	400	0.517	0.201	0.326	1.953	0.539	0.224	-1.016	0.830
8	100	0.506	0.190	0.305	1.906	0.551	0.203	-0.680	0.839
	200	0.505	0.189	0.308	1.882	0.551	0.204	-0.647	0.839
	400	0.504	0.187	0.308	1.862	0.553	0.201	-0.601	0.840
10	100	0.505	0.184	0.296	1.841	0.551	0.196	-0.351	0.846
	200	0.503	0.183	0.297	1.841	0.552	0.201	-0.340	0.846
	400	0.502	0.182	0.297	1.830	0.553	0.200	-0.329	0.849

4.2. Experimentos Numéricos do Poço 2

Observa-se pela Tabela 5 que os modelos com profundidade máxima 6 têm menor erro normalizado, e enquanto apresentam menor R^2 que os modelos com profundidade máxima 3, a diferença não é significativa, o que não pode ser dito sobre a diferença entre os erros normalizados dos modelos. Com isso, os modelos com profundidade máxima 6 são preferíveis. A variação no número de árvores utilizados para estimar não altera significativamente o desempenho do modelo. Novamente, o valor mínimo de R^2 negativo ocorre somente em um par treino-teste, perto do final do poço, e tem uma influência desproporcional, mas pouco significativa, no valor médio.

Tabela 5. Resultados da floresta aleatória — poço 2

Prof.	# Arv.	Erro Normalizado				R^2			
		méd	dev	mín	máx	méd	dev	mín	máx
3	100	0.546	0.140	0.395	1.452	0.426	0.181	-0.197	0.788
	200	0.546	0.141	0.395	1.452	0.426	0.181	-0.206	0.788
	400	0.546	0.140	0.394	1.447	0.427	0.180	-0.187	0.787
6	100	0.528	0.109	0.385	1.213	0.424	0.219	-0.315	0.840
	200	0.528	0.107	0.384	1.192	0.424	0.216	-0.249	0.845
	400	0.529	0.110	0.385	1.219	0.423	0.218	-0.255	0.849
8	100	0.538	0.128	0.378	1.355	0.409	0.248	-0.469	0.809
	200	0.538	0.129	0.380	1.367	0.408	0.250	-0.470	0.825
	400	0.538	0.129	0.378	1.356	0.406	0.256	-0.466	0.826
10	100	0.551	0.157	0.379	1.650	0.386	0.289	-0.616	0.816
	200	0.552	0.163	0.382	1.704	0.385	0.295	-0.645	0.816
	400	0.552	0.161	0.378	1.672	0.383	0.298	-0.751	0.815

4.3. Experimentos Numéricos do Poço 3

A Tabela 6 revela um melhor desempenho em termos de erro normalizado para os modelos com profundidade máxima 3, e quanto ao R^2 , observa-se que, em média, os modelos de profundidade máxima 3 são melhores. Em ambos os casos, não há diferenças significativas quando varia-se o número de árvores da floresta. Neste poço, observa-se um valor mínimo de R^2 bastante negativo e, como nos outros poços, ocorre somente uma vez perto do final do poço. Entretanto, como a magnitude do valor negativo é muito grande, e o poço 3 possui somente 9 pares de conjunto treino-teste, o efeito do valor negativo é grande na média, que passa a não ser representar bem o comportamento geral dos dados; neste caso, a mediana seria mais representativa.

4.4. Experimentos Numéricos do Poço 4

Conforme apontado pela Tabela 7, os melhores modelos são aqueles com profundidade máxima igual a 3, tanto em R^2 quanto em erro normalizado, com pouca influência do número de árvores no desempenho. Diferentemente dos outros poços, os valores de R^2 são todos positivos, o que facilita a análise. Um possível motivo para essa diferença no valores de R^2 é que para o poço 4 havia somente dados do final do poço, provavelmente de uma única litologia, e portanto de comportamentos semelhantes entre si.

