

Escola Politécnica da Universidade de São Paulo
Departamento de Engenharia de Minas e de Petróleo

**Predição da Taxa de Perfuração de Poços Utilizando Modelos de Aprendizado
de Máquina**

Trabalho de formatura apresentado
referente a obtenção de diploma de
graduação em Engenharia de Petróleo
na Escola Politécnica da Universidade
de São Paulo.

Denise Lidiane da Silva

Santos, 06 de janeiro de 2023

Predição da Taxa de Perfuração de Poços Utilizando Modelos de Aprendizado de Máquina

Discente	Denise Lidiane da Silva
Instituição Sede	Escola Politécnica da Universidade de São Paulo Departamento de Engenharia de Minas e de Petróleo
Curso	Engenharia de Petróleo
Docentes orientadores	Dr. Marcio Augusto Sampaio Dra. Nara Angélica Policarpo
Área de concentração preferencial	Perfuração; Inteligência Computacional

Doutor Marcio Augusto Sampaio

Doutora Nara Angélica Policarpo

Denise Lidiane da Silva

Santos - SP

Autorizo a reprodução e divulgação total ou parcial deste trabalho, por qualquer meio convencional ou eletrônico, para fins de estudo e pesquisa, desde que citada a fonte.

Catálogo-na-publicação

da Silva, Denise Lidiane

Predição da Taxa de Perfuração de Poços Utilizando Modelos de Aprendizado de Máquina /

D. L. da Silva, M. A. Sampaio, N. A. Policarpo -- São Paulo, 2022.

41 p.

Trabalho de Formatura - Escola Politécnica da Universidade de São Paulo.
Departamento de Engenharia de Minas e de Petróleo.

1.Perfuração 2.Inteligência Artificial I.Universidade de São Paulo. Escola Politécnica. Departamento de Engenharia de Minas e de Petróleo II.t.

III.Sampaio, Marcio Augusto IV.Policarpo, Nara Angélica

AGRADECIMENTOS

Agradeço à Universidade de São Paulo, por conceder acesso acadêmico para discentes e docentes a plataformas como Elsevier, Google, Matlab usadas ao longo do desenvolvimento deste trabalho. Assim, também explico meus agradecimentos à Escola Politécnica da Universidade de São Paulo, por promover disciplinas tão importantes para o desenvolvimento atual da abordagem moderna da Engenharia quanto PMI3927 - Inteligência Computacional Aplicada à Indústria de Petróleo, na qual foi iniciado este trabalho, e ao Capítulo Estudantil da Sociedade de Engenharia de Petróleo, por auxiliar na formação dos alunos ao longo da graduação e conceder acesso estudantil gratuito a plataformas de referência na Engenharia de Petróleo como a OnePetro, também usada para o desenvolvimento da revisão bibliográfica deste trabalho. Finalmente, agradeço aos Professores Doutor Marcio Sampaio e Doutora Nara Angélica Policarpo, referente aos ensinamentos lecionados sobre Inteligência Computacional e Perfuração respectivamente durante o curso, bem como pelo esforço e paciência durante a orientação deste trabalho. Agradeço também aos meus pais e minha irmã, Rosângela, Edmilson e Erika, que são meu lar e meu refúgio, por fornecerem o apoio emocional durante toda a graduação, dando o suporte em todos os momentos, mesmo sem perceber o incentivo que aquilo proporcionou. Ao Victor, agradeço por segurar minha mão e caminhar junto comigo durante essa trajetória, onde nos desenvolvemos individualmente e em conjunto, em intelecto e em coração.

RESUMO

Com a crescente preocupação com sustentabilidade nas indústrias e preservação da natureza, muito tem-se discutido sobre a necessidade de tornar as atividades da Indústria de Óleo e Gás mais seguras e harmoniosas possível. Ao encontro desta demanda, o uso de recursos digitais como redes neurais e os conceitos de inteligência computacional está se tornando cada vez mais necessário para o avanço e desenvolvimento de práticas industriais mais eficientes e sustentáveis. O setor de perfuração possui um grande potencial para implementação de tecnologias mais avançadas devido ao seu alto custo operacional, ao grande volume de informações obtidas em tempo real e, também, em relação ao risco envolvido nas atividades. Assim, este trabalho visa apresentar metodologia de predição da taxa de penetração (*Rate of Penetration*) utilizando as ferramentas *Random Forest*, *Support Vector Machine* e Redes Neurais Artificiais *Multiple-Layer Perceptron*, discutir sobre exemplos de implementação de modelos de aprendizado de máquina para tais cálculos e, finalmente, propor o desenvolvimento de um código que seja capaz de prever tais taxas a partir dos dados de perfuração do Campo de Volve, disponibilizados pela Equinor, após o processo de manipulação de dados. Com isso, os resultados obtidos pelo trabalho foram avaliados por meio do erro quadrático médio e do tempo de execução, que indicam viabilidade do uso quando comparados com a bibliografia, sendo que o modelo de regressão *Random Forest* seria o mais indicado graças ao baixo erro resultante.

Palavras chave: Perfuração de Poços; Taxa de Penetração; Random Forest; Support Vector Machine; Redes Neurais Artificiais; Aprendizado de Máquina.

ABSTRACT

With the growth of the sustainable mindset at industries and the worry about nature preservation, there have been many discussions about the necessity of turning O&G industry activities as safer and harmonious as possible. To comply with this request, digital resources usage as neural networks and computational intelligence concepts are becoming more and more necessary for the advance and development of more efficient and sustainable industrial practices. Drilling sector has a great potential for the implementation of more advanced technologies thanks to its high operational cost, high amount of real-time information and, also, due to the risk involved in the activities. Thus, this work aims to present Rate of Penetration prediction methods using Random Forest, Support Vector Machine and Artificial Neural Network Multiple-Layer Perceptron tools, discuss examples of machine learning models implementation for such calculations and, finally, suggest the development of a code capable of predicting such rates from Volve Field drilling data, provided by Equinor, after data wrangling. With this, results from the paper were evaluated through mean squared error and execution time, which indicated feasibility of applying the model when compared with the literature, being the Random Forest regression the most feasible model due to its low resulting error.

Keywords: Wells Drilling; Rate of Penetration; Random Forest; Support Vector Machine; Artificial Neural Network; Machine Learning

LISTA DE FIGURAS

- [Figura 1 - Representação esquemática do MPD. Fonte: Park et al. \(2020\)](#)
- [Figura 2 - Organograma das atividades exercidas. Fonte: Elaboração Própria](#)
- [Figura 3 - Localização geográfica do Campo de Volve. Fonte: Ravasi et al \(2015\)](#)
- [Figura 4 - Diagrama de caixa do parâmetro de profundidade. Fonte: Elaboração Própria](#)
- [Figura 5 - Diagrama de caixa do parâmetro de gravidade rotacional contínua \(GS_G_C3\). Fonte: Elaboração Própria](#)
- [Figura 6 - Diagrama de caixa do parâmetro de quantidade de gás \(GS_GASA\). Fonte: Elaboração Própria](#)
- [Figura 7 - Diagrama de caixa do parâmetro de fluxo total de bombeamento de lama \(GS_TFLO\). Fonte: Elaboração Própria](#)
- [Figura 8 - Diagrama de caixa do parâmetro de peso superficial sobre a broca \(GS_SWOB\). Fonte: Elaboração Própria](#)
- [Figura 9 - Diagrama de caixa do parâmetro de densidade da lama retirada do poço \(GS_MDOA\). Fonte: Elaboração Própria](#)
- [Figura 10 - Diagrama de caixa do parâmetro de densidade da lama dentro do poço \(GS_MWTI\). Fonte: Elaboração Própria](#)
- [Figura 11 - Diagrama de caixa do parâmetro de velocidade rotacional \(GS_RPM\). Fonte: Elaboração Própria](#)
- [Figura 12 - Diagrama de caixa do parâmetro de torque superficial \(GS_TQA\). Fonte: Elaboração Própria](#)
- [Figura 13 - Diagrama de caixa do parâmetro de perda de carga \(GS_SPPA\). Fonte: Elaboração Própria](#)
- [Figura 14 - Diagrama de caixa do parâmetro de taxa de penetração da formação rochosa \(GS_ROP\). Fonte: Elaboração Própria](#)
- [Figura 15 - Diagrama de caixa do parâmetro de carga do gancho \(GS_HKLD\). Fonte: Elaboração Própria](#)
- [Figura 16 - Diagrama de caixa do parâmetro de profundidade após remoção de outliers. Fonte: Elaboração Própria](#)
- [Figura 17 - Diagrama de caixa do parâmetro de gravidade rotacional contínua \(GS_G_C3\) após remoção de outliers. Fonte: Elaboração Própria](#)
- [Figura 18 - Diagrama de caixa do parâmetro de quantidade de gás \(GS_GASA\) após remoção de outliers. Fonte: Elaboração Própria](#)
- [Figura 19 - Diagrama de caixa do parâmetro de fluxo total de bombeamento de lama \(GS_TFLO\) após remoção de outliers. Fonte: Elaboração Própria](#)
- [Figura 20 - Diagrama de caixa do parâmetro de peso superficial sobre a broca \(GS_SWOB\) após remoção de outliers. Fonte: Elaboração Própria](#)
- [Figura 21 - Diagrama de caixa do parâmetro de densidade da lama retirada do poço \(GS_MDOA\) após remoção de outliers. Fonte: Elaboração Própria](#)
- [Figura 22 - Diagrama de caixa do parâmetro de densidade da lama dentro do poço \(GS_MWTI\) após remoção de outliers. Fonte: Elaboração Própria](#)
- [Figura 23 - Diagrama de caixa do parâmetro de velocidade rotacional \(GS_RPM\) após remoção de outliers. Fonte: Elaboração Própria](#)
- [Figura 24 - Diagrama de caixa do parâmetro de torque superficial \(GS_TQA\) após remoção de outliers. Fonte: Elaboração Própria](#)

[Figura 25 - Diagrama de caixa do parâmetro de perda de carga \(GS_SPPA\) após remoção de outliers. Fonte: Elaboração Própria](#)

[Figura 26 - Diagrama de caixa do parâmetro de taxa de penetração da formação rochosa \(GS_ROP\) após remoção de outliers. Fonte: Elaboração Própria](#)

[Figura 27 - Diagrama de caixa do parâmetro de carga do gancho \(GS_HKLD\) após remoção de outliers. Fonte: Elaboração Própria](#)

[Figura 28 - Matriz de Correlação \(Heatmap\). Fonte: Elaboração Própria](#)

[Figura 29 - Importância dos parâmetros considerados para predição do ROP. Fonte: Elaboração Própria](#)

[Figura 30 - Cumulativo da importância dos parâmetros. Fonte: Elaboração Própria](#)

[Figura 31 - ROP x profundidade \(Regressão Random Forest\). Fonte: Elaboração Própria](#)

[Figura 32 - ROP x profundidade \(Regressão SVM\). Fonte: Elaboração Própria](#)

[Figura 33 - ROP x profundidade \(Multiple-Layer Perceptron\). Fonte: Elaboração Própria](#)

[Figura 34 - Importância dos parâmetros sem remoção do torque superficial. Fonte: Elaboração Própria](#)

[Figura 35 - Cumulativo da importância dos parâmetros sem remoção do torque superficial. Fonte: Elaboração Própria](#)

[Figura 36 - ROP x profundidade \(Regressão Random Forest\) sem remoção do torque superficial. Fonte: Elaboração Própria](#)

[Figura 37 - ROP x profundidade \(Regressão SVM\) sem remoção do torque superficial. Fonte: Elaboração Própria](#)

[Figura 38 - ROP x profundidade \(Multiple-Layer Perceptron\) sem remoção do torque superficial. Fonte: Elaboração Própria](#)

LISTA DE TABELAS

[Tabela 1 - Variáveis que influenciam estimativa da taxa de penetração e operação em reservatórios de óleo e gás. Fonte: Ossai & Duru \(2020\)](#)

[Tabela 2 - Avaliação entre os modelos de predição aplicados. Fonte: Elaboração Própria](#)

[Tabela 3 - Avaliação entre os modelos de predição aplicados sem remoção da tensão superficial. Fonte: Elaboração Própria](#)

SUMÁRIO

1. INTRODUÇÃO	10
2. OBJETIVOS	11
3. REVISÃO BIBLIOGRÁFICA	11
3.1. Inteligência Artificial na Indústria de Óleo e Gás	11
3.2. Aplicações de Redes Neurais Artificiais (ANN), Random Forest (RF) e Support Vector Machine (SVM)	12
3.3. Importância do Aprendizado de Máquina para a Perfuração	13
3.4. Previsão da Taxa de Penetração	15
3.4.1. Métodos empíricos	15
3.4.2. Métodos utilizando inteligência artificial	16
4. METODOLOGIA	17
4.1. Regressão Random Forest (RF)	18
4.2. Regressão Support Vector Machine (SVM)	18
4.3. Redes Neurais Artificiais (ANN-MLP)	18
5. ESTUDO DE CASO	19
6. RESULTADOS	20
7. CONCLUSÃO	38
8. REFERÊNCIAS BIBLIOGRÁFICAS	39

1. INTRODUÇÃO

Com o avanço da tecnologia em diversas áreas da Engenharia, pôde-se notar que ainda hoje existem atividades que demandam tempo e esforço a níveis maiores do que os exercidos pela indústria tradicional. Sendo assim, muitas das atividades de monitoramento e análise dos dados geofísicos gerados não são tão bem aproveitadas quanto poderiam devido à limitação envolvida em atividades manuais, como é o caso do controle da pressão durante a perfuração (Rooki 2016). Torna-se um dever, então, das empresas buscar cada vez mais conhecimento acerca da implementação destas tecnologias essenciais para o crescimento e otimização das atividades de perfuração devido à grande quantidade de dados gerados e obtidos em tempo real, bem como ao grande custo e risco envolvido nos procedimentos.

Assim, para lidar com altas complexidades, uma prática recorrente dentre os métodos de análise de problemas de Engenharia é o estabelecimento de simplificações que viabilizem sua solução manual sob as condições de não perder completamente sua contextualização. Essa prática implica, no entanto, na desconsideração de propriedades que influenciam no comportamento real daquele fenômeno (Park et al. 2020). Na indústria O&G, muitos dos problemas relacionados também são complexos e, assim, necessitam de simplificações para que seja passível de solução, sendo necessária alguma ferramenta que ajuste tal necessidade e possibilite o encontro de soluções para problemas o mais próximo possível da realidade.

Levando em consideração tal problemática, o uso de ferramentas da Inteligência Computacional tem ganhado destaque em pesquisas e aplicações da indústria. Uma delas é a ferramenta de Redes Neurais Artificiais (do inglês, *Artificial Neural Networks* - ANN) que, Segundo Rooki (2016), são modelos não lineares que simulam sistemas neurais biológicos, onde o objetivo é o processamento de informações complexas do organismo. Os sistemas neurais consistem em pequenas frações, chamadas de neurônios, onde a informação é processada paralelamente e, conectando-os, forma-se uma rede neural. Do mesmo modo, as redes neurais artificiais têm o objetivo do processamento de informações não lineares complexas de maneira mais otimizada e, para isso, usam-se diversos parâmetros, que seriam as frações menores, que podem ser otimizados por algoritmos genéticos (do inglês *Genetic Algorithm* - GA), por exemplo, em um sistema de múltiplas camadas que devem ser treinadas.

Uma das atividades exercidas durante a perfuração é o controle de pressão do poço durante a perfuração que visa manter a pressão natural da formação para evitar que ocorram problemas como o comprometimento da estabilidade das paredes do poço (Guilherme et al. 2010). Para isso, é necessário criar modelos físicos complexos para simular as propriedades termodinâmicas da formação durante a dinâmica da perfuração. Dentre os parâmetros de monitoramento do gerenciamento da pressão de perfuração (do inglês *management pressure drilling* - MPD), está a taxa de penetração da formação rochosa (do inglês, *Rate of Penetration* - ROP). Para a indústria de petróleo, este conceito trata da extensão perfurada do poço pela broca por unidade de tempo, resultando em uma taxa comumente usada para caracterizar a eficiência da atividade de perfuração.

Segundo Guilherme et al. (2010), muitas vezes, para modelar esse sistema complexo de perfuração às condições de análise da indústria, com as limitações operacionais, é necessário que sejam feitas simplificações como a linearização de processos. Infelizmente, ao mesmo tempo que tais simplificações tornam o problema mais fácil de ser resolvido, podem ocasionar na perda de efeitos físicos do problema real.

2. OBJETIVOS

Este trabalho visa apresentar uma metodologia de predição da taxa de penetração (*Rate of Penetration*) como um dos parâmetros utilizados para controlar a pressão de poço durante a perfuração, discorrer sobre exemplos de implementação de modelos de aprendizado de máquina para tais cálculos e, finalmente, propor o desenvolvimento de um código que seja capaz de prever tais taxas a partir dos dados de perfuração utilizando as ferramentas *Random Forest*, *Support Vector Machine* e *ANN Multiple-Layer Perceptron*. Assim, na revisão bibliográfica serão apresentados diversos estudos que se dedicaram ao cálculo citado, dando enfoque nos modelos existentes que fazem uso de redes neurais. O estudo, então, servirá de motivação para o desenvolvimento deste trabalho e do código com ferramentas da Inteligência Artificial para a estimativa da taxa.

3. REVISÃO BIBLIOGRÁFICA

3.1. Inteligência Artificial na Indústria de Óleo e Gás

Aplicações de ferramentas da inteligência artificial têm se tornado cada vez mais frequentes devido à capacidade de solucionar problemas com maior acurácia, velocidade e, também, maior volume de dados de entrada. Este avanço em tecnologias autônomas melhora significativamente a funcionalidade de empresas que, até então, investem apenas em monitoramentos remotos e otimização de processos (Toorajipour et al. 2021). Neste sentido, a Inteligência Computacional tem ganhado espaço em diferentes áreas da indústria, como é o caso de Óleo e Gás. Sabe-se que a produção de petróleo em campos *offshore* são atividades economicamente custosas, complexas e que demandam tempo e alto nível de qualificação dos operadores (Guilherme et al. 2010). Por isso, novas estratégias de produção em reservatórios estão sendo apresentadas para melhorar tais técnicas já conhecidas (Reginato et al. 2019).

Assim, são diversos os exemplos de aplicações destas ferramentas no setor petrolífero. Em Toorajipour et al. (2021), pode-se notar o estudo de diversos trabalhos cuja implementação tinha como foco o setor de *Supply Chain*, analisando as muitas aplicações que poderiam ser feitas em diferentes atividades de logística. Já em Agostini & Sampaio (2020), discorre-se sobre a importância da aplicação na estimativa de desgaste das brocas usadas durante a perfuração de poços devido às diferentes variáveis do sistema litológico. Por outro lado, em Reginato et al (2019), estuda-se sobre o uso de ferramentas para estimar curvas de permeabilidade relativa de acordo com os efeitos da alteração de molhabilidade junto à composição iônica da água injetada, abrangendo uma aplicação química e geológica.

De um modo geral, o investimento em sistemas e equipamentos de alto desempenho têm crescido em meio às operadoras de campos de petróleo, junto com o interesse em fazer monitoramento da perfuração e produção do poço em tempo real com uso de algoritmos de aprendizado de máquina (Agostini & Sampaio 2020). Isto se deve a tal técnica tornar possível a resolução de equações e modelos não lineares, com maior versatilidade e habilidade para agregar informações e estabelecer previsões de determinado dado (Reginato et al. 2019). Tais sistemas são capazes de receber informações de sinais e dados colhidos durante a produção e transformá-los em conhecimento para auxiliar tomadas de decisão, prevenindo, assim, grandes esforços a cada vez que haja uma alteração no modelo simplificado analisado (Agostini & Sampaio 2020).

3.2. Aplicações de Redes Neurais Artificiais (ANN), Random Forest (RF) e Support Vector Machine (SVM)

Neste sentido, o uso de Redes Neurais Artificiais e de Algoritmos Genéticos (GA) são muito frequentes em implementações no setor de produção. O GA tem amplo leque de aplicações desde contribuições científicas, quanto em desenvolvimento empresarial com processos que envolvem tomadas de decisão, até mesmo na eficiência do setor de suprimentos, tudo isso graças à aplicabilidade em otimização com múltiplos objetivos, podendo ser usada em circuitos fechados (Toorajipour et al. 2021). Seu objetivo é de simular aspectos da biologia e genética, como a evolução de organismos, em modelos numéricos, expressões linguísticas a partir da varredura do espaço de soluções, encontrando aquelas que são mais próximas da solução ótima.

Já as ANNs são sistemas de processamento de informações que, com seu aprendizado, são capazes de determinar as relações entre dados de entrada e saída utilizando bases de dados mais complexas. Durante a transformação digital da indústria, tal ferramenta tem sido usada em diferentes campos devido à capacidade de extrair relações complexas e não lineares, visto que consegue atribuir hipóteses mais próximas da realidade (Rooki, 2016). Para isso, é necessário especificar o tipo de rede utilizada, como por exemplo, a ANN-MLP, rede neural artificial simples em modelo *perceptron* de múltiplas camadas. Existem, também, outros modelos como: rede neural probabilística (PNN), uma implementação de arquitetura neural de rápido treinamento usando função de densidade probabilística (PDF) como função classificatória, neste modelo existem quatro camadas - *Input, Pattern, Summation, Decision* - (Agostini & Sampaio 2020); rede neural com retropropagação (BP), um modelo tradicional que consiste na atualização dos pesos de cada neurônio através do cálculo de seus gradientes para minimização de erros; rede neural de ajuste - *Neural Net Fitting* (NNF) - cuja arquitetura é conhecida e separada em três principais camadas - *Input, Hidden, Output* - sendo simples e propagando o sinal em uma única direção, que seria a camada de output (Reginato et al. 2019).

As redes neurais podem utilizar validação cruzada apenas durante o treinamento para que seus dados sejam generalizados, fazendo o aprendizado e validação. Após feito isso, é possível iniciar os testes. Algo importante a se notar, neste caso, é que o tamanho da rede neural é uma métrica sensível e importante para um bom ajuste de bons dados, como mostrado em Reginato et al. (2019), onde foi feita comparação entre redes com 15,

17 e 26 neurônios, chegando à conclusão de que os valores encontrados para 17 neurônios foram suficientemente bons para a pesquisa.

Para avaliar os modelos usados, é necessário que sejam estabelecidas equações ou análises classificatórias. Um exemplo é o uso de *Receiver Operating Characteristics* (ROC), que são úteis para domínios com distribuição de classe assimétrica graças às suas propriedades e da área sobre a curva ROC (AUC), como mostrado em Reginato et al. (2019). Existem também aplicações que usam modelos de alto fluxo de fidelidade, visando tornar problemas mais compreensíveis e sem muitas simplificações.

Em Machado et al (2021), foram analisadas outras duas ferramentas da inteligência computacional; tais como máquinas de vetores de suporte (do inglês, *Support Vector Machine* - SVM) para classificar linear ou não-linearmente e em meio de regressão; e florestas aleatórias (do inglês, *Random Forest* - RF), que combina vários preditores fracos e apresenta resultados robustos.

Segundo Breiman (2001), a ferramenta *Random Forest* implica em uma combinação de preditores de árvores, onde cada uma delas depende de um vetor aleatório independente. Ou seja, para cada árvore de decisão, um vetor é gerado independente dos vetores anteriores, mas seguindo a mesma distribuição de todas as árvores da floresta. Assim, à medida que o número de árvores aumenta, o erro do modelo converge para um limite, dependendo do peso de cada árvore no modelo e da correlação entre elas.

Já o *Support Vector Machine* é explicado por Vapnik (1999) como uma ferramenta que, a partir do mapeamento não linear de vetores de entrada, visa obter uma solução que minimize o risco. Tais modelos têm sido amplamente utilizados para problemas reais e uma de suas vantagens é que a base de dados de treinamento é manipulada de modo que apenas os vetores de suporte contribuam para a predição do modelo (Pepper et al. 2022).

3.3. Importância do Aprendizado de Máquina para a Perfuração

Neste trabalho, o foco é a implementação de modelos de aprendizado de máquina na predição da taxa de penetração, usada durante atividades de monitoramento da pressão do poço a partir do *Managed Pressure Drilling* (MPD). Tal tecnologia permite o controle da pressão de poço, usando modelos que, na prática, possuem certas limitações como a consideração falha de algumas perturbações no sistema físico. Esta análise é de extrema importância uma vez que a margem entre a pressão do fluido presente nos poros da formação e a pressão de fratura pode ser estreita.

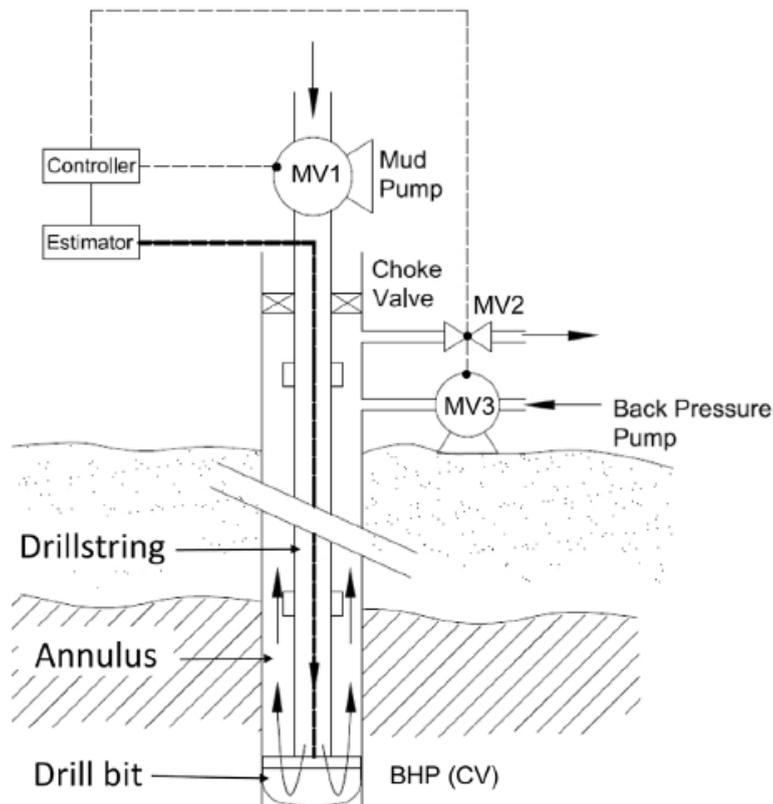


Figura 1 - Representação esquemática do MPD. Fonte: Park et al. (2020)

Sabe-se que a perfuração do poço envolve o bombeamento de fluido de perfuração continuamente através da coluna e brocas. A segurança da operação depende de qual a densidade correta que deve ser utilizada no fluido de perfuração de modo a fornecer pressão hidrostática maior do que aquela existente nos reservatórios para que não ocorra perda de circulação severa ou total, ou seja, que o fluido bombeado seja perdido para a formação e, possivelmente, ocasione um *blowout* (Machado et al. 2021). Geralmente tais previsões são feitas por especialistas, sem um sistema de apoio à decisão.

Essa análise, no entanto, envolve o uso de modelos que simulam o sistema físico de maneira simplificada para adaptar às limitações operacionais como entendimento incompleto do processo, limitação computacional ou falta de especialistas no assunto. Assim, geralmente são realizadas reduções para que este modelo seja linearizado, como é o caso do “*gray box model*”, um modelo derivado empiricamente, incorporando funções lineares e não lineares. Para melhorar esse tipo de modelos, o uso de redes neurais artificiais em sistemas off-line são implementados, sem que seja necessário partir para resolução manual de modelos físicos altamente detalhados e complexos, conhecidos como *high-fidelity flow models* (HFMs) (Park et al. 2020).

Durante a pesquisa de Park et al (2020), foi usado um modelo HFM, porém com simplificações para aplicações específicas de perfuração (do inglês *real-time high-fidelity flow models*, RT-HFM). Este método se mostrou efetivo ao considerar situações em que as informações do sensor de fundo de poço eram disponibilizadas para retornar ao modelo e, também, onde estas informações eram limitadas ou inacessíveis.

As variáveis consideradas durante a análise são o fluxo de bombeamento de lama e a pressão a partir da abertura e fechamento da válvula *choke*, resultando em funções objetivo que visam minimizar a variação de pressão. Cada um dos parâmetros das equações mencionadas é considerado de maneira diferente em cada um dos três casos a seguir: perfuração normal (fluxo de lama, perfuração rotacional e penetração da formação rochosa), conexão de tubos (parada no fluxo de lama e na perfuração rotacional) e deslocamento da densidade dentro de um período fixo.

No estudo de Park et al. (2020), foi possível analisar que, para a perfuração normal, a função objetivo utilizou a pressão *choke* e fluxo dos fluidos de perfuração com uma calibração inicial de maneira bem-sucedida. Para a conexão de tubos, pôde-se notar que houve compensação de pressão *choke* à medida em que o fluxo de lama era cessado. Já para a mudança na densidade da lama, pôde-se notar que o gerenciamento da abertura da válvula *choke* não foi suficiente para ajustar a pressão do sistema, sendo necessário o gerenciamento do fluxo da lama. Ambas as análises foram, também, bem-sucedidas.

Existem também registros de estudos sobre estimativa de perdas de pressão de fricção em escoamentos laminar, transiente e turbulento usando ANNs com retropropagação no treinamento, onde são considerados o diâmetro do tubo, a densidade e a velocidade do fluido e a tensão de cisalhamento nas paredes no sistema, segundo Rooki (2016). Neste caso, as funções objetivo usadas para o método de retropropagação são diferentes das mencionadas anteriormente e podem ser encontradas no modelo citado em Rooki (2016).

3.4. Previsão da Taxa de Penetração

3.4.1. Métodos empíricos

Existe grande quantidade de estudos envolvendo modelos matemáticos para estimativa da ROP, porém a maior parte faz uso de modelos empíricos desenvolvidos durante a última década (Hohn et al. 2022). O modelo matemático de Bourgoyne & Young (1974) possui boa documentação, evidenciando o seguinte cálculo para a taxa de penetração da formação rochosa (dD/dt), como representado na Equação 1.

$$\frac{dD}{dt} = \exp(a_1 + \sum_{j=2}^8 a_j x_j) \quad (1)$$

Neste modelo, cada constante representa o efeito de uma variável da operação sobre a ROP, a_1 representa o efeito da resistência da formação rochosa, $a_2 x_2$ e $a_3 x_3$ representam o efeito da compactação, visto que a ROP diminui exponencialmente de acordo com a profundidade em uma formação naturalmente compactada, $a_4 x_4$ modela o efeito do diferencial de pressão ao longo do fundo do poço, $a_5 x_5$ representa o efeito do peso e diâmetro da broca (do inglês *bit weight and diameter*), $a_6 x_6$ representa o efeito da velocidade de rotação, $a_7 x_7$ modela o efeito do desgaste dos dentes da broca, $a_8 x_8$ representa o efeito da hidráulica da broca na ROP, cujos valores são obtidos experimentalmente.

Outros estudos consideram parâmetros adicionais, como o de Motahhari et al. (2010), no qual foi feita a conexão entre o desempenho da broca PDC com o desempenho do motor de bombeamento da lama de perfuração, visando adicionar o torque da broca à análise.

$$ROP = W_f \times \left(\frac{G \times RPM_t^y \times WOB^a}{D_B \times S} \right) \quad (2)$$

Neste modelo, representado pela Equação 2, D_B representa o diâmetro da broca, S representa a resistência mecânica da rocha confinada, sendo G um coeficiente determinado pela geometria da broca, tamanho e design dos cortadores da broca (do inglês *cutters*).

3.4.2. Métodos utilizando inteligência artificial

A previsão da taxa de penetração pode ser utilizada para auxiliar o controle da pressão do poço durante a atividade de perfuração visto que sofre influência do peso aplicado sobre a broca, da densidade da lama, das propriedades da formação, entre outros parâmetros a depender da complexidade e objetivo específico do modelo (Gan et al. 2022). Segundo Ossai & Duru (2020), as características químicas e físicas do fluido de perfuração devem ser consideradas para a otimização do ROP, pois tornam a perfuração da rocha mais rápida e fácil através da sinergia entre a lama e os diversos aditivos utilizados para controlar as condições de subsuperfície como temperatura e pressão, garantindo a segurança operacional.

Sendo assim, o estudo sobre a taxa de penetração inclui monitoramento contínuo das propriedades da lama de perfuração, as condições de temperatura e pressão da formação e seus parâmetros de mecânica. Tais dados podem ser analisados de forma *offline*, calculados por modelos matemáticos estáticos onde não se considera a mudança de litologia da formação rochosa. Outra possibilidade é a análise *online*, considerando um modelo dinâmico, onde a quantidade de dados usados se torna maior e mais complexa, segundo Gan et al. (2022), devido a problemas como mudanças bruscas de litologia, baixa qualidade dos dados recebidos durante a perfuração e não linearidade dos parâmetros analisados.

Tabela 1 - Variáveis que influenciam a estimativa da taxa de penetração e operação em reservatórios de óleo e gás.

Propriedades do poço	Propriedades do reservatório	Propriedades hidráulicas	Propriedades da broca
Pressão de poro	Profundidade vertical total (TVD)	Composição da formação da coluna rochosa	Tamanho e tipo da broca
Diâmetro do revestimento do poço	Ângulo de perfuração	Composição da lama de perfuração	Peso sobre broca (WOB)
Pressão de Standpipe (perda de carga)	Resistência da rocha	Vazão da lama de perfuração	Rotação por minuto (RPM)
Resistência da formação rochosa	Composição mineralógica do reservatório	Densidade da lama de perfuração	Torque
Permeabilidade e porosidade da formação rochosa		Viscosidade plástica da lama de perfuração	Taxa de desgaste da broca

Fonte: Ossai & Duru (2020)

4. METODOLOGIA

A partir de dados retirados da literatura, foi ajustado um modelo artificial para estimativa da taxa de penetração da formação rochosa (ROP) fazendo uso de ferramentas de inteligência artificial no ambiente Colab com linguagem de programação Python. Para isso, foram utilizadas as bibliotecas Numpy, Pandas, Seaborn, Matplotlib, Scikit-learn e Bs4 e os pacotes Azure-storage-blob e Pandas nas versões 0.37.1 e 1.3.5, respectivamente. O código foi dividido em dois arquivos para poder ser executado no ambiente Google Colab sem problemas de memória para armazenamento.

Graças à possibilidade de consultar os dados de perfuração por meio do banco de dados SQL da Microsoft Azure, este método foi utilizado para prevenir problemas com memória da máquina para extrair e salvar uma grande quantidade de dados, sem antes selecioná-los. Isto foi feito com o auxílio do guia de acesso e funções elaboradas por Duque & Jardim (2020) para acessar contêineres e blobs na nuvem. Após tal acesso, os dados foram reconhecidos como *data frame*, com o auxílio das funções da biblioteca Pandas e foram submetidos ao processo de manipulação (*data wrangling*), onde as colunas e linhas indesejadas, ou seja, com grande quantidade de valores nulos e duplicados, foram removidas a partir das funções específicas da biblioteca Pandas.

Para melhor compreensão dos dados que foram utilizados, foi realizada uma análise estatística com as bibliotecas Pandas, Matplotlib e Seaborn por meio de diagramas de caixa (*boxplot*), para estudo da distribuição dos dados e remoção de valores

discrepantes (*outliers*), e da matriz de correlação (*heatmap*), para identificação e remoção de colunas com alta dependência ao parâmetro de interesse; ROP. Após tais tratamentos, a base de dados foi exportada para a nuvem do Google Drive para ser executada pela segunda parte do código, onde foram realizadas as previsões supervisionadas.

A base de dados limpa foi submetida à etapa de preparação de dados, sendo separada a coluna com os valores reais de ROP no *data frame* “y” e as demais colunas no “X”, ambos *dataframes* foram utilizados para a separação dos conjuntos de treino e de teste dos modelos a partir da função *train_test_split* da biblioteca Sklearn. Após essa etapa, foi realizado inicialmente o treinamento e teste usando regressão Random Forest, sendo utilizado o parâmetro erro quadrático médio para avaliar a previsão realizada. Para diminuição desse valor, foram identificadas as importâncias das colunas do *data frame* “X” durante a previsão de “y” e, os parâmetros com melhor importância foram excluídos dos dados.

Utilizando os conjuntos de treino e de teste atualizados apenas com os parâmetros mais importantes, foi executada a Regressão *Random Forest* com combinação dos parâmetros da biblioteca Sklearn. Em seguida, a mesma base de dados foi utilizada para executar os modelos de Regressão *Support Vector Machine* e de Redes Neurais Artificiais *Multiple-Layer Perceptron* (MLP) usando *backpropagation*.

4.1. Regressão Random Forest (RF)

Para treinar o modelo, foi usado o método *Grid Search*, no qual a função *RandomForestRegressor* da biblioteca Scikit-learn foi executada com diferentes parâmetros, de modo a encontrar aqueles que retornaram os melhores resultados. Foram informados os seguintes valores para os parâmetros: 200, 250, 300, 350 ou 400 árvores de decisão; *None*, 5 ou 10 unidades de profundidade máxima para cada árvore. Como parâmetros do método *Grid Search*, foram informadas 5 subdivisões para a estratégia de validação cruzada e verbosidade igual a 1. A avaliação informada para o modelo foi o erro quadrático médio.

4.2. Regressão Support Vector Machine (SVM)

Para treinar o modelo, foi usado o método *Grid Search* com a função *SVR* da biblioteca Scikit-learn como estimador. Os possíveis valores informados por parâmetro foram: coeficiente do kernel (*gamma*) em modo *scale* (padrão) ou *auto*; 0,01, 0,1, 1 ou 100 como parâmetro de regularização; *epsilon* igual a 0,1, 0,2 ou 0,3. Os parâmetros do método *Grid Search* não foram alterados entre os modelos.

4.3. Redes Neurais Artificiais (ANN-MLP)

Para o treinamento iterativo do modelo, foi utilizada a função *MLPRegressor* da biblioteca Scikit-learn. Os seguintes parâmetros foram informados no código: 100, 150, 200, 250, 300 ou 350 camadas ocultas; 500 possíveis iterações no máximo; verbosidade igual a 1. Todos os demais parâmetros da função foram mantidos nos seus valores padrões.

Para todos os valores encontrados, os modelos foram avaliados de acordo com o erro quadrático médio relativo e, para facilitar a comparação entre tais resultados, foram feitos gráficos verticais de ROP em função da profundidade.

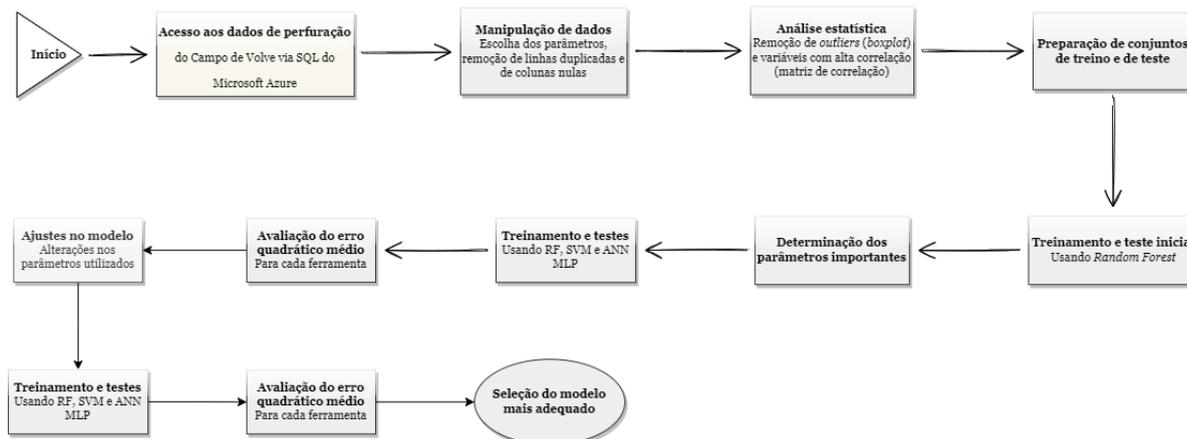


Figura 2 - Organograma das atividades exercidas. Fonte: Elaboração Própria

5. ESTUDO DE CASO

Os modelos foram treinados usando a base de dados do Campo de Volve, fornecido pela Equinor a partir de licenças gratuitas. O campo foi perfurado e produzido por meio de uma plataforma *jack-up*, tendo sido descomissionado em 2006. O campo se situa no bloco 15/9, na porção sudeste do Mar do Norte, a 200 km da cidade de Stavanger, na Noruega, próximo aos campos de Sleipner Ost e Vest, conforme apresentado na Figura 2. O acesso aos seus dados de perfuração foi realizado por meio do Microsoft Azure.

O poço escolhido para treinamento dos modelos foi o Norway-StatoilHydro-15_47\$9-F-10 15/9-F-10 W-924688, devido à maior quantidade de dados de perfuração em função da profundidade em relação aos demais analisados na mesma base de dados. Tal poço foi perfurado em uma lâmina d'água de 91 m, até a profundidade vertical de 2963 m abaixo do leito marinho. Foi necessário realizar a manipulação desses dados, filtrando os parâmetros a serem analisados visto que muitos estavam nulos para todas as profundidades, ou possuíam dados mais espaçados se comparados aos demais. Assim, foram escolhidos os seguintes parâmetros, resultando em uma base de dados com 90.000 linhas: *Depth* (m), *Continuous Rotational Gravity X-axis* (ppm), *FA Gas Amount* (%), *Total flow rate of all active pumps* (L/min), *Surface Weight On Bit* (kkgf), *Mud Density Out* (g/cm³), *Measured density of mud going into hole* (g/cm³), *Rotational Speed* (rpm), *Surface Torque* (ppm), *Standpipe Pressure* (kPa), *Rate of Penetration* (m/h), *Hook Load* (kkgf).

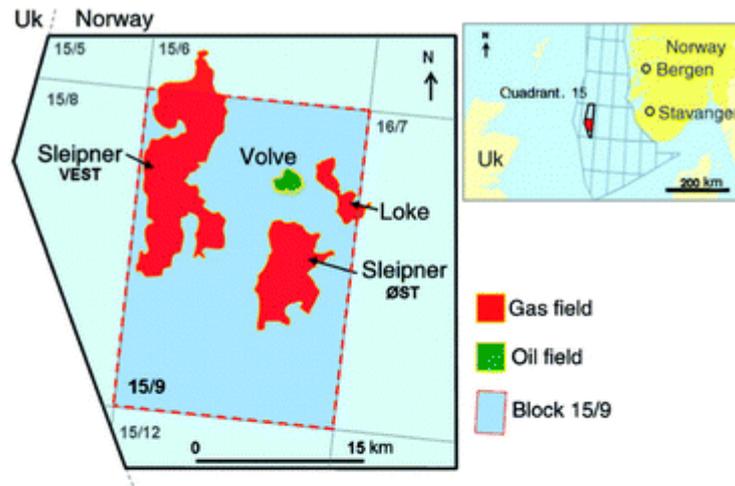


Figura 3 - Localização geográfica do Campo de Volve. Fonte: Ravasi et al (2015)

6. RESULTADOS

Após a seleção das colunas de interesse apresentadas anteriormente e a remoção das linhas nulas e duplicatas, a base de dados resultante possuía uma dimensão 1678 x 12 e estava pronta para a análise estatística, cujos diagramas de caixa estão apresentados:

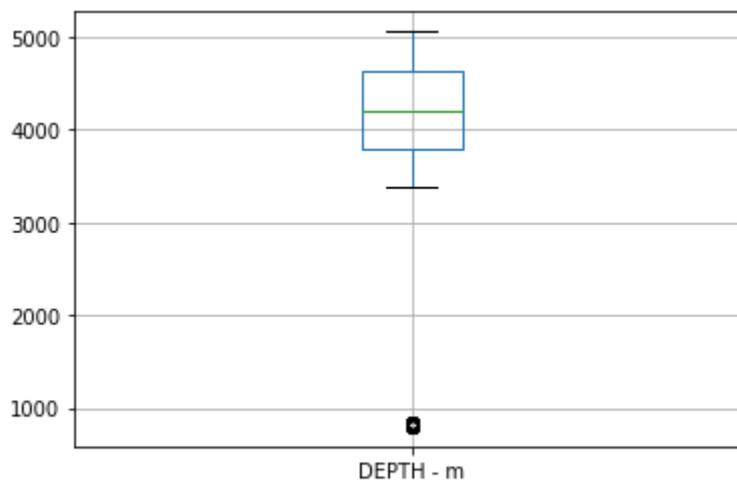


Figura 4 - Diagrama de caixa do parâmetro de profundidade. Fonte: Elaboração Própria

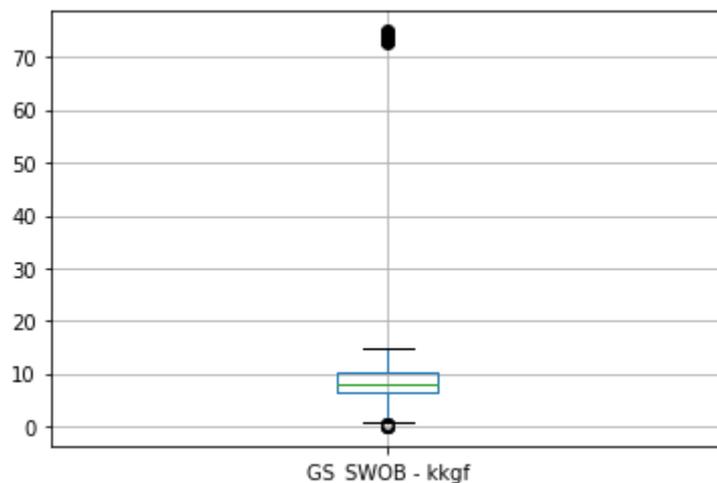


Figura 8 - Diagrama de caixa do parâmetro de peso superficial sobre a broca (GS_SWOB).
Fonte: Elaboração Própria

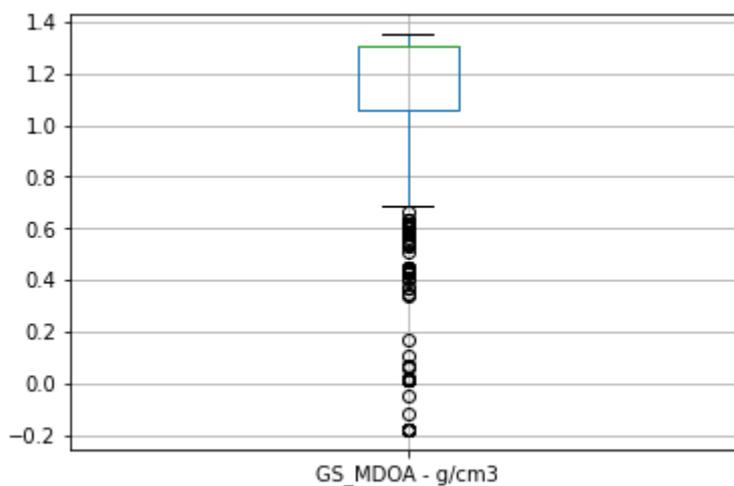


Figura 9 - Diagrama de caixa do parâmetro de densidade da lama retirada do poço (GS_MDOA). Fonte: Elaboração Própria

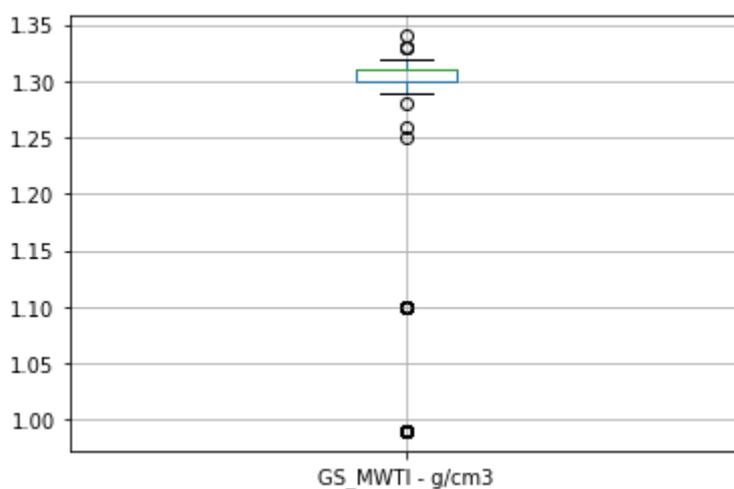


Figura 10 - Diagrama de caixa do parâmetro de densidade da lama dentro do poço (GS_MWTI). Fonte: Elaboração Própria

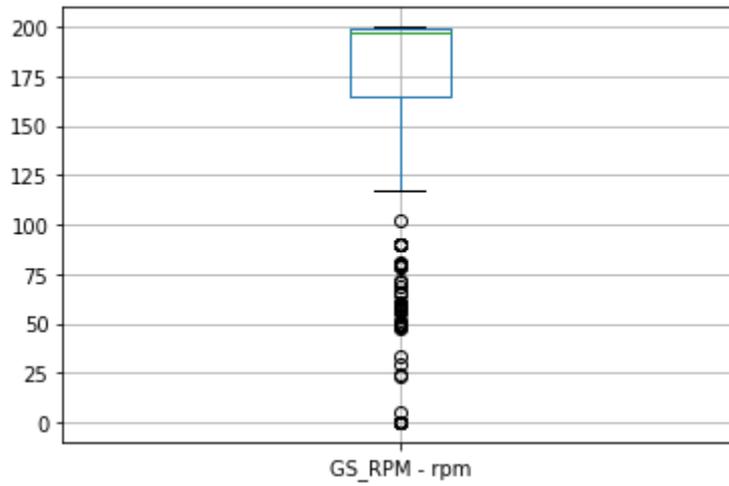


Figura 11 - Diagrama de caixa do parâmetro de velocidade rotacional (GS_RPM). Fonte: Elaboração Própria

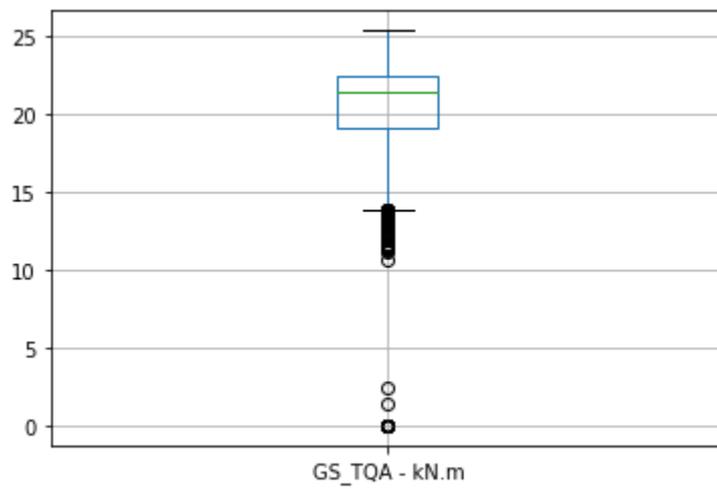


Figura 12 - Diagrama de caixa do parâmetro de torque superficial (GS_TQA). Fonte: Elaboração Própria

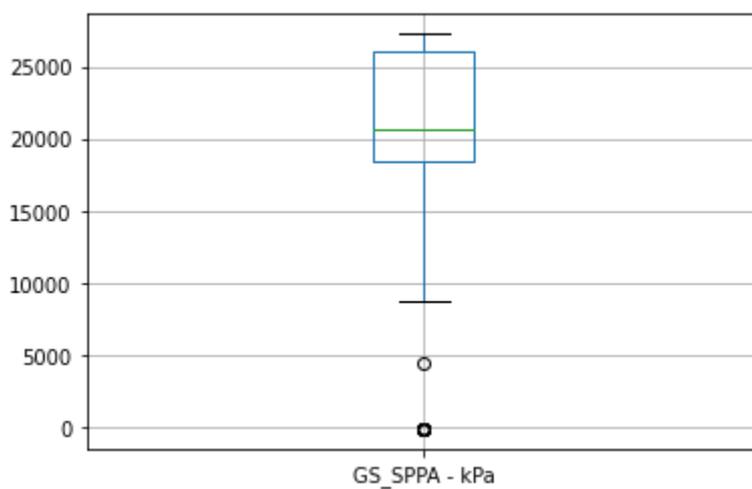


Figura 13 - Diagrama de caixa do parâmetro de perda de carga (GS_SPPA). Fonte: Elaboração Própria

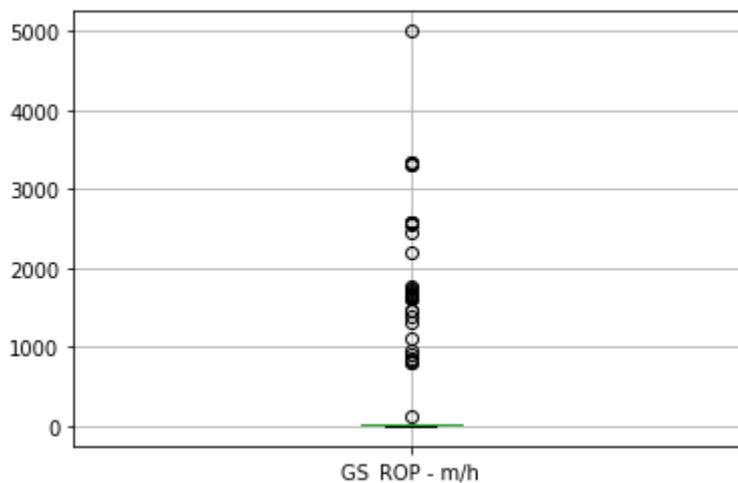


Figura 14 - Diagrama de caixa do parâmetro de taxa de penetração da formação rochosa (GS_ROM). Fonte: Elaboração Própria

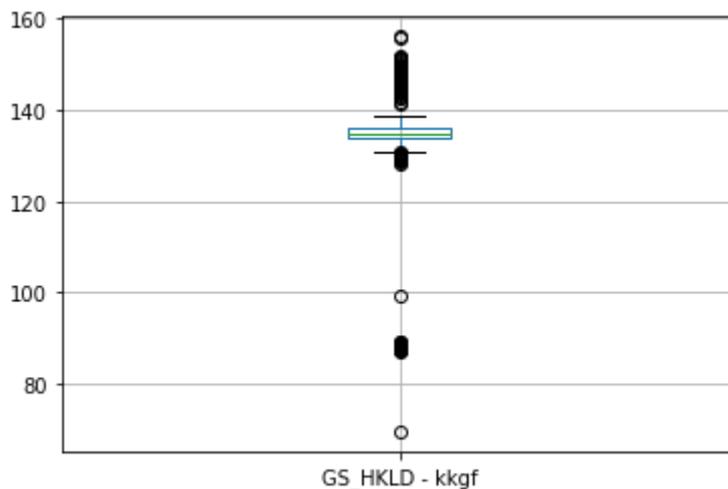


Figura 15 - Diagrama de caixa do parâmetro de carga do gancho (GS_HKLD). Fonte: Elaboração Própria

Após analisar as figuras 4 a 15, foram removidos os *outliers* dos seguintes parâmetros: valores de ROP maiores que 3000 m/h; valores de profundidade menores que 1000 m; valores de FA *Gas Amount* maiores que 7%; valores de gravidade rotacional contínua maiores que 2000 ppm, valores de carga do gancho menores que 100 kkgf e maiores que 155 kkgf. Com isso, foram geradas as figuras 16 a 27, representando os diagramas de *boxplot* por parâmetro já ajustados.

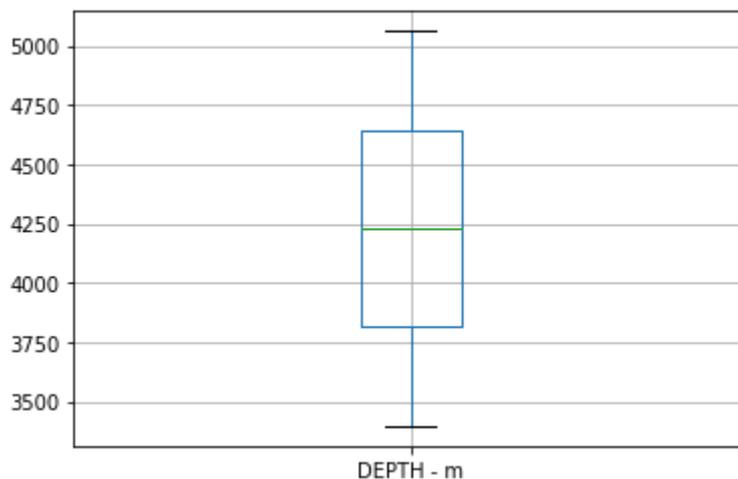


Figura 16 - Diagrama de caixa do parâmetro de profundidade após remoção de *outliers*.
 Fonte: Elaboração Própria

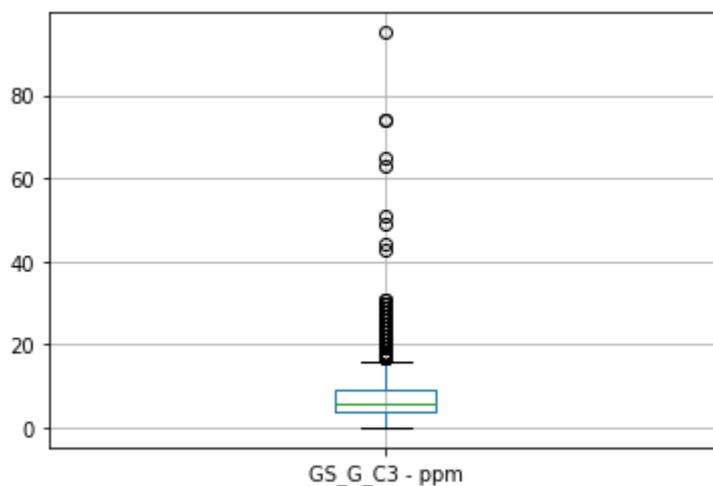


Figura 17 - Diagrama de caixa do parâmetro de gravidade rotacional contínua (GS_G_C3) após remoção de *outliers*. Fonte: Elaboração Própria

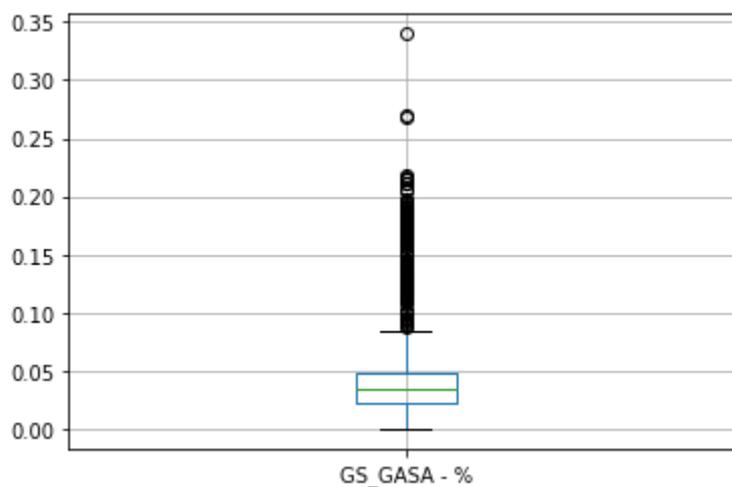


Figura 18 - Diagrama de caixa do parâmetro de quantidade de gás (GS_GASA) após remoção de *outliers*. Fonte: Elaboração Própria

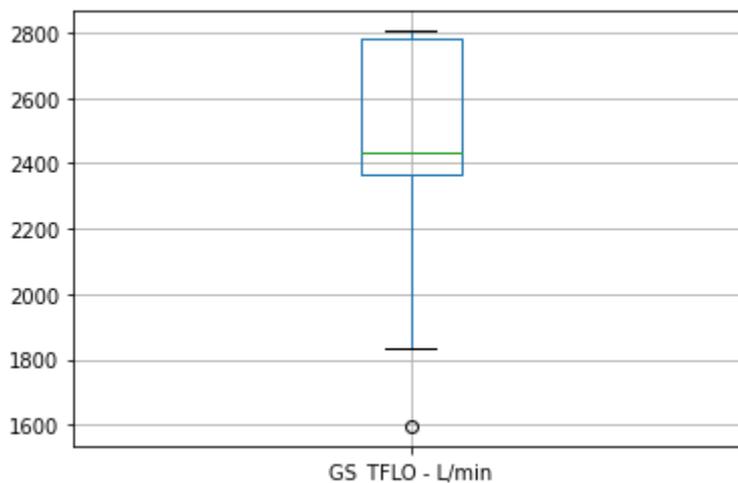


Figura 19 - Diagrama de caixa do parâmetro de fluxo total de bombeamento de lama (GS_TFLO) após remoção de *outliers*. Fonte: Elaboração Própria

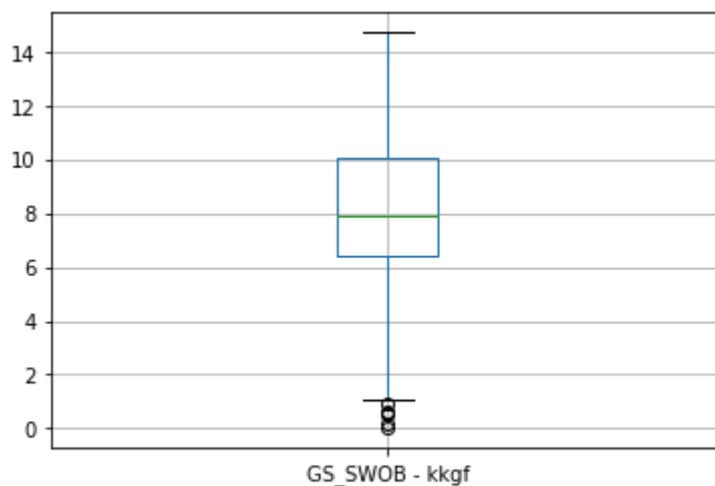


Figura 20 - Diagrama de caixa do parâmetro de peso superficial sobre a broca (GS_SWOB) após remoção de *outliers*. Fonte: Elaboração Própria

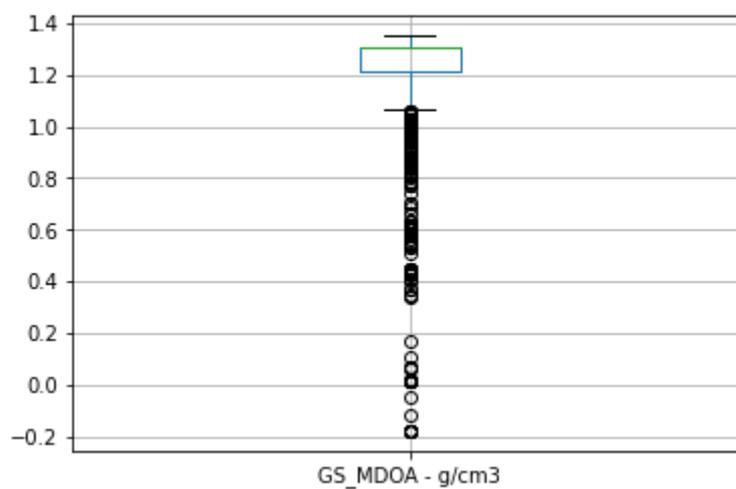


Figura 21 - Diagrama de caixa do parâmetro de densidade da lama retirada do poço (GS_MDOA) após remoção de *outliers*. Fonte: Elaboração Própria

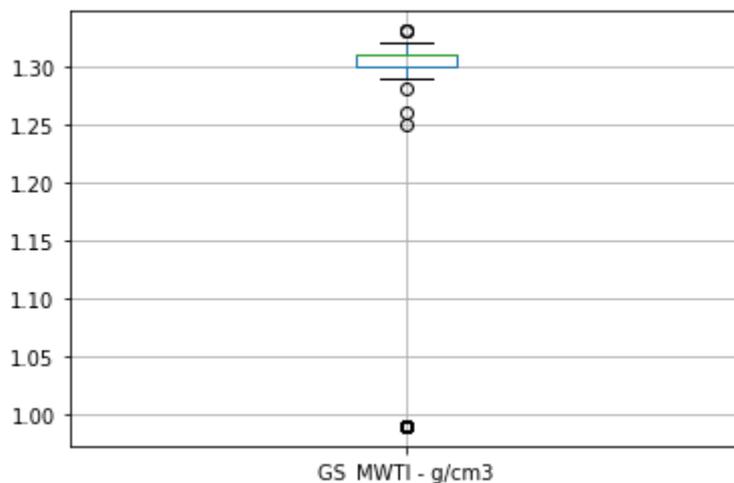


Figura 22 - Diagrama de caixa do parâmetro de densidade da lama dentro do poço (GS_MWTI) após remoção de *outliers*. Fonte: Elaboração Própria

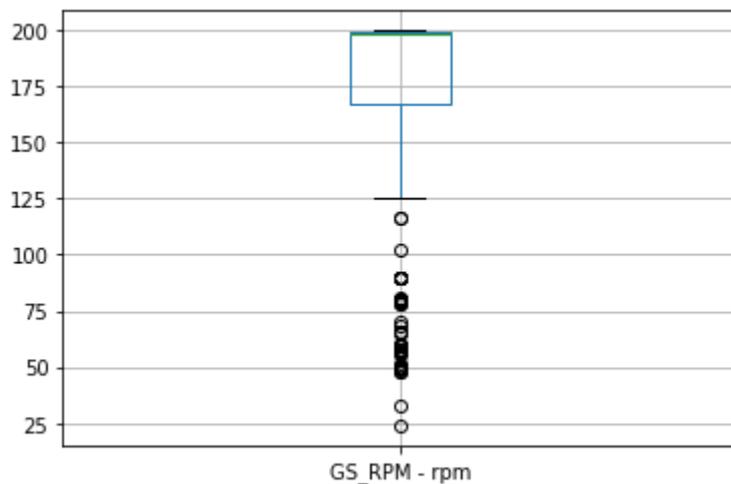


Figura 23 - Diagrama de caixa do parâmetro de velocidade rotacional (GS_RPM) após remoção de *outliers*. Fonte: Elaboração Própria

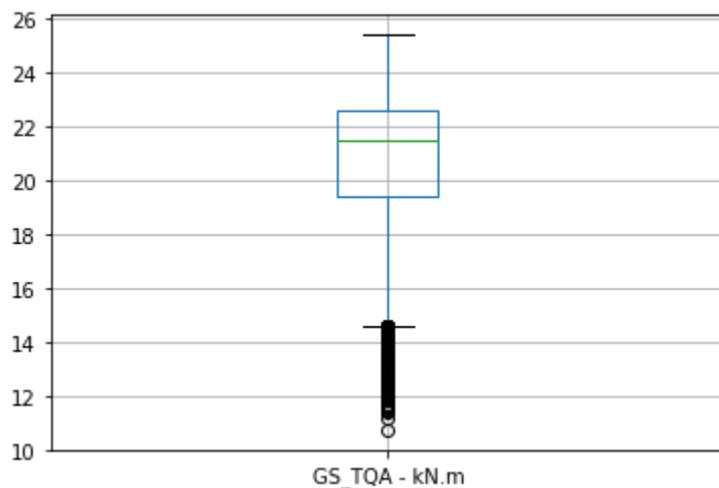


Figura 24 - Diagrama de caixa do parâmetro de torque superficial (GS_TQA) após remoção de *outliers*. Fonte: Elaboração Própria

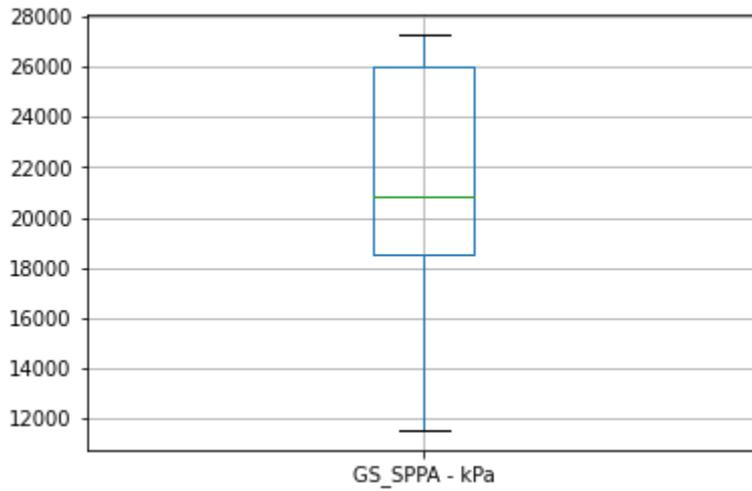


Figura 25 - Diagrama de caixa do parâmetro de perda de carga (GS_SPPA) após remoção de *outliers*. Fonte: Elaboração Própria

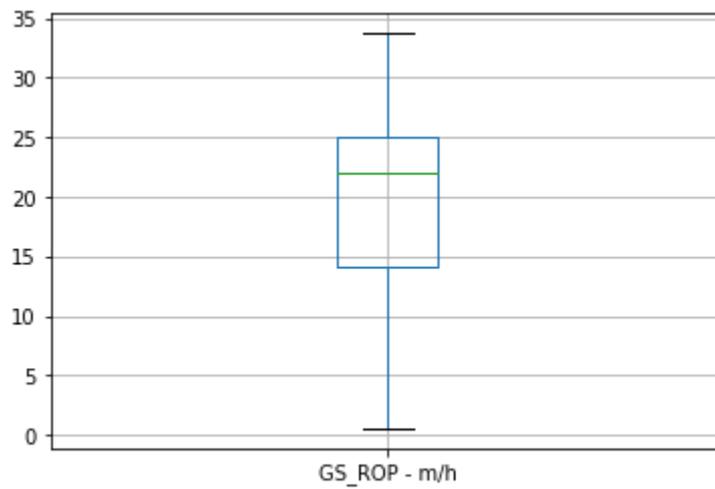


Figura 26 - Diagrama de caixa do parâmetro de taxa de penetração da formação rochosa (GS_ROP) após remoção de *outliers*. Fonte: Elaboração Própria

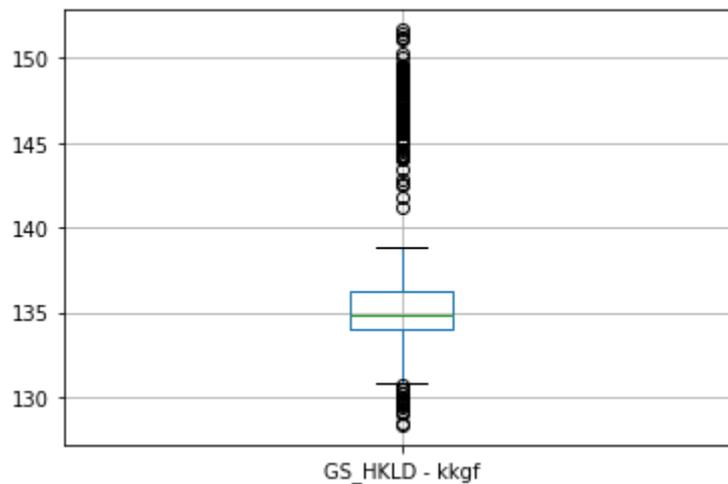


Figura 27 - Diagrama de caixa do parâmetro de carga do gancho (GS_HKLD) após remoção de *outliers*. Fonte: Elaboração Própria

Analisando-se a matriz de correlação da base de dados, conforme apresentado na Figura 27, pôde-se notar que o parâmetro de maior correlação com o ROP é torque de superfície com coeficiente igual a 0,844. Além disso, os parâmetros de carga do gancho e de densidade de lama dentro do poço possuem coeficiente de correlação igual a -0,871.

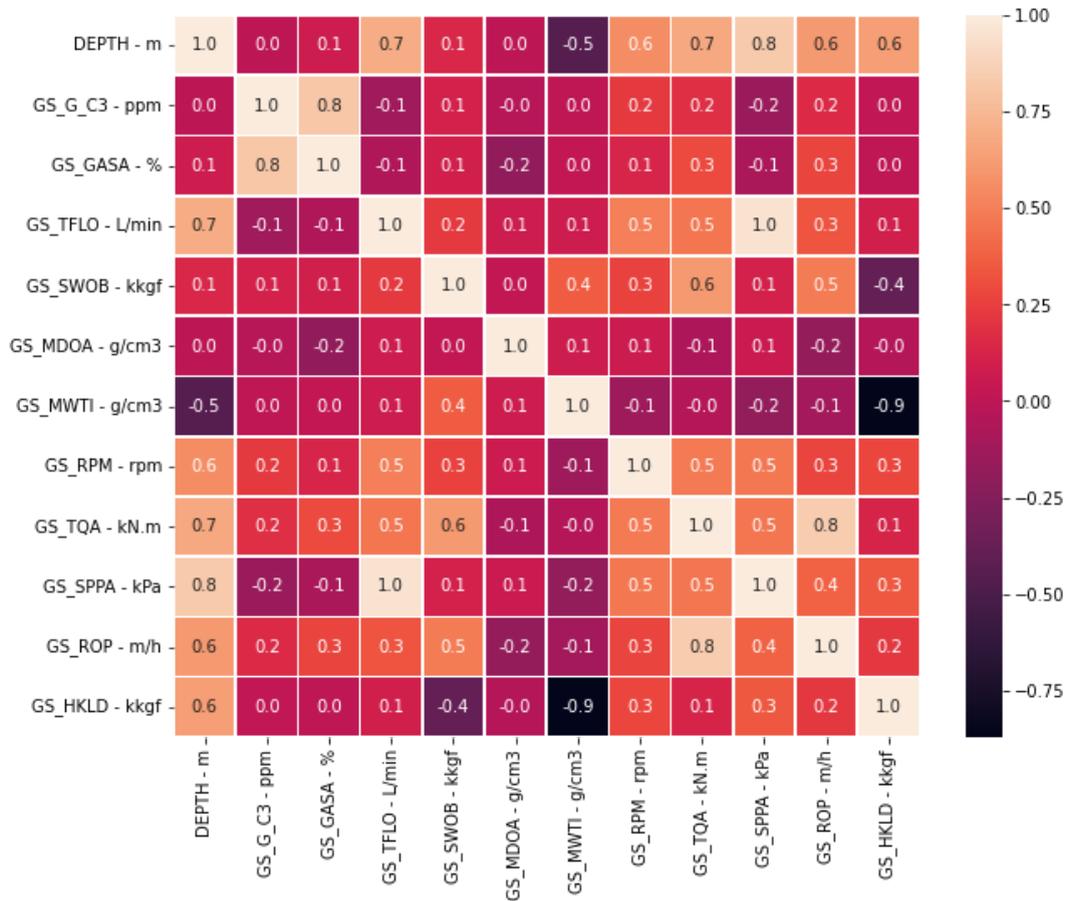


Figura 28 - Matriz de Correlação (*Heatmap*). Fonte: Elaboração Própria

Removendo-se os parâmetros de torque de superfície e de densidade de lama dentro do poço, resultou-se em um *data frame* de dimensão 1634 x 10.

Assim, após treinamento e teste inicial com o modelo de regressão *Random Forest*, o erro quadrático médio resultante foi de 5,55. Comparando-se tal valor com a bibliografia, foi possível notar a necessidade de diminuir o erro a partir da seleção dos parâmetros mais importantes, cuja análise está presente nas Figuras 28 e 29.

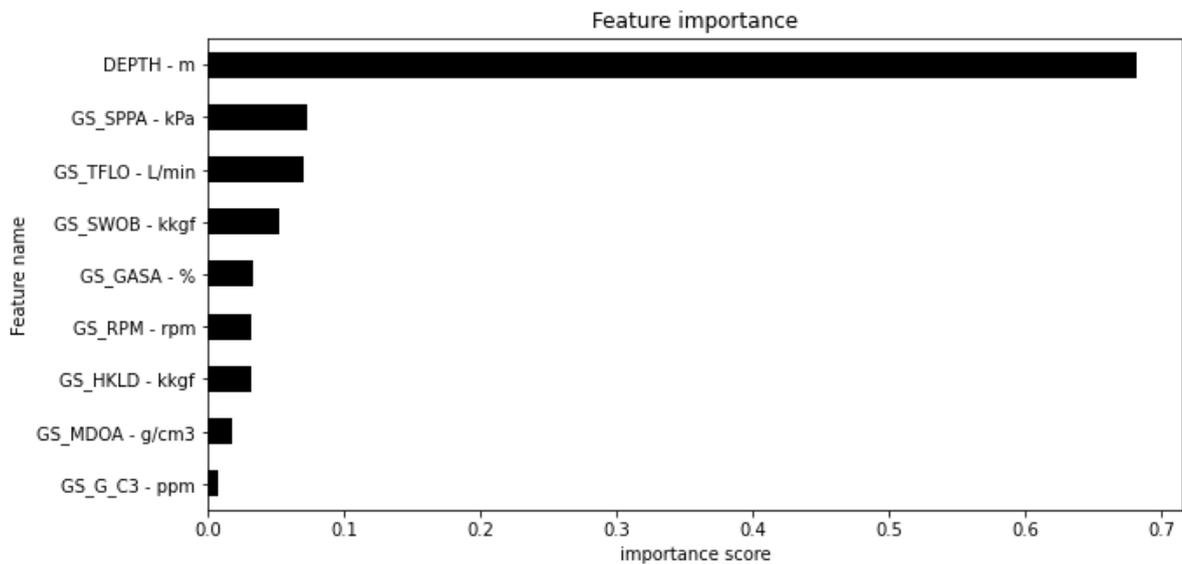


Figura 29 - Importância dos parâmetros considerados para predição do ROP. Fonte: Elaboração Própria

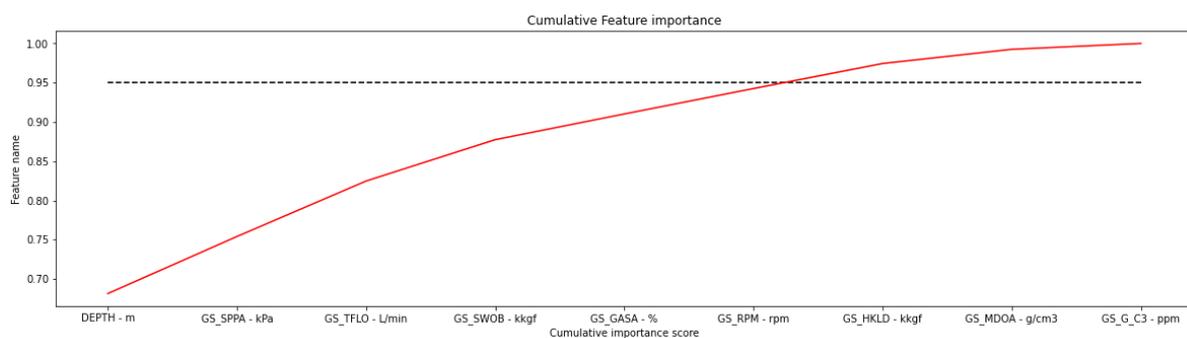


Figura 30 - Cumulativo da importância dos parâmetros. Fonte: Elaboração Própria

Portanto, foram removidos do *data frame* os parâmetros de densidade da lama retirada do poço e de gravidade rotacional contínua pois possuem baixa importância para a predição, sendo inferiores a 5% do total.

Assim, passando a base de dados atualizada novamente por treinamento (cuja duração foi de 2 minutos, resultando em 75 combinações) e teste com regressão *Random Forest*, o erro quadrático médio resultante foi de 2,39. Os resultados da predição estão evidenciados na Figura 30.

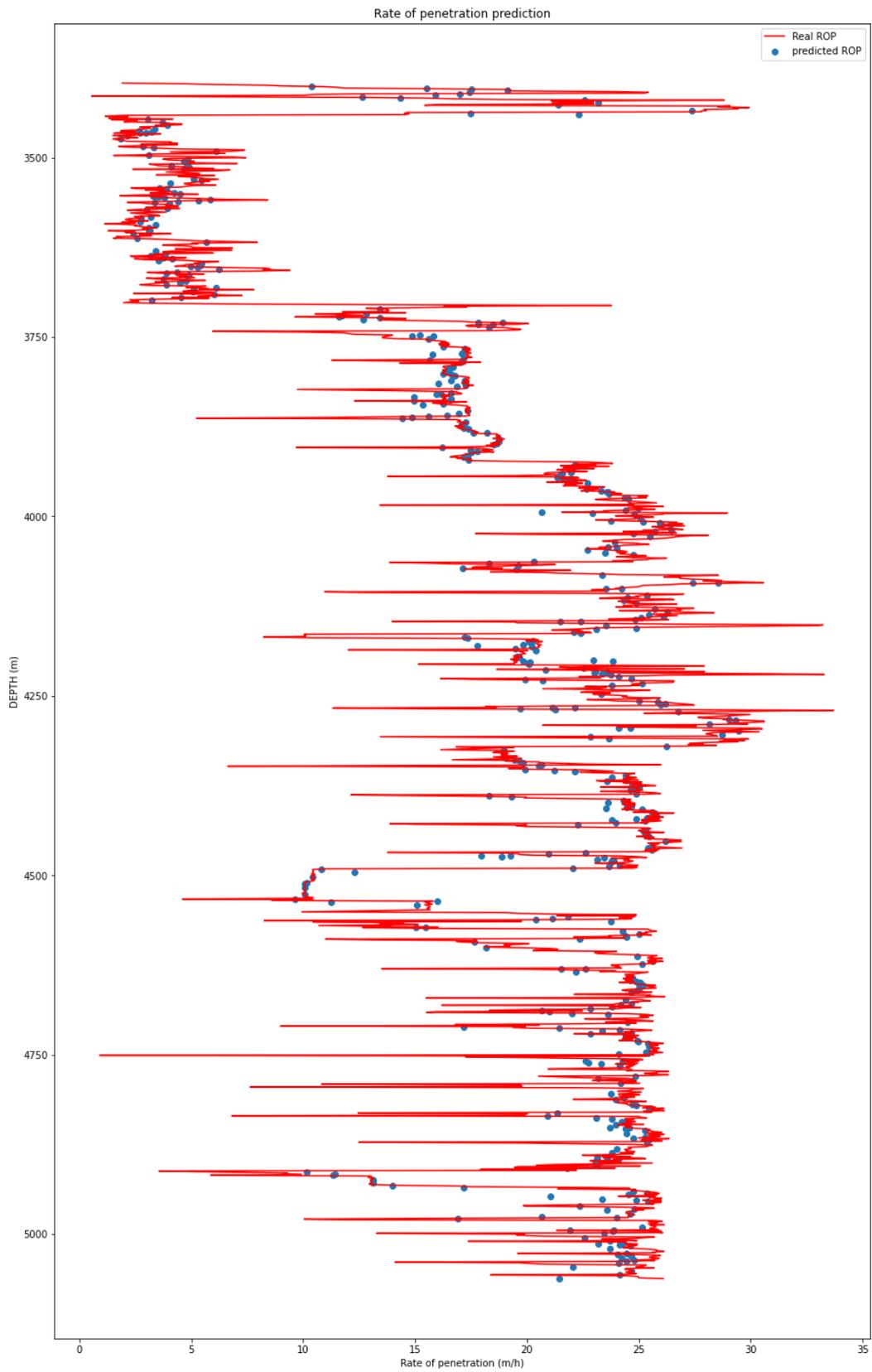


Figura 31 - ROP x profundidade (Regressão *Random Forest*). Fonte: Elaboração Própria

Em seguida, o *data frame* foi treinado e testado com o modelo de regressão *Support Vector Machine*, resultando no erro quadrático médio de 3,21. A duração do treinamento usando tal modelo foi de 13 segundos.

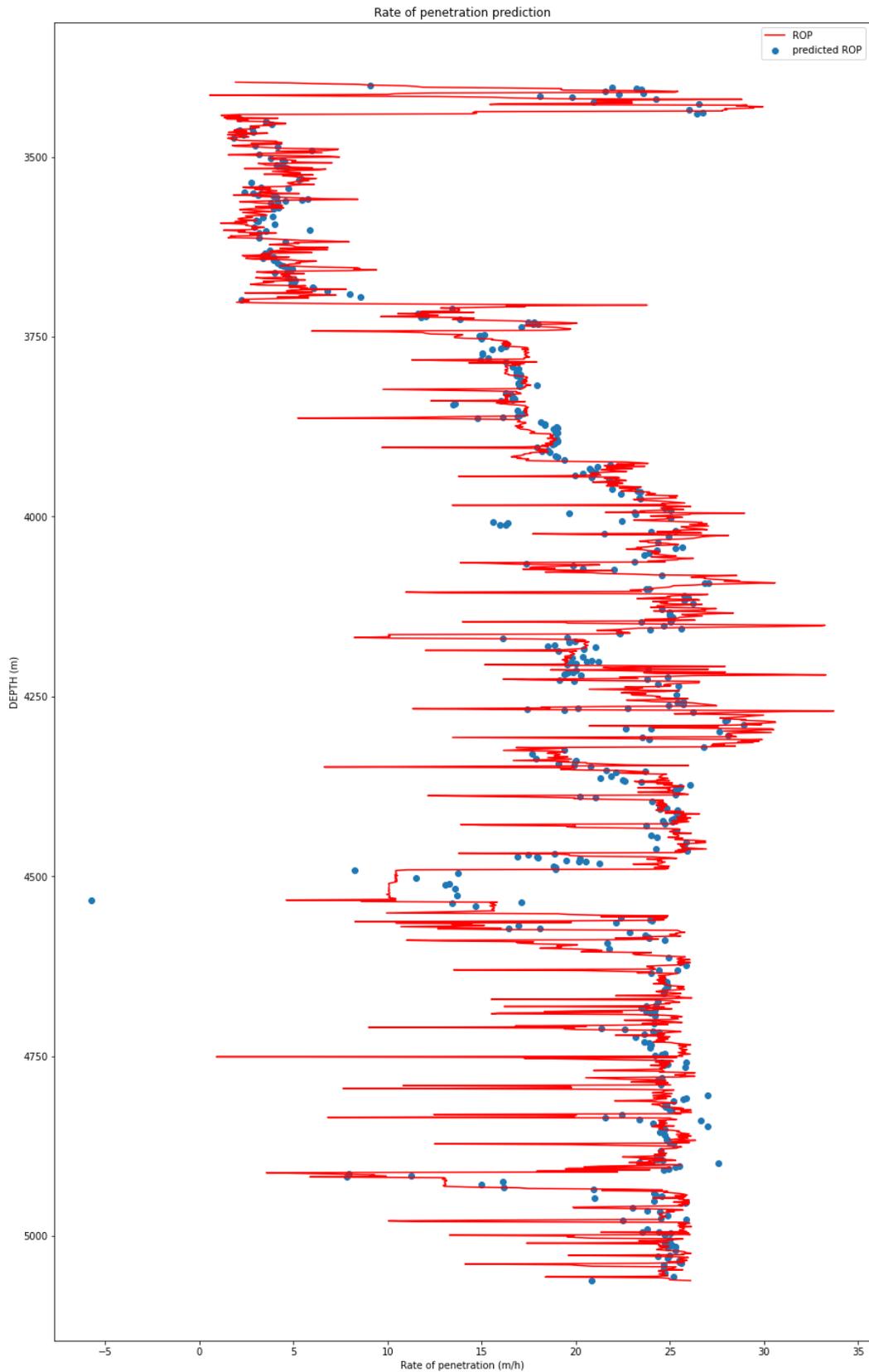


Figura 32 - ROP x profundidade (Regressão SVM). Fonte: Elaboração Própria

Finalmente, foi possível passar a base de dados pelo modelo de Redes Neurais Artificiais do tipo MLP, resultando no erro quadrático médio de 2,79. A duração do treinamento usando tal modelo foi de 27 segundos, totalizando 145 iterações.

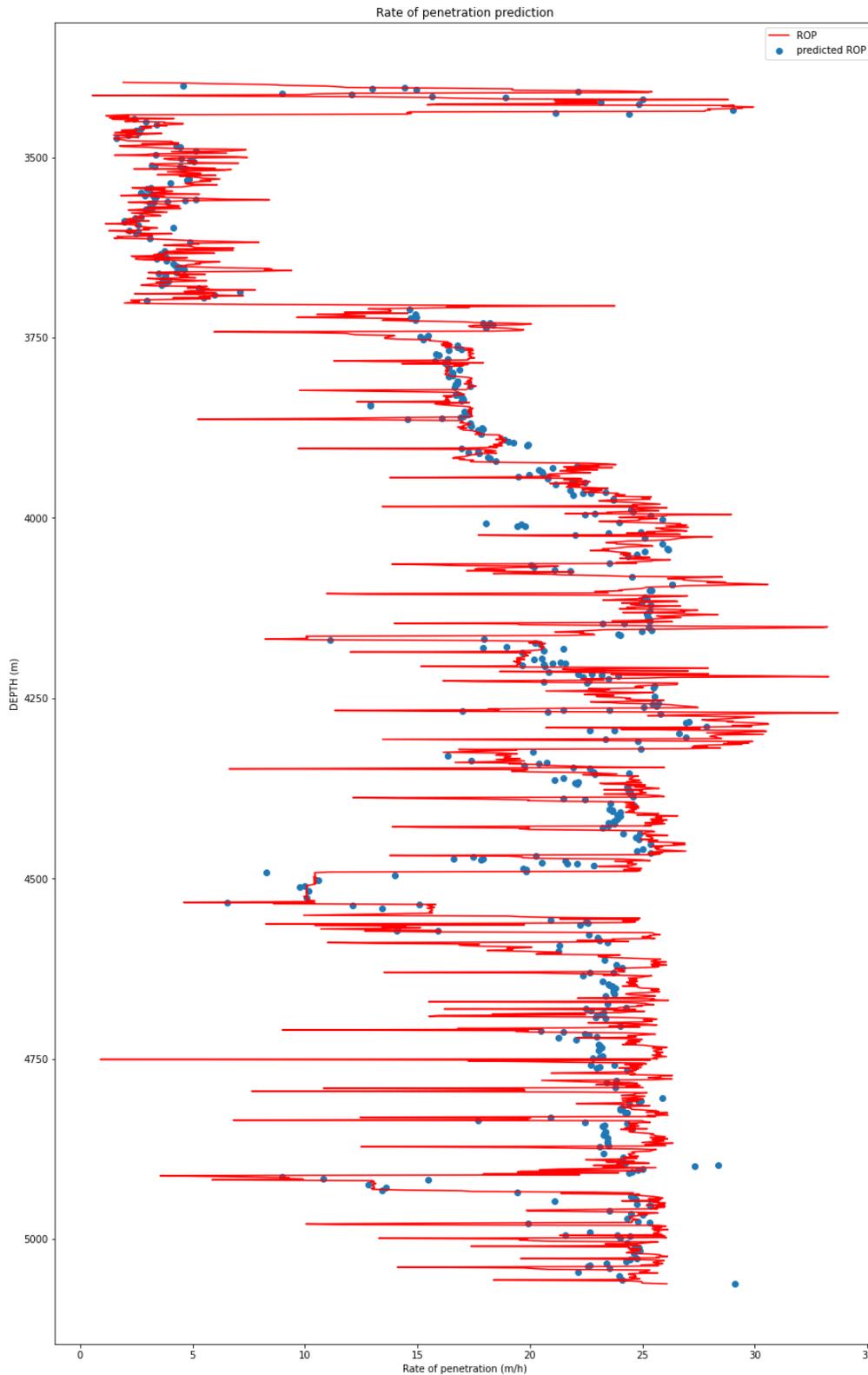


Figura 33 - ROP x profundidade (*Multiple-Layer Perceptron*). Fonte: Elaboração Própria

Assim, foi possível construir a Tabela 2, comparando os modelos utilizados.

Tabela 2 - Avaliação entre os modelos de predição aplicados.

Modelo	Regressão <i>Random Forest</i>	Regressão <i>Support Vector Machine</i>	<i>ANN Multiple-Layer Perceptron</i>
Erro Quadrático Médio	2,39	3,21	2,79
Duração do Treinamento (s)	128	13	27

Fonte: Elaboração Própria

Após realizar outros testes na etapa de análise estatística, foi possível observar que, mantendo-se o parâmetro de torque superficial, sua importância para a predição seria a maior dentre todas as colunas, então todas as etapas seguintes foram realizadas novamente com a base de dados com dimensão 1634 x 11.

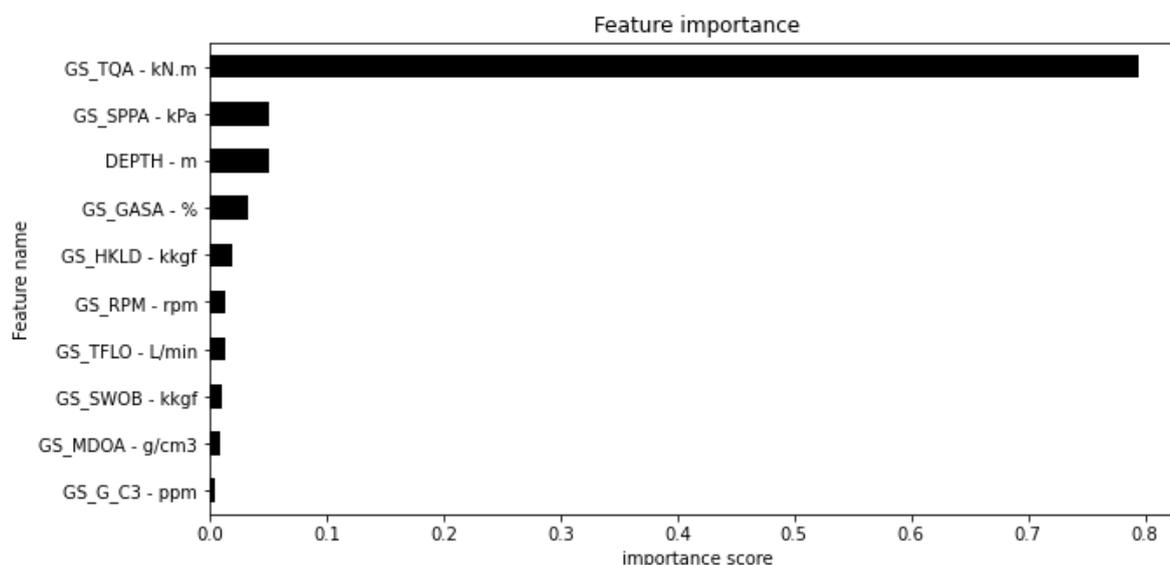


Figura 34 - Importância dos parâmetros sem remoção do torque superficial. Fonte: Elaboração Própria

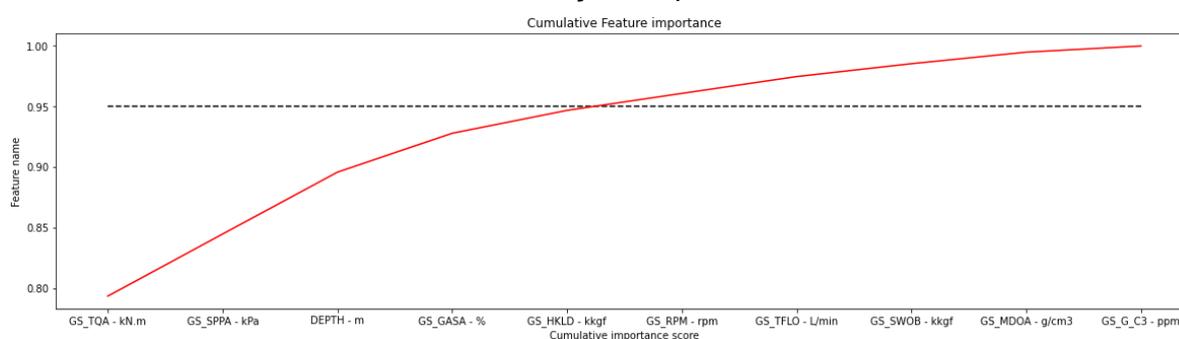


Figura 35 - Cumulativo da importância dos parâmetros sem remoção do torque superficial. Fonte: Elaboração Própria

Após realizar o treinamento com o modelo de regressão *Random Forest* durante 2 minutos, resultando em 75 combinações realizadas, o resultado do segundo teste está apresentado na Figura 35, com erro quadrático médio igual a 1,88.

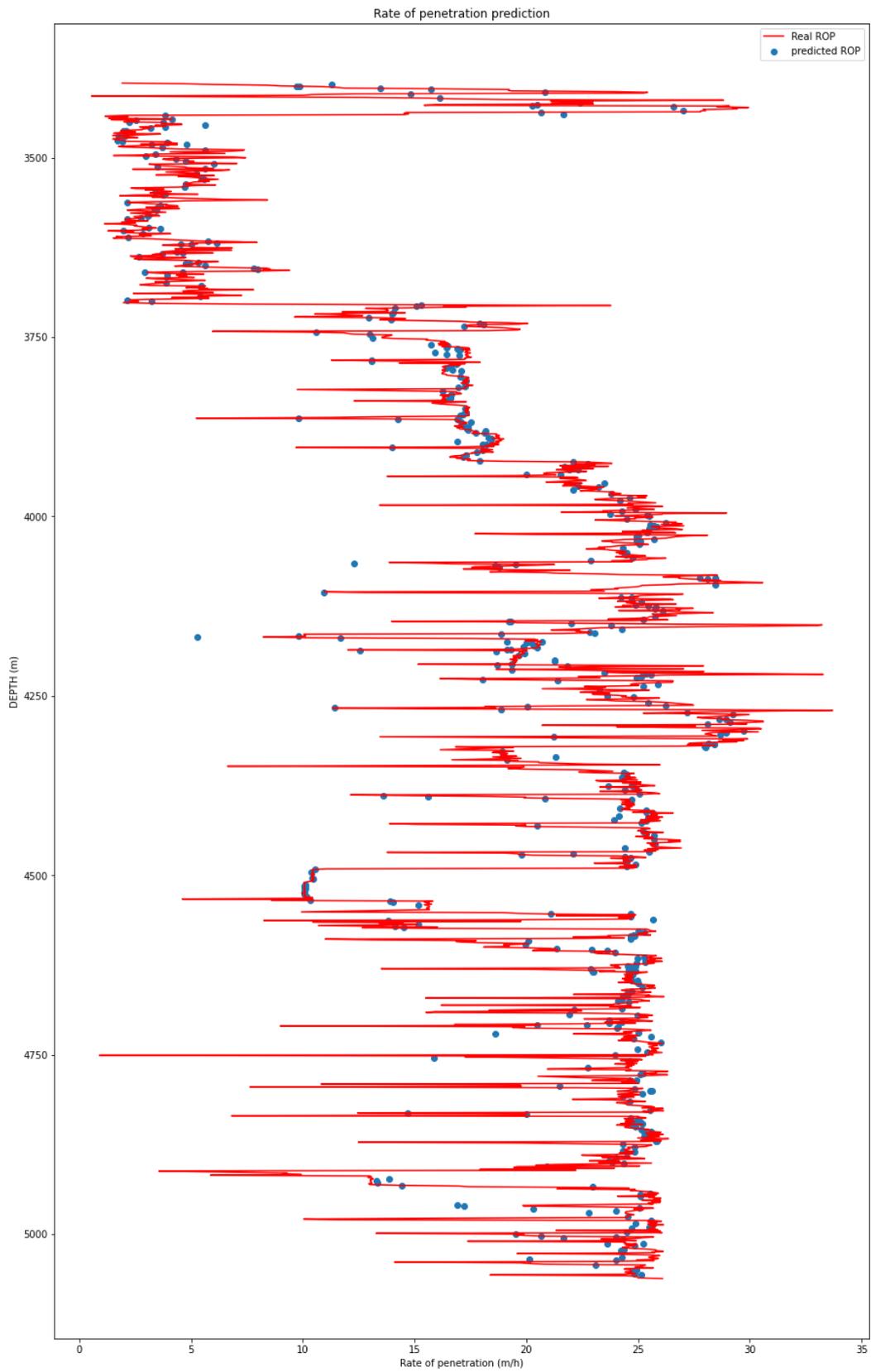


Figura 36 - ROP x profundidade (Regressão *Random Forest*) sem remoção do torque superficial. Fonte: Elaboração Própria

Aplicando-se o modelo de regressão *Support Vector Machine*, com treinamento de 12 segundos, o resultado do teste está presente na Figura 36 e seu erro quadrático médio foi igual a 2,15.

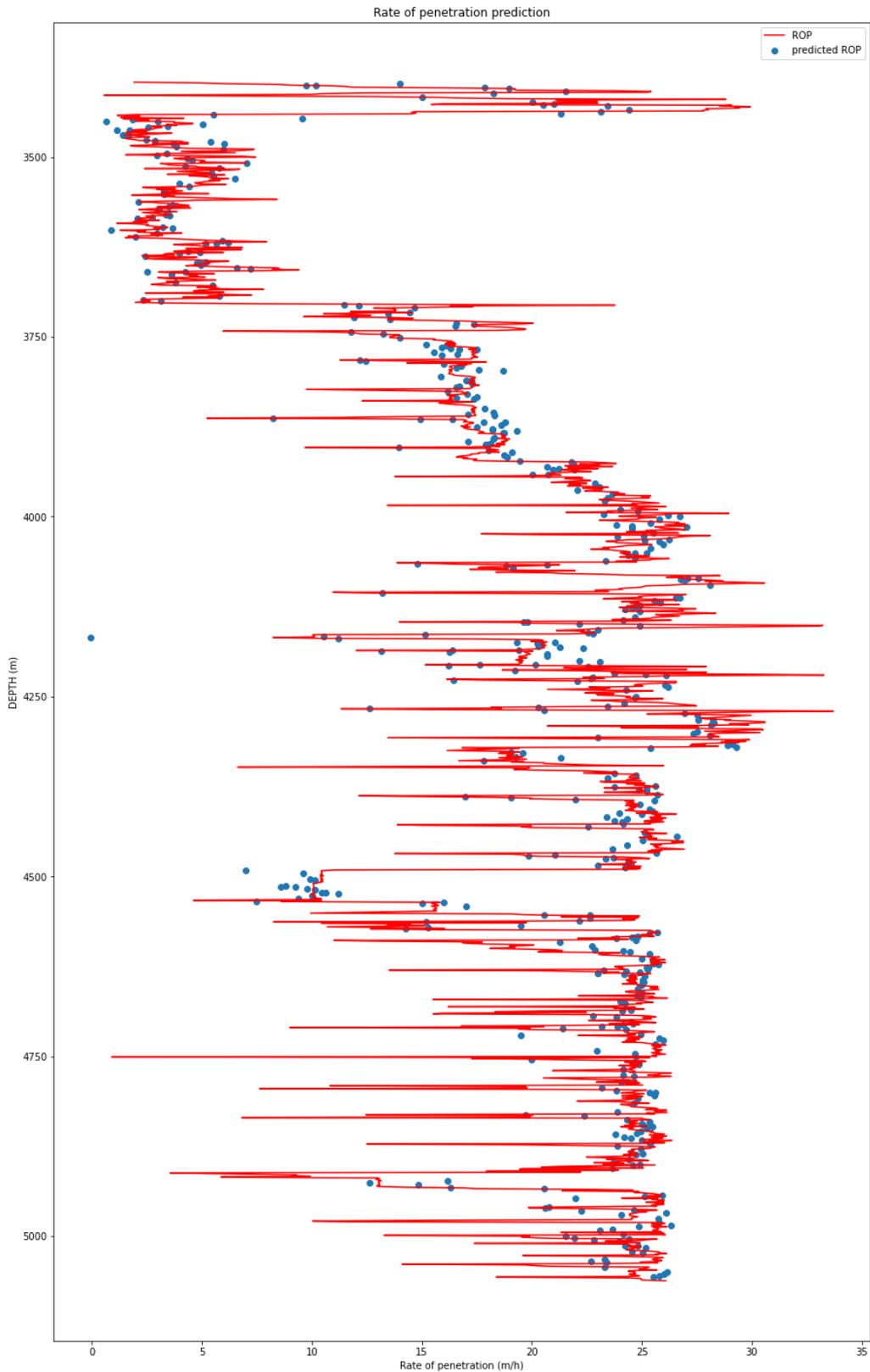


Figura 37 - ROP x profundidade (Regressão *Support Vector Machine*) sem remoção do torque superficial. Fonte: Elaboração Própria

Após treinar o modelo de predição usando Redes Neurais Artificiais MLP durante 15 segundos, totalizando 75 iterações, o erro quadrático médio do teste foi igual a 2,29.

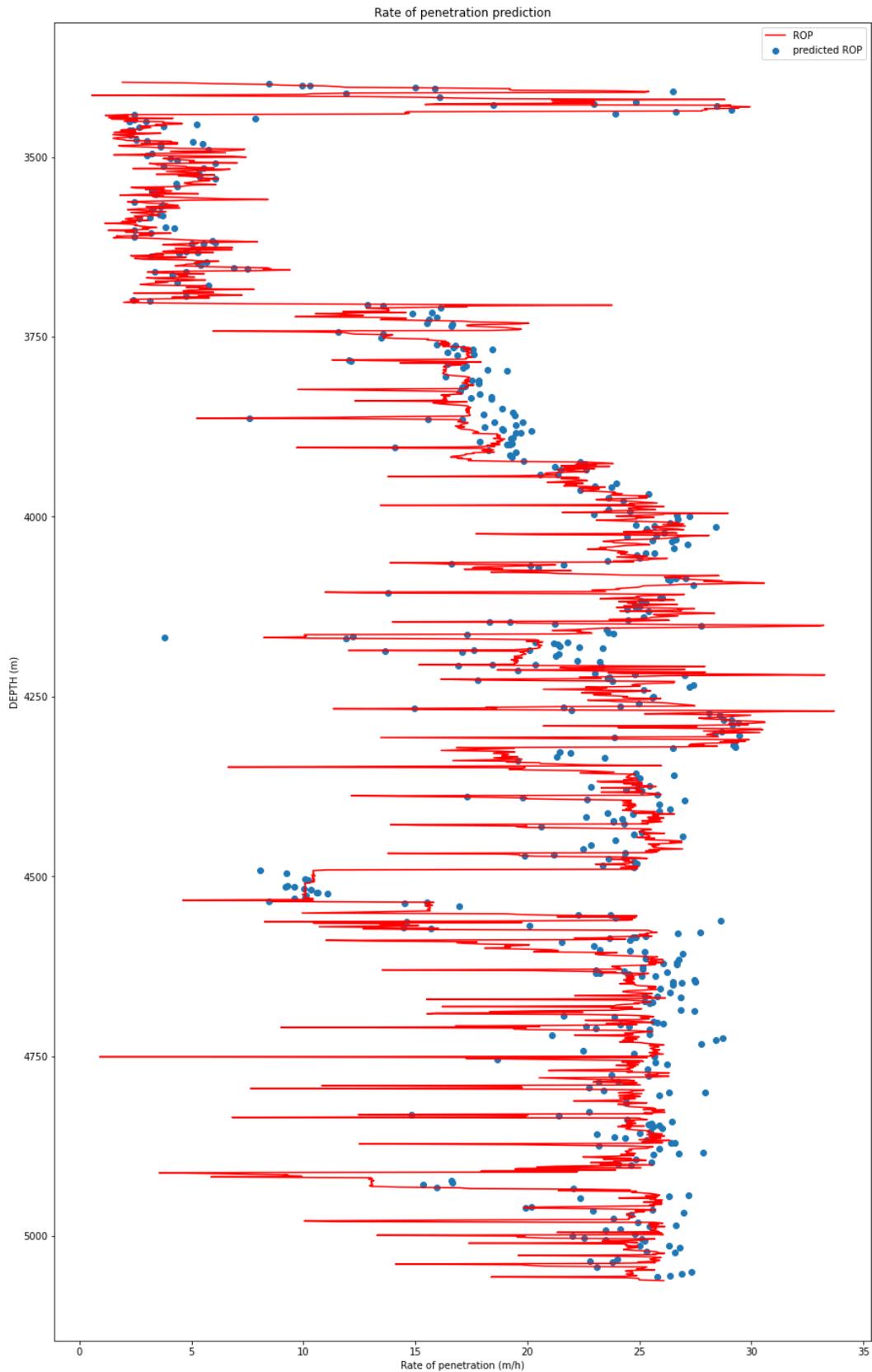


Figura 38 - ROP x profundidade (*Multiple-Layer Perceptron*) sem remoção do torque superficial. Fonte: Elaboração Própria

Assim, foi possível construir a Tabela 3, apresentada abaixo, que compara os modelos de aprendizado de máquina aplicados considerando a tensão superficial.

Tabela 3 - Avaliação entre os modelos de predição aplicados sem remoção da tensão superficial.

Modelo	Regressão <i>Random Forest</i>	Regressão <i>Support Vector Machine</i>	ANN <i>Multiple-Layer Perceptron</i>
Erro Quadrático Médio	1,88	2,15	2,29
Duração do Treinamento (s)	143	12	15

Fonte: Elaboração Própria

7. CONCLUSÃO

Com os resultados obtidos, foi possível analisar a comparação entre os valores de erro quadrático médio para cada um dos três modelos de predição utilizados após o tratamento dos dados, apresentada na Tabela 2. Nota-se que, dentre os três modelos de predição utilizados, a regressão *Random Forest* seria a mais indicada para o estudo de caso visto que seu erro quadrático médio foi o menor dentre todos. Além disso, é possível notar que tal modelo foi o mais demorado a ser executado, com uma grande diferença para os demais, ainda se comparado com o MLP, que foi capaz de realizar 145 iterações enquanto foram feitas 75 combinações com o *Random Forest*.

Após testes feitos no código, mantendo a coluna com os valores de tensão superficial na base de dados, foi possível avaliar que os erros quadráticos médios das predições realizadas diminuiriam, como apresentado na Tabela 3. A melhora dos valores de avaliação das predições com a adição da tensão superficial reforça que tal parâmetro possui grande importância para os modelos, assim como apresentado nas Figuras 33 e 34 previamente, sendo indicado mantê-lo no *data frame* para este caso.

Além disso, o tempo necessário para treinar o modelo *Random Forest* aumentou com a adição de uma coluna a mais, o que pode ser explicado pela necessidade de estabelecer mais combinações entre os parâmetros. Enquanto isso, as durações dos outros dois modelos (*Support Vector Machine* e MLP) diminuiriam em relação à execução anterior.

Foi possível notar, também, a importância da análise estatística dos dados, visto o impacto da remoção de uma única coluna nos resultados finais. Além do mais, ao realizar a remoção de *outliers* de apenas alguns parâmetros, houve melhora na distribuição da maioria dos demais. Por exemplo, após a remoção dos valores discrepantes da profundidade, ROP, quantidade de gás e gravidade rotacional contínua, grande parte dos *outliers* da carga do gancho já haviam sido removidas também.

Finalmente, os resultados obtidos, quando comparados aos valores presentes na bibliografia, indicam viabilidade de uso dos modelos para predição da taxa de penetração da formação rochosa para o estudo de caso com dados do Campo de Volve. Ainda, a

baixa duração de execução e poucos requisitos operacionais tornam a operação mais atrativa.

8. REFERÊNCIAS BIBLIOGRÁFICAS

1. AGOSTINI, Cristiano Eduardo; SAMPAIO, Marcio Augusto. **Probabilistic Neural Network with Bayesian-based, spectral torque imaging and Deep Convolutional Autoencoder for PDC bit wear monitoring**. Elsevier, Departamento de Engenharia de Minas e de Petróleo, USP, Santos, Brasil. Journal of Petroleum Science and Engineering 193 (2020) 107434. Disponível em <https://bit.ly/3uBeVTu>
2. BOURGOYNE, A.T. JR., YOUNG, F.S. JR. **A multiple regression approach to optimal drilling and abnormal pressure detection**. United States of America. SPE J. 14 (04): 371–384, SPE-4238-PA, 1974. Disponível em <https://bit.ly/3atrkFi>
3. BREIMAN, L. **Random Forests**. Machine Learning, 45, 5-32. Kluwer Academic Publishers. Statistics Department, University of California, 2001. Disponível em <https://bit.ly/3WU6RMm>
4. DUQUE, Maria Clara. JARDIM, Thonny. **Dados Volve: Guia de acesso e primeiros passos**. Instituto Alberto Luiz Coimbra de Pós-Graduação e Pesquisa de Engenharia, Universidade Federal do Rio de Janeiro (COPPE/UFRJ). Laboratório LORDE, 2020.
5. F. MACHADO, Giovani; F. ALMEIDA, Luciana; L. LAZO, Juan Guillermo. **Técnicas de aprendizado de máquina para previsão de perdas severas em rochas carbonáticas de reservatórios do pré-sal**. Brazilian Journal of Development, PUC-RJ, Rio de Janeiro, Brasil; Centro Federal de Educação Tecnológica Celso Suckow da Fonseca, Rio de Janeiro, Brasil; Universidad del Pacífico, Lima, Perú. 2021. Disponível em <https://bit.ly/3uAmJVI>
6. F. REGINATO, Leonardo; C. CARNEIRO, Cleyton; S. GIORIA, Rafael; S. PINTO, Marcio. **Prediction of Wettability Alteration Using the Artificial Neural Networks in the Salinity Control of Water Injection in Carbonate Reservoirs**. Offshore Technology Conference, Rio de Janeiro, Brasil. 2019. Disponível em <https://bit.ly/34vIVGX>
7. GAN, C., CAO, W., LIU, K., WU, M. **A novel dynamic model for the online prediction of rate of penetration and its industrial application to a drilling process**. Wuhan, China. Journal of Process Control 109 (2022) 83-92. Disponível em <https://bit.ly/3wIDpNW>
8. HOHN, P., ODEBRETT, F., SHAHID, K., PAZ, C., OPPELT, J. **Framework for automated generation of real-time rate of penetration models**. Lower Saxony, Germany. Journal of Petroleum Science and Engineering 213 (2022). Disponível em <https://bit.ly/3wIDqS0>
9. LINDEN, Ricardo. **Algoritmos Genéticos: Uma importante ferramenta da Inteligência Computacional**. Rio de Janeiro: Editora Brasport, 2006.
10. MOTAHHARI, H.R., HARELAND, G., JAMES, J.A. **Improved drilling efficiency technique using integrated PDM and PDC bit parameters**.

- Calgary, Canada. Journal of Canadian Petroleum Technology 49 (10): 45–52, SPE-141651-PA, 2010. Disponível em <https://bit.ly/3anedFs>
11. OSSAI, C., DURU U. **Applications and theoretical perspectives of artificial intelligence in the rate of penetration**. Owerri, Nigeria. KeAi Advancing Research Evolving Science, 2020. Disponível em <https://bit.ly/3bYk3xs>
 12. PARK, Junho; PRICE, Cameron; PIXTON, David; AGHITO, Manuel; NYBO, Roar; BJORKEVOLL, Knutt; D. HEDENGREN, John. **Model predictive control and estimation of managed pressure drilling using a real-time high fidelity flow model**. Elsevier, Department of Chemical Engineering, Brigham Young University, Provo, Utah, USA. 2020. Disponível em <https://bit.ly/3c6valb>
 13. PEPPER, N., CRESPO, L., MONTOMOLI, F. **Adaptive learning for reliability analysis using Support Vector Machines**. United Kingdom. Reliability Engineering and System Safety 226, 2022. Disponível em <https://bit.ly/3GIA6Bc>
 14. RAVASI, M., VASCONCELOS, I., CURTIS, A., KRIST, A.: Vector-acoustic reverse time migration of Volve ocean-bottom cable data set without up/down decomposed wavefields, Geophysics, 80, 137–150, 2015. Disponível em <https://bit.ly/3WPFnrl>
 15. ROOKI, Reza. **Application of general regression neural network (GRNN) for indirect measuring pressure loss of Herschel–Bulkley drilling fluids in oil drilling**. Elsevier, Birjand University of Technology, Birjand, Iran. 2016. Disponível em <https://bit.ly/2S0jBFc>
 16. R. GUILHERME, Ivan; N. MARANA, Aparecido; P. PAPA, João; CHIACHIA, Giovanni; C.S. AFONSO, Luis; MIURA, Kazuo; V.D. FERREIRA, Marcus; TORRES, Francisco. **Petroleum well drilling monitoring through cutting image analysis and artificial intelligence techniques**. Elsevier, Departamento de Estatística, Matemática Aplicada e Computação, UNESP, Rio Claro, Brasil; Departamento de Computação, UNESP, Bauru, Brasil; CENPES, PETROBRAS, Brasil. 2010. Disponível em <https://bit.ly/3fTlpH7>
 17. SANTANA, Alan Felipe. **Treinamento de redes neurais artificiais utilizando algoritmos genéticos em plataforma distribuída**. Monografia apresentada à Universidade de Lavras, Minas Gerais, Brasil. 2012. Disponível em <https://bit.ly/2SKbJrp>
 18. TOORAJIPOUR, Reza; SOHRABPOUR, Vahid; NAZARPOUR, Ali; OGHAZI, Pejvak; FISCHI, Maria. **Artificial intelligence in supply chain management: A systematic literature review**. Elsevier, School of Innovation, Design and Engineering, Malardalen University, Eskilstuna, Sweden; Department of Operations Management, Copenhagen Business School. Copenhagen, Denmark; SAVEGGY AB, Lund, Sweden; School of Business, Maynooth University, Maynooth, Co. Kildare, Ireland; School of Social Sciences, Sodertorn University, Stockholm, Sweden; Siemens Gas and Power, Berlin, Germany. Journal of Business Research 122 (2021) 502-517. Disponível em <https://bit.ly/3c4Qbww>

19. VAPNIK, V. N. **An overview of statistical learning theory**. IEEE Transactions on Neural Networks, Vol. 10, Nº 5. September, 1999. Disponível em <https://bit.ly/2JtXBNM>
20. **Volve**. Norwegian Petroleum, Norwegian Ministry of Petroleum and Energy. Acesso em 7 de novembro de 2022. Disponível em <https://bit.ly/3TlxWG1>
21. **Volve Oil Field, North Sea**. Offshore Technology, 2013. Acesso em 7 de novembro de 2022. Disponível em <https://bit.ly/3fKTGgK>
22. **15/9-F-10**. Norwegian Petroleum Directorate, 2009. Acesso em 7 de novembro de 2011. Disponível em <https://bit.ly/3G2bIFY>