

# Aplicação das Redes Neurais Artificiais para classificação das operações de perfuração: O caso de poços deepwater de Exploração e Produção

Valter Chaile<sup>1</sup>, Sergio Moro<sup>2</sup>, Aristides Carneiro<sup>3</sup>, Ricardo F. Ramos<sup>2,4,5</sup>

[vjchaile@gmail.com](mailto:vjchaile@gmail.com); [scmoro@gmail.com](mailto:scmoro@gmail.com); [aristides.carneiro@galp.com](mailto:aristides.carneiro@galp.com);  
[ricardo\\_ramos@iscte-iul.pt](mailto:ricardo_ramos@iscte-iul.pt)

<sup>1</sup> ISCTE-IUL, Instituto Universitário de Lisboa, Av. das Forças, 1649-026, Lisboa, Portugal

<sup>2</sup> Instituto Universitário de Lisboa (ISCTE-IUL), ISTAR-IUL, Lisboa, Portugal;  
Av. Forças Armadas, 1649-026 Lisboa, Portugal

<sup>3</sup> Galp, Rua Tomas da Fonseca – Torre A, 1600-209, Lisboa, Portugal

<sup>4</sup> Instituto Politécnico de Coimbra, ESTGOH, Rua General Santos Costa, 3400-124 Oliveira do Hospital

<sup>5</sup> CICEE – Centro de Investigação em Ciências Económicas e Empresariais, Universidade Autónoma de Lisboa, Portugal; Address: Rua de Santa Marta, Palácio dos Condes do Redondo 56, 1169-023, Lisboa. Portugal

DOI: [10.17013/risti.43.5-20](https://doi.org/10.17013/risti.43.5-20)

**Resumo:** A aplicação de métodos automáticos para classificação de texto não estruturadas são extremamente valiosas para a indústria de *Oil&Gas*. A perfuração é uma operação que acarreta custos elevados que são proporcionais à duração das atividades. A classificação das diversas operações durante a perfuração é muito importante para gerar premissas de duração para o projeto de novos poços. Para este artigo, dois procedimentos independentes foram realizados para identificar o melhor modelo de NPT (*Non-Productive Time*) e PT (*Productive Time*). As conclusões apontam o modelo *Multi-layer Perceptron* (MLP) como o melhor modelo. O sistema de classificação pode ser utilizado para produzir um relatório preciso e detalhado sobre as atividades realizadas durante a perfuração de um poço. Através desse trabalho é possível concluir que os relatórios diários de perfuração atualmente disponíveis representam uma fonte rica de informação e podem ser utilizados para melhorar o processo de construção de poços de petróleo.

**Palavras-chave:** redes neuronais artificiais; inteligência artificial; classificação; aprendizagem de máquina; perfuração; completação.

## ***Application of Artificial Neural Networks for Classification of Drilling Operations: The deepwater wells case of exploration and production***

**Abstract:** The application of automatic methods for the classification of unstructured text is precious for the *Oil&Gas* industry. Drilling is an operation that entails high costs that demands efficiency. A classification of the various operations during drilling is vital to generate assumptions of duration for the design of new wells. For this paper, two classification analyses for operation classification were

conducted to identify the Non-Productive Time (NPT) and Productive Time (PT) best model. Conclusions led to Multi-layer Perceptron (MLP) as the best model. The classification system can produce an accurate and detailed report on the activities performed during the drilling of a well. Through this work, it is possible to conclude that the currently available daily drilling report represents a rich source of information and can optimize the oil well construction process.

**Keywords:** Artificial Neural Network; Artificial Intelligence; Classification; Machine Learning; Drilling; Completion.

## 1. Introdução

As atividades de perfuração na indústria de petróleo e gás são uma preocupação partilhada entre as *Joint Ventures*, agências governamentais e o público em geral (Silva, Costa, & Barreiros, 2006). Estas podem afetar a rentabilidade das empresas e o seu meio ambiente (Hoffmann, Mao, Wesley, & Taylor, 2018). As operações de perfuração envolvem um planeamento e execução detalhados para obter uma exploração viável do poço. O planeamento permitirá atingir a eficiência da perfuração, minimizando o custo de tempo improdutivo (NPT – *Non-Productive Time*).

Os programas de vigilância ativos utilizam grandes volumes de dados gerados em vários estágios do processo de perfuração (Silva, Martins, Doria Neto, Rodrigues, & Da Mata, 2016). Apesar do aumento da confiança nas medições quantitativas em tempo real no local da perfuração, os dados gerados em relatórios diários de perfuração (DDR) e gravações de atividades não têm sido devidamente aproveitados para uma melhoria do processo de perfuração. Estas fontes não são exploradas para monitorização e alerta por ferramentas convencionais. Nesse sentido, através dos dados gerados em DDR, o objetivo deste estudo passa por construir um modelo de Inteligência Artificial (AI) usando Redes Neurais Artificiais (ANN) para classificar as operações de perfuração de um poço de petróleo.

O desenvolvimento bem-sucedido de novos campos de petróleo e a extensão da vida útil dos campos de produção são o resultado de programas de perfuração eficiente (Sidahmed, Coley, & Shirzadi, 2015). A institucionalização de tais programas diminui a estrutura de custos substancialmente através do processo sistemático e advento de novas tecnologias. Ao longo da última década, o setor reconheceu o valor de automatizar partes significativas das tarefas de perfuração para manter eficiência, reduzir a sobrecarga e superar o erro humano (Iversen, Gressgård, Thorogood, Balov, & Hepsø, 2013).

O objetivo deste estudo é desenvolver um classificador através de aprendizagem automática que permita classificar se um poço em prospeção tem ou não petróleo. Para atingir o objetivo definido, foram conduzidas duas análises de classificação de operações para identificar as atividades de NPT e tempo produtivo (PT – *Productive Time*). A primeira, usando as redes *Multi-Layer Perceptron* (MLP). A segunda, usando as redes *Long Short-term Memory* (LSTM). Posteriormente, os dois modelos foram comparados com o objetivo de identificar o modelo com melhores resultados.

O sistema de classificação pode ser utilizado para produzir um relatório preciso e detalhado sobre as atividades realizadas durante a perfuração de um poço. Através

deste, será possível identificar as atividades que gastam mais tempo na sonda e otimizar os custos em futuros poços.

## 2. Revisão de literatura

A classificação automática de texto consiste na tarefa de atribuir uma ou mais categorias a um documento ou a um conjunto de dados eletrónicos com base no seu conteúdo. Atualmente, a classificação é amplamente utilizada em muitos contextos. As abordagens por padrões de categorização utilizam estatística ou métodos de IA para executar uma tarefa (Sanchez-Pi, Martí, & Garcia, 2014). Os métodos mais usados podem ser ANN, máquinas de Vetores de Suport (SVM), lógica Fuzzy, algoritmos genéticos (GA) e *K-Nearest Neighbor* (Han, Kamber, & Pei, 2011). Todos estes métodos requerem um conjunto de treino de documentos pré-classificados que são usados para atribuir classificações automaticamente (Sanchez-Pi et al., 2014).

A perfuração de um poço de petróleo é feita a partir da utilização de uma sonda de perfuração que, por definição, consta como um conjunto de sistemas, equipamentos e ferramentas que tem a principal finalidade de perfurar diversas camadas de rochas, até encontrar um reservatório de petróleo (Thomas, 2001). O percurso criado através das camadas geológicas (poço) irá ligar o reservatório à superfície, permitindo assim a correta drenagem dos hidrocarbonetos. Os eventos que podem ocorrer durante a operação de perfuração de um poço de petróleo são a perda de circulação, prisão da coluna de perfuração, desmoronamento do poço, alargamento do poço, influxo de fluidos indesejados (*kick*), falha no BHA (vibração), entrada de cascalhos nos jatos quando a circulação é interrompida (*pack-off*), falta de hidratação de argilas (*bit-balling*) e vazamento no sistema de circulação (*washout*) (Heriot Watt, 2013).

### 2.1. Evolução da IA na indústria de Oil&Gas

A IA é uma área de grande interesse e importância para o setor de *Oil&Gas*. Esta tecnologia tem provocado grandes impactos na indústria e a sua aplicação tem continuado a crescer (Bello, Holzmann, Yaqoob, & Teodoriu, 2015). A aplicação de IA nesta indústria tem mais de 30 anos, sendo voltada para a interpretação de *logs* de poços, diagnóstico de brocas usando ANN e interfaces inteligentes de simuladores de reservatórios (Bello et al., 2015). Os valores mensuráveis da IA incluem a possibilidade de fazer máquinas resolverem problemas difíceis para entender tendências e fazer previsões de cenários futuros com o menor esforço e desperdício de tempo (Rable, 2017).

### 2.2. Aplicação da IA na engenharia de perfuração

A classificação de textos através de técnicas de IA criou desafios. Estas técnicas estão concentradas na forma como os relatórios estão escritos. Em muitos casos, não existe um padrão a ser seguido pelas empresas de *Joint Venture*. De modo a identificar os estudos mais relevantes realizados no âmbito do objeto de estudo, foi aplicada uma query na base de dados Scopus a partir do ano 2000. O resultado devolveu 50 artigos científicos. Através de uma análise cuidada, foram excluídos aqueles que não estavam relacionados com a aplicação de abordagens de dados para o problema específico da perfuração de petróleo, conduzindo a 15 artigos (Tabela 1).

Através desta tabela, podemos verificar que vários autores se concentraram em técnicas de regressão para resolver problemas ou estimar resultados de perfuração. Nenhuma técnica de classificação de operações de perfuração é aplicada.

De acordo com análise da Tabela 1, podemos concluir que cerca de 75% dos estudos estão relacionados com o uso de ANN (50%) e SVM (25%), seguidos da lógica *Fuzzy* (12%) e GA (13%). Tais resultados justificam a escolha de ANN para treinar o classificador. Adicionalmente, em comparação com técnicas mais tradicionais para classificação como regressão logística e árvores de decisão, as ANN tipicamente obtêm melhores resultados (Moro et al., 2018). No entanto, poucos estudos apresentaram resultados utilizando a ANN com dados extraídos de DDRs.

Para além disto, verificou-se que os estudos apresentados se concentram maioritariamente em estimar a velocidade em operações de perfuração de um poço de petróleo, não observado a classificação automática das operações de perfuração com o objetivo de estimar a duração de perfuração de poços. Nesse sentido, propõe-se um novo estudo com base na topologia de ANN. Das técnicas de IA destacadas, é pertinente questionar se há alguma que possa ser 100% fiável e adequada para o uso em todas as circunstâncias. A Tabela 2 resume os pontos fortes e fracos das técnicas de IA (Agwu, Akpabio, Alabi, & Dosunmu, 2018).

Título	Referência	Objetivo	Método	Resultado
Determinação da alteração da densidade da lama de perfuração com pressão e temperatura simplificadas e precisas	(Osman & Aggour, 2003)	Fornecer previsões precisas da densidade da lama em função do tipo, pressão e temperatura da lama	ANN	$R^2 = 0.9998$
Abordagem de ANN para estimar propriedades de filtragem de fluidos de perfuração	(Jeirani & Mohebbi, 2006)	Estimar o volume de filtro e a permeabilidade do bolo de filtro usando os dados de filtragem estática.	ANN	$R^2$ (Volume do filtro) = 0.9815 $R^2$ (Permeabilidade) = 0.9433
Previsão e Prevenção de Tubos Presos: Uma Abordagem de Rede Neural Convolucional	(Siruvuri, Nagarakanti, & Samuel, 2006)	apresentar uma aplicação de métodos de IA para entender e estimar a ocorrência de tubos diferencialmente presos durante a perfuração.	ANN	$R^2$ (Tubos presos) = 0.063 $R^2$ (Tubos não presos) = 0.1619
Estimando padrões de fluxo e perdas de pressão por atrito de fluidos bifásicos em poços horizontais usando ANN	(Ozbayoglu & Ozbayoglu, 2009)	Estimar os padrões de fluxo e as perdas por pressão de atrito de fluidos bifásicos que fluem através de geometrias anulares horizontais usando ANN, em vez de usar modelos mecanicistas convencionais.	ANN	MSE=0.006 [FPL with BP] MSE=0.005 [FPL with J/E] MSE=0.005 [FP with BP] MSE=0.005 [FP with J/E]

<b>Título</b>	<b>Referência</b>	<b>Objetivo</b>	<b>Método</b>	<b>Resultado</b>
Tomada de decisão para redução do tempo improdutivo por meio de uma previsão integrada de circulação perdida	(Moazzeni, Nabaei, & Jegarluei, 2012)	Prever a gravidade da perda de lama durante a perfuração ao longo de diferentes setores do campo petrolífero.	ANN	$R^2 = 0.82$
Pesquisa de colagem de tubos com pré-aquecimento baseada em rede neural	(Zhu, Liu, & Zhang, 2013)	Propor o uso da tecnologia de IA para realizar o pré-aviso de acidente de tubos presos durante a perfuração.	ANN	-
Novo método para prever e resolver o problema da perfuração e perda de fluidos usando ANN modular e exame de partículas algoritmo de otimização	(Toreifi, Rostami, & Manshad, 2014)	Prever a perda de circulação durante a perfuração em qualidade e quantidade.	ANN	$R^2 = 0.94$
Previsão em tempo real de parâmetros reológicos do fluido de perfuração à base de água KCl usando ANN	(Elkakatny, 2017)	Usar as frequentes de medições de densidade da lama, viscosidade do funil de Marsh e percentagem sólida para prever as propriedades reológicas desenvolvendo correlações empíricas	ANN	AAPE < 6% $R^2 > 0.90$
Previsão e prevenção de aderência de tubulação usando modelagem lógica difusa adaptativa	(Murillo, Neuman, & Samuel, 2009)	Estimar o risco de ocorrência de tubos presos no procedimento de planeamento de poços e durante a perfuração em tempo real	Lógica Fuzzy	-
Um modelo preciso para prever a densidade do fluido de perfuração em condições de poço	(Ahmadi, Shadizadeh, Shah, & Bahadori, 2018)	Sugerir um modelo preditivo rigoroso para estimar a densidade do fluido de perfuração (g / cm <sup>3</sup> ) em condições de poço	Logica Fuzzy	$R^2 = 0.7237$ MSE = 69.0907
Uma abordagem de aprendizagem de máquina para a previsão de settling	(Goldstein & Coco, 2014)	Utilizar uma abordagem de aprendizagem de máquina baseada em programação genética para prever a velocidade de assentamento de partículas não coesas.	GA	RMSE = 0.26 $R^2 = 0.97$

Título	Referência	Objetivo	Método	Resultado
Determinação ideal de parâmetros reológicos para fluidos de perfuração de espigão-sela usando GA	(Rooki et al., 2012)	Determinar comportamento reológico não newtoniano de fluidos de perfuração, a fim de determinar os três parâmetros do modelo de Herschel-Bulkley com mais precisão.	GA	$R^2 = 0.9972$
Uma abordagem de SVM para a previsão da densidade do fluido de perfuração em alta temperatura e pressão	(Wang, Pu, & Tao, 2012)	Prever a densidade do fluido de perfuração em alta temperatura e pressão (HTHP).	SVM	MAPE = 0.872 $R^2 = 0.9994$
Aplicação do algoritmo SVM para o cálculo da perda por pressão por atrito do fluxo trifásico em anéis inclinados	(Shahdi & Arabloo, 2014)	Uso de Lease Square (LS-SVM), para cálculo de perdas por atrito de fluidos de perfuração bifásicos baseados em gás.	SVM	$R^2 = 0.9862$
Estimação da densidade do fluido de perfuração na tecnologia de lama: Aplicação em poços de petróleo de alta temperatura e alta pressão	(Kamari, Gharagheizi, Shokrollahi, Arabloo, & Mohammadi, 2017)	Desenvolver um modelo confiável para prever a densidade de quatro fluidos de perfuração, incluindo à base de água, à base de óleo, Coloidal Gás Aphron (CGA) e sintético.	SVM	$R^2 = 0.999$

Tabela 1 – Técnicas de IA aplicadas na engenharia de perfuração

### 3. Metodologia

O estudo é desenvolvido segundo a metodologia *Cross-Industry Standard Process of Data Mining* (CRISP-DM) (Laureano, Caetano, & Cortez, 2014). A escolha deste modelo deve-se ao facto de ser considerado o padrão de maior aceitação e por ter sido usado para problemas semelhantes. A metodologia CRISP-DM é um modelo que possui processos iterativos, com sequências não mandatárias, possuindo um ciclo de vida, que ocorre nas fases que têm as suas tarefas (Sergio Moro, Laureano, & Cortez, 2011).

Foi utilizado o *Spyder* 4.0 (do projeto Anaconda) como IDE<sup>1</sup> para o desenvolvimento do projeto (Smith, 2020).

<sup>1</sup> IDE (Integrated Development Environment), ou ambiente de desenvolvimento integrado, é um software que combina ferramentas comuns de desenvolvimento numa interface gráfica do utilizador (GUI), facilitando o desenvolvimento de aplicações (Chapman et al., 2000).

Em termos de bibliotecas, foram utilizadas o *Keras* e o *TensorFlow*, configurado para uso de GPUs. Adicionalmente, foi adotado o *scikit-learn* para os algoritmos de *Machine Learning* (MLP; LSTM) e validação cruzada.

### 3.1. As operações de perfuração e classificação

Durante a fase de perfuração de um poço de petróleo, compete aos engenheiros de perfuração fazer um relatório diário de operações com o objetivo de controlar o monitorizar o processo de perfuração. Através da inspeção dos relatórios, é possível identificar operações que estão a consumir tempo excessivo de sonda e, a partir dessa observação, adotar medidas que melhorem a operação de perfuração. O sistema de classificação proposto identifica a operação que está a ser executada através da interpretação de dados.

Atualmente, o registo de operações realizadas é feito através do DDR. O DDR é um relatório preenchido diariamente que descreve as operações executadas nas últimas 24 horas. A descrição das operações é feita em linguagem natural e o responsável pelo preenchimento elabora um pequeno texto no qual descreve de forma resumida as atividades executadas. Além da descrição, existe um sistema de codificação que classifica a atividade.

### 3.2. Construção do dataset

Antes de proceder ao treino do modelo, foi necessário recolher os dados. Esses dados foram extraídos do DDR (Figura 1) através de um script desenvolvido em linguagem VBA no MS excel. Depois da construção do *dataset*, foram obtidos 20.390 registos de entrada. Os dados foram exportados para uma lista de Python. A Tabela 3 resume as variáveis usadas para o presente estudo.

DESCRIÇÃO DAS OPERAÇÕES						
DAS	ATÉ	HORAS	DE (m)	ATÉ (m)	ANORM	DESCRIÇÃO DAS OPERAÇÕES
05:00	07:00	2	0,00	0,0		Movendo e cortando 112 ft de cabo de perfuração.
07:00	08:00	1	0,00	0,0		Deitando restante BHA 8 1/2", posicionando na plataforma MF e HO 36", montando e estaleirando RT do alojador de 30".
08:00	10:30	2,5	0,00	0,0		Descendo BHA c/ br 26" / HO 36" / MF 9 5/8", testando MF.
10:30	14:30	4	0,00	0,0		Descendo coluna c/ br 26" / HO 36".
14:30	18:00	3,5	0,00	0,0		Posicionando coluna com BR/HO na locação, verificando coordenadas.

Figura 1 – Exemplo de um DDR

#	Column	Not-Null Count	Dtype
0	<i>PDF</i>	20390 non-null	object
1	<i>Type</i>	20390 non-null	object

Tabela 2 – Resumo das variáveis



A variável PDF representa os dados extraídos do DDR. Trata-se da variável dependente que contém o texto que ajuda a classificar as operações. Cada linha desta variável representa uma atividade e é com base nestas atividades que os engenheiros conseguem classificar as operações de perfuração.

### 3.3. Modelo de redes neuronais

O modelo de ANN tenta imitar processos simplificados de aprendizagem biológica e simular algumas funções do sistema nervoso humano. Uma ANN consiste em unidades de processamento simples, chamadas de neurónios (Bishop, 1996). As ANNs consistem num modo de abordar problemas de AI (Barreto, 1997). As ANN provaram fornecer um alto nível de competência na solução de muitos problemas complexos de engenharia que estão além da capacidade computacional da matemática clássica e dos procedimentos tradicionais (Agwu et al., 2018).

Para tornar o trabalho robusto, os modelos de ANN escolhidos foram os que obedecem a critérios de multicamadas: redes recorrentes (*recurrent*) e redes de propagação para frente (*feedforward*), designados de LSTM e MLP, respetivamente. A escolha destes modelo deveu-se ao facto de outros modelos de ANN, como ADALINE e *Perceptron*, apresentarem resultados com pouca relevância em comparação com os modelos escolhidos (Jiang, Tang, Chen, Wang, & Huang, 2019). O estudo obedeceu a dois cenários diferentes (MLP e LSTM), com experimentação de cada atributo classificador para cada modelo de treino.

### 3.4. Multi-Layer Perception

O modelo MLP proposto é de três camadas com oito neurónios para cada camada. O número de camadas depende da complexidade do problema em estudo (Khan & Yu, 2016). Redes maiores, compostas por muitas camadas, conseguem “aprender” mais padrões. No entanto, assume um custo computacional elevado para além de que o *overfitting*<sup>2</sup> pode surgir. Para o treino do modelo MLP foram definidos os parâmetros de entrada da Tabela 4.

Parâmetro	Descrição	Valores
<i>hidden_layer_sizes</i>	Número de neurónios na enésima camada oculta	(8, 8, 8)
<i>alpha</i>	Constante que multiplica o termo L1.	1e-5
<i>max_iter</i>	O número de iterações	200
<i>solver</i>	Tamanho do mini bach	lbfgs (é um otimizador na família de métodos quasi-Newton.)
<i>Activation</i>	Ativação para a camada oculta	Relu (a função de unidade linear retificada, retorna $f(x) = \max(0, x)$ )

<sup>2</sup> *Overfitting* ocorre quando o modelo se adaptou muito bem aos dados. Porém, não generaliza corretamente para novos dados. Ou seja, o modelo “decorou” o conjunto de dados de treino, mas não aprendeu o que diferencia os dados de novos testes (Kohn & Manaris, 2020).



Parâmetro	Descrição	Valores
<i>Verbose</i>	Níveis de log WARNING e INFO	True
<i>random_state</i>	Determina a geração de números aleatórios para inicialização de pesos e desvios	40

Tabela 3 – Parâmetros de entrada MLP

### 3.5. Long Short-Term Memory

O modelo LSTM é de 64 camadas e épocas igual a 60. Em ANN, uma época corresponde a uma passagem completa pelo conjunto de treino. Para o treino do modelo LSTM foram definidos os parâmetros indicados na Tabela 5.

Parâmetro	Descrição	Valores
<i>input_length</i>	Tamanho do vetor de sentença de entrada	300
<i>units</i>	Quantidade de células na camada LSTM	64
<i>dropout_rate</i>	Taxa de dropout da camada de entrada	0.5
<i>batch_size</i>	número de amostras a serem utilizadas em cada atualização do gradiente	128
<i>optimizer</i>	Tipo de otimizador	adam
<i>epochs</i>	Número de épocas	10
<i>word_embedding_dim</i>	dimensionalidade do word embedding pré-treinado	50
<i>max_fatures</i>	Quantidade máxima de palavras mantidas no vocabulário	5000
<i>embed_dim</i>	dimensão de saída da camada Embedding	128
<i>loss</i>	calcula a quantidade que um modelo deve procurar minimizar durante o treinamento.	binary_crossentropy

Tabela 4 – Parâmetros de entrada LSTM

### 3.6. Avaliação

Para avaliar os modelos de classificação as seguintes métricas foram adotadas (Ian Witten, Eibe Frank, Mark Hall, 2016): precisão (1), sensibilidade (2), F1-Score (calculado a partir da precisão e do recall) (3) e curva ROC (Receiver operator characteristic curve). TP, FP e FN correspondem ao número de verdadeiros positivos, falsos positivos e falsos negativos, respetivamente. A precisão é intuitivamente a capacidade do classificador de não rotular como positiva uma amostra negativa.

$$\text{Precisão} = \frac{TP}{TP + FP}$$

A sensibilidade é a capacidade de o classificador de encontrar todas as amostras positivas.

$$Sensibilidade = \frac{TP}{TP + FN}$$

O F1-score, pode ser interpretada como uma média ponderada da precisão e recuperação

$$F_1 = 2 \times \frac{Precisão \times Sensibilidade}{Precisão + Sensibilidade}$$

A sensibilidade e a especificidade são características difíceis de conciliar, sendo complicado aumentar a sensibilidade e a precisão de um teste simultaneamente. As curvas ROC representam a relação entre a precisão e a sensibilidade de um teste diagnóstico quantitativo. Outra forma de avaliar o modelo é através da *accuracy* (ACC). É comum interpretar a qualidade dos valores da ACC como: 0,5 – igual a um classificador aleatório; 0,6 – razoável; 0,7 – bom; 0,8 – muito bom; 0,9 – excelente; 1 – perfeito (Landis & Koch, 1977).

## 4. Resultados e discussão

### 4.1. Classificador Multi-Layer Perceptron

O ACC foi 98%, com uma precisão de acerto de 98% para o NPT e 97% para o PT (Tabela 6).

	Precisão	Sensibilidade	F1-Score	Support
<i>NPT</i>	0.98	0.91	0.94	3118
<i>PT</i>	0.97	0.99	0.98	12205
<i>ACC</i>			0.98	16334
<i>Macro avg</i>	0.98	0.95	0.98	16334
<i>Weighted</i>	0.97	0.97	0.97	16334

Tabela 5 – Accuracy do modelo MLP

### 4.2. Long Short-Term Memory

O ACC foi de 91%, com uma precisão de acerto de 93% para o NPT e de 91% para o PT (Tabela 7).

	Precisão	Sensibilidade	F1-Score	Support
<i>NPT</i>	0.93	0.71	0.80	4108
<i>PT</i>	0.91	0.98	0.94	12226
<i>ACC</i>			0.91	16334
<i>Macro avg</i>	0.92	0.85	0.97	16334
<i>Weighted</i>	0.92	0.91	0.90	16334

Tabela 6 – Accuracy do Modelo LSTM

### 4.3. Análise dos modelos

A Tabela 8 apresenta um resumo dos resultados dos modelos.

Modelo	ACC	Precisão	Sensibilidade	F1-Score
<i>MLP</i>	0.98	0.98	0.95	0.96
<i>LSTM</i>	0.91	0.86	0.85	0.87

Tabela 7 – Resumo da análise do Modelo

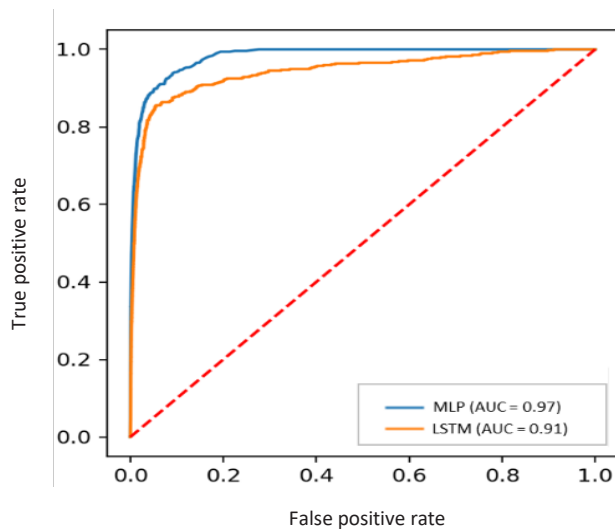


Figura 2 – Curva ROC dos modelos MLP e LSTM

Face aos resultados obtidos, constatou-se que o modelo que apresentou melhor performance foi o modelo MLP (Figura 2). Por este motivo, o modelo MLP foi o escolhido para o presente estudo.

### 4.4. Interpretação do modelo

A curva ROC possibilita determinar a capacidade de previsão do modelo medida em termos dos acertos na classe real. A diagonal a tracejado representa um modelo aleatório. Logo, quanto maior for a área debaixo da curva ROC, melhor a capacidade do modelo de distinguir a classe objetivo.

Este estudo propõe modelos de previsão que, sendo ANN, não podem ser diretamente interpretados (Sérgio Moro, Cortez, & Rita, 2018). Como tal, para validar o modelo, foi usado o *package Local Interpretable Model-agnostic Explanation* (LIME) que permite validar a classificação de modelos tipo *black-box*. O LIME é um algoritmo que tem a

capacidade de explicar as previsões de problemas de classificação ou regressão de maneira fiel, através de um modelo interpretável (Ribeiro, Singh, & Guestrin, 2016).

No modelo MLP, o LIME previu com uma certeza de 100% a classificação de NPT. Isto significa que há determinadas palavras que foram identificadas na MLP e que univocamente permitem classificar o NPT versus o PT. Portanto, relatórios que contenham as palavras “vazamento”, “apresentou”, “localizado”, “indicativo” e “queda”, claramente identificam tempo não produtivo (Figura 3). Estas palavras permitem a equipa de perfuração segregar os problemas e entender onde ocorrem e se aparecem nas mesmas frases. Se se entender que estas palavras estão sempre associadas à mesma frase, podem permitir à equipa de perfuração rever o programa e promover melhorias no processo.

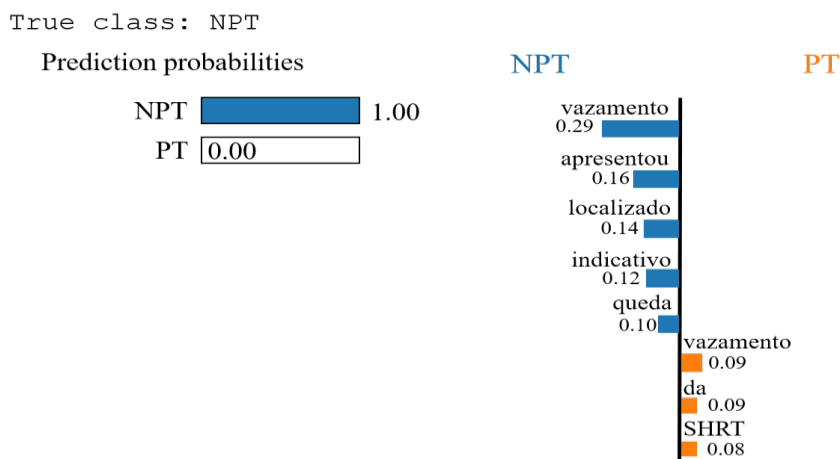


Figura 3 – Validação do modelo

## 5. Conclusões

O objetivo deste estudo passou por construir um modelo de IA usando ANN para classificar as operações de perfuração num poço de petróleo. Para atingir este objetivo, foram testados dois modelos para identificar o modelo com melhores resultados, validando o modelo ANN-MLP. Submetidas as análises de validação deste modelo através de gráficos ACC e de uma validação através do *package* LIME para garantir a interpretação do modelo, este foi considerado aceite para classificação de operações de perfuração. No que diz respeito ao desempenho do modelo MLP, os resultados obtidos nas métricas de precisão, sensibilidade e F1-score são satisfatórios e com uma avaliação de excelente (Landis & Koch, 1977).

O resultado deste estudo pode ser considerado uma vantagem competitiva para uma empresa que opera na área de *Oil&Gas*, pois consegue classificar as operações de perfuração, tornando o processo mais eficiente.

A descrição fornecida pelo modelo permite identificar eventos que estão a consumir tempo excessivo de sonda e contribui para o processo de minimização do NPT na perfuração do poço. Este modelo não só poderá trazer benefícios na redução de horas

de trabalho à equipa de *Drilling & Completion* (D&C) e economizar recursos financeiros às empresas de *Oil&Gas*. Segundo números fornecidos pela equipa de D&C da Galp Exploração e Produção, em média, uma classificação normal é feita em cerca de 45 minutos, divididos em 15 minutos para leitura do relatório e 30 minutos para classificar manualmente as operações. Com o sistema de IA proposto, prevê-se uma poupança de 30 minutos por relatório. A leitura do relatório torna-se irrelevante e exigirá apenas validar a classificação fornecida pelo algoritmo e rever possíveis falhas. Estima-se que este processo dure cerca de 15 minutos. Isto significa que, para cada cinco relatórios diários, o sistema consegue poupar cerca de 2h30 de trabalho.

É de realçar que o ganho poderá aumentar com a evolução da aprendizagem contínua do modelo e, consoante a confiança e quantidade de dados que o modelo pode aprender, maior será a sua performance.

### 5.1. Limitações e propostas futuras

O presente estudo deparou-se com limitações, sendo que a maior foi o facto de não ser possível identificar, através do modelo, os problemas que podem ocorrer durante a perfuração. Esta limitação deveu-se ao facto de os dados serem confidenciais. Num futuro estudo, sugere-se identificar os problemas de perfuração e incluir junto destes resultados.

Com base nos dados do mesmo campo de exploração e, onde a geologia é semelhante, é possível, através do histórico, prever o tempo de duração de novos poços. Assim sendo, para trabalhos futuros, seria interessante apresentar um modelo de regressão capaz de prever o tempo de perfuração de um poço, com o objetivo de diminuir o tempo de sonda e otimizar a perfuração de novos poços.

Através do procedimento de identificação de problemas, será possível identificar o comportamento dos parâmetros para um caso real de *Pack-off*. O procedimento pode ser utilizado para evitar a ocorrência de problemas durante a perfuração do poço.

## Referências

- Agwu, O. E., Akpabio, J. U., Alabi, S. B., & Dosunmu, A. (2018). Artificial intelligence techniques and their applications in drilling fluid engineering: A review. *Journal of Petroleum Science and Engineering*, 167, 300–315. <https://doi.org/10.1016/j.petrol.2018.04.019>
- Ahmadi, M. A., Shadzadeh, S. R., Shah, K., & Bahadori, A. (2018). An accurate model to predict drilling fluid density at wellbore conditions. *Egyptian Journal of Petroleum*, 27(1), 1–10. <https://doi.org/10.1016/j.ejpe.2016.12.002>
- Barreto, J. M. (1997). Introdução às Redes Neurais Artificiais. *V Escola Regional de Informática*, (5), 47–71.
- Bello, O., Holzmann, J., Yaqoob, T., & Teodoriu, C. (2015). Application Of Artificial Intelligence Methods In Drilling System Design And Operations: A Review Of The State Of The Art. *Journal of Artificial Intelligence and Soft Computing Research*, 5(2), 121–139. <https://doi.org/10.1515/jaiscr-2015-0024>

- Bishop, C. M. (1996). *Neural Networks: A Pattern Recognition Perspective*. Aston University.
- Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., & Shearer, C. (2000). CRISP-DM 1.0 Step-by-step data mining guide. *SPSS Inc*, 9, 13.
- Elkhatatny, S. (2017). Real-Time Prediction of Rheological Parameters of KCl Water-Based Drilling Fluid Using Artificial Neural Networks. *Arabian Journal for Science and Engineering*, 42(4), 1655–1665. <https://doi.org/10.1007/s13369-016-2409-7>
- Goldstein, E. B., & Coco, G. (2014). A machine learning approach for the prediction of settling velocity. *Water Resources Research*, 50(4), 3595–3601. <https://doi.org/10.1002/2013WR015116>
- Han, J., Kamber, M., & Pei, J. (2011). The Morgan Kaufmann Series in Data Management Systems third edition. *Data Mining Concepts and Techniques*, 5(4), 83–124.
- Heriot, W. (2013). Drilling Engineering. In: *Material de apoio ao curso de MSc Petroleum Engineering*. Heriot Watt University.
- Hoffmann, J., Mao, Y., Wesley, A., & Taylor, A. (2018). Sequence Mining and Pattern Analysis in Drilling Reports with Deep Natural Language Processing. *SPE Annual Technical Conference and Exhibition*. Society of Petroleum Engineers. <https://doi.org/10.2118/191505-MS>
- Ian, W., Eibe, F., & Mark Hall, C. P. (2016). *Data mining: practical machine learning tools and techniques*. 4th ed.
- Iversen, F., Gressgård, L., Thorogood, J., Balov, M., & Hepsø, V. (2013). Drilling Automation: Potential for Human Error. *SPE Drilling & Completion*, 28(01), 45–59. <https://doi.org/10.2118/151474-PA>
- Jeirani, Z., & Mohebbi, A. (2006). Artificial Neural Networks Approach for Estimating Filtration Properties of Drilling Fluids. *Journal of the Japan Petroleum Institute*, 49(2), 65–70. <https://doi.org/10.1627/jpi.49.65>
- Jiang, Q., Tang, C., Chen, C., Wang, X., & Huang, Q. (2019). Stock Price Forecast Based on LSTM Neural Network. *Proceedings of the Twelfth International Conference on Management Science and Engineering Management*. *ICMSEM 2018*, 393–408. [https://doi.org/10.1007/978-3-319-93351-1\\_32](https://doi.org/10.1007/978-3-319-93351-1_32)
- Kamari, A., Gharagheizi, F., Shokrollahi, A., Arabloo, M., & Mohammadi, A. H. (2017). Estimating the drilling fluid density in the mud technology: Application in high temperature and high pressure petroleum wells. In *Heavy Oil: Characteristics, Production and Emerging Technologies* (pp. 285–295). Nova Science Publishers, Inc.
- Khan, S. A. R., & Yu, Z. (2016). Logistics forecasting method based on a hybrid quantum particle swarm optimization and RBF neural network model. *RISTI - Revista Iberica de Sistemas e Tecnologias de Informacao*, (18B), 317–236.

- Kohn, T., & Manaris, B. (2020). Tell Me What's Wrong: A Python IDE with Error Messages. *Proceedings of the 51st ACM Technical Symposium on Computer Science Education*, 1054–1060. ACM. <https://doi.org/10.1145/3328778.3366920>
- Landis, J. R., & Koch, G. G. (1977). The Measurement of Observer Agreement for Categorical Data. *Biometrics*, 33(1), 159. <https://doi.org/10.2307/2529310>
- Laureano, R. M., Caetano, N., & Cortez, P. (2014). Predicting inpatient length of stay in a Portuguese hospital using the CRISP-DM methodology. *RISTI - Revista Iberica de Sistemas e Tecnologias de Informacao*, (13), 83–99.
- Moazzeni, A., Nabaei, M., & Jegarluei, S. G. (2012). Decision Making for Reduction of Nonproductive Time through an Integrated Lost Circulation Prediction. *Petroleum Science and Technology*, 30(20), 2097–2107. <https://doi.org/10.1080/10916466.2010.495961>
- Moro, S., Cortez, P., & Rita, P. (2018). A divide-and-conquer strategy using feature relevance and expert knowledge for enhancing a data mining approach to bank telemarketing. *Expert Systems*, 35(3), e12253. <https://doi.org/10.1111/exsy.12253>
- Moro, S., Laureano, R., & Cortez, P. (2011). *Using data mining for bank direct marketing: An application of the crisp-dm methodology*.
- Murillo, A., Neuman, J., & Samuel, R. (2009). Pipe Sticking Prediction and Avoidance Using Adaptive Fuzzy Logic Modeling. *SPE Production and Operations Symposium*. Society of Petroleum Engineers. <https://doi.org/10.2118/120128-MS>
- Osman, E. A., & Aggour, M. A. (2003). Determination of Drilling Mud Density Change with Pressure and Temperature Made Simple and Accurate by ANN. *Middle East Oil Show*. Society of Petroleum Engineers. <https://doi.org/10.2118/81422-MS>
- Ozbayoglu, E. M., & Ozbayoglu, M. A. (2009). Estimating Flow Patterns and Frictional Pressure Losses of Two-Phase Fluids in Horizontal Wellbores Using Artificial Neural Networks. *Petroleum Science and Technology*, 27(2), 135–149. <https://doi.org/10.1080/10916460701700203>
- Rable, B. (2017). *The Future is Here: 3 Ways AI Roots Itself in O&G in the Surge Magazine*.
- Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). “Why Should I Trust You?”: Explaining the Predictions of Any Classifier. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 1135–1144. ACM. <https://doi.org/10.1145/2939672.2939778>
- Rooki, R., Ardejani, F. D., Moradzadeh, A., Mirzaei, H., Kelessidis, V., Maglione, R., & Norouzi, M. (2012). Optimal determination of rheological parameters for herschel-bulkley drilling fluids using genetic algorithms (GAs). *Korea-Australia Rheology Journal*, 24(3), 163–170. <https://doi.org/10.1007/s13367-012-0020-3>
- Sanchez-Pi, N., Martí, L., & Garcia, A. C. B. (2014). Text Classification Techniques in Oil Industry Applications. In *International Joint Conference SOCO'13-CISIS'13-ICEUTE'13* (pp. 211–220). [https://doi.org/10.1007/978-3-319-01854-6\\_22](https://doi.org/10.1007/978-3-319-01854-6_22)



- Shahdi, A., & Arabloo, M. (2014). Application of SVM Algorithm for Frictional Pressure Loss Calculation of Three Phase Flow in Inclined Annuli. *Journal of Petroleum & Environmental Biotechnology*, 05(03). <https://doi.org/10.4172/2157-7463.1000179>
- Sidahmed, M., Coley, C. J., & Shirzadi, S. (2015). Augmenting Operations Monitoring by Mining Unstructured Drilling Reports. *SPE Digital Energy Conference and Exhibition*, 403–415. <https://doi.org/SPE-173429-MS>
- Silva, C., & Barreiros, J. (2006). Intelligent Analysis Program Applied to Production Logs in Oil and Gas Wells. *IEEE Latin America Transactions*, 4(5), 353–358. <https://doi.org/10.1109/TLA.2006.4472135>
- Silva, F., Martins, D., Doria Neto, A., Rodrigues, M., & Da Mata, W. (2016). Optimization of the Oil Production Fields Submitted the Water Injection, Using the Algorithm NSGA-II. *IEEE Latin America Transactions*, 14(9), 4166–4172. <https://doi.org/10.1109/TLA.2016.7785948>
- Siruvuri, C., Nagarakanti, S., & Samuel, R. (2006). Stuck Pipe Prediction and