Tabela 6. Resultados da floresta aleatória — poço 3

Prof.	# Arv.	Erro Normalizado				R ²			
		méd	dev	mín	máx	méd	dev	mín	máx
3	100	0.810	0.412	0.449	1.706	-0.278	2.382	-6.592	0.762
	200	0.819	0.417	0.450	1.708	-0.291	2.396	-6.639	0.762
	400	0.812	0.411	0.447	1.672	-0.269	2.327	-6.429	0.767
6	100	0.892	0.452	0.452	1.714	-0.319	2.399	-6.633	0.785
	200	0.896	0.452	0.453	1.738	-0.342	2.445	-6.774	0.782
	400	0.882	0.438	0.447	1.746	-0.375	2.537	-7.052	0.787
8	100	0.982	0.551	0.448	1.887	-0.377	2.527	-7.026	0.791
	200	1.009	0.596	0.450	2.050	-0.434	2.672	-7.466	0.783
	400	0.992	0.569	0.446	1.934	-0.467	2.770	-7.762	0.788
10	100	1.051	0.648	0.437	2.215	-0.431	2.691	-7.511	0.801
	200	1.059	0.661	0.443	2.241	-0.504	2.874	-8.071	0.790
	400	1.051	0.646	0.439	2.182	-0.527	2.934	-8.256	0.794

Tabela 7. Resultados da floresta aleatória — poço 4

Prof.	# Arv.	Erro Normalizado				R ²			
		méd	dev	mín	máx	méd	dev	mín	máx
3	100	0.850	0.466	0.514	1.665	0.561	0.288	0.095	0.886
	200	0.856	0.479	0.512	1.697	0.555	0.303	0.060	0.887
	400	0.851	0.474	0.511	1.681	0.559	0.296	0.077	0.886
6	100	0.939	0.475	0.512	1.707	0.516	0.286	0.058	0.844
	200	0.926	0.476	0.504	1.708	0.518	0.290	0.055	0.849
	400	0.930	0.484	0.514	1.727	0.513	0.295	0.040	0.846
8	100	0.985	0.466	0.549	1.722	0.497	0.280	0.042	0.805
	200	0.976	0.468	0.545	1.719	0.502	0.283	0.047	0.821
	400	0.984	0.477	0.559	1.744	0.497	0.290	0.025	0.815
10	100	1.015	0.469	0.606	1.760	0.488	0.291	0.011	0.804
	200	1.001	0.460	0.597	1.735	0.496	0.284	0.033	0.807
	400	1.006	0.465	0.600	1.749	0.491	0.287	0.021	0.798

4.5. Importância das Características

A importância das características foi calculada para os melhores modelos para cada poço. Para todos os casos, a proximidade da broca é a característica que mais causa redução no erro cometido pelos modelos. Para os poços 1 e 2 a profundidade também teve importância significativa sendo a segunda característica mais importante. A importância maior dada à profundidade para os poços 1 e 2 em comparação com os poços 3 e 4 pode ser explicada pela maior amplitude de profundidade deles. O mesmo pode ser observado para a inclinação no poço 2, que varia até um valor de 14,52°, significativamente maior que as inclinações dos outros poços.

4.6. Aprendizado por Transferência

Para avaliar a possibilidade de aprendizado por transferência, são utilizados os poços 1 e 2, com o melhor modelo. Como os melhores modelos encontrados para os poços 1 e

2 não foram comuns entre eles, escolheu-se o modelo de floresta aleatória para avaliar o aprendizado por transferência por ter tido o melhor desempenho geral, com a combinação de hiper-parâmetros de profundidade máxima da árvore igual a 8 (intermediário entre os melhores resultados obtidos para cada poço) e 100 árvores.

Foram utilizados dois modelos para avaliar o aprendizado por transferência: o modelo simples corresponde à aplicação direta do modelo treinado em todo o primeiro poço para a previsão no segundo poço e o modelo composto que utiliza a média entre as previsões do modelo treinado em todo o primeiro poço e do modelo treinado com os dados de profundidades menores para o segundo poço para a previsão no segundo poço. Observando a Tabela 8, observa-se que o modelo composto tem pior performance que o

Tabela 8. Resultados para o aprendizado por transferência

Poço Teste	Modelo	Erro Normalizado				R ²			
		méd	dev	mín	máx	méd	dev	mín	máx
2 → 1	Simples	0.594	0.318	0.385	2.599	0.451	0.329	-1.330	0.804
	Composto	0.606	0.363	0.410	3.484	0.358	0.204	-1.156	0.616
1 → 2	Simples	0.693	0.247	0.421	2.334	0.232	0.391	-1.071	0.753
	Composto	0.611	0.201	0.389	2.127	0.359	0.274	-0.556	0.781

modelo simples, especialmente em termos de R². Enquanto isto para previsão do poço 2 o desempenho foi significativamente melhor para o modelo composto.

Além disso, comparando-se os resultados obtidos nos experimentos numéricos de transferência de aprendizado com os resultados dos experimentos anteriores, verifica-se que os resultados de transferência foram todos inferiores, mostrando uma transferência negativa de aprendizado. Para a previsão no poço 1, o melhor resultado médio entre os experimentos numéricos originais foram erro normalizado de 0,502 e R² de 0,553, comparados aos 0,594 de erro normalizado e 0,451 de R² do aprendizado por transferência. Já para a previsão no poço 2, o melhor resultado médio entre os experimentos numéricos originais foi erro normalizado de 0,524 e R² de 0,417, enquanto os resultados de aprendizado por transferência foram 0,611 de erro normalizado e 0,359 de R².

5. Conclusão

Este trabalho teve como objetivo avaliar o uso de técnicas de aprendizado de máquina para a previsão da taxa de perfuração (ROP) durante a perfuração de poços de petróleo *offshore* de pré-sal da Bacia de Santos.

A metodologia utilizada no trabalho considerou a coleta e preparação dos dados, e a definição dos experimentos numéricos que foram realizados. O conjunto de dados obtido é formato por 4 poços *offshore* da Bacia de Santos com dados disponíveis para diferentes extensões de profundidade do poço. A etapa de preparação dos dados tratou os problemas encontrados nos dados coletados. Para os experimentos foi avaliado o algoritmo de floresta aleatória variando os hiper-parâmetros, e o aprendizado por transferência.

Os resultados obtidos foram inferiores a trabalhos semelhantes verificados em trabalhos relacionados; um motivo provável para isso foi a falta de alguns dados que, como

apontado na subseção 3.1, têm papel importante nos métodos já existentes de estimação da ROP, mas não estavam disponíveis para a realização deste trabalho.

Nos experimentos numéricos realizados a característica com maior importância para todos os poços é a proximidade da broca. O melhor resultado obtido para previsão da ROP foi para o poço 1 com os hiper-parâmetros profundidade máxima 10 e 400 estimadores, com erro normalizado de $0,502 \pm 0,182$ e R^2 de $0,553 \pm 0,200$. Apesar disso, a variação do número de estimadores teve pouco efeito no desempenho do algoritmo, o que justificaria a escolha de um número menor de árvores em favor de menor custo computacional.

Como trabalhos futuros, pretende-se utilizar outros parâmetros considerados pela literatura como determinantes na previsão da ROP, como de propriedades das formações perfuradas, do fluido de perfuração e das brocas utilizadas que não estavam disponíveis durante a execução do trabalho. Além disso, a utilização de métodos de identificação automática das operações de perfuração como sugerido em [Barbosa et al. 2019] poderia também ser implementada.

Referências

- Al Sairafi, F., Al Ajmi, K., Yigit, A., and Christoforou, A. (2016). Modeling and control of stick slip and bit bounce in oil well drill strings. In *SPE/IADC Middle East Drilling Technology Conference and Exhibition*. OnePetro.
- Barbosa, L. F. F., Nascimento, A., Mathias, M. H., and de Carvalho Jr, J. A. (2019). Machine learning methods applied to drilling rate of penetration prediction and optimization-a review. *Journal of Petroleum Science and Engineering*, 183:106332.
- GANDELMAN, R. A. (2012). Predição da rop e otimização em tempo real de parâmetros operacionais na perfuração de poços de petróleo offshore. Master's thesis, Universidade Federal do Rio de Janeiro.
- Hegde, C., Daigle, H., Millwater, H., and Gray, K. (2017). Analysis of rate of penetration (ROP) prediction in drilling using physics-based and data-driven models. *Journal of Petroleum Science and Engineering*, 159:295–306.
- Hegde, C. and Gray, K. (2017). Use of machine learning and data analytics to increase drilling efficiency for nearby wells. *Journal of Natural Gas Science and Engineering*, 40:327–335.
- Hegde, C., Pyrcz, M., Millwater, H., Daigle, H., and Gray, K. (2020). Fully coupled end-to-end drilling optimization model using machine learning. *Journal of Petroleum Science and Engineering*, 186:106681.
- Morais, J. M. d. (2013). *Petróleo em águas profundas: uma história tecnológica da Petrobras na exploração e produção offshore*. Instituto de Pesquisa Econômica Aplicada (Ipea).
- Shi, X., Liu, G., Gong, X., Zhang, J., Wang, J., and Zhang, H. (2016). An efficient approach for real-time prediction of rate of penetration in offshore drilling. *Mathematical Problems in Engineering*, 2016:1–13.
- Thomas, J. (2001). *Fundamentos de engenharia de petróleo*. Interciência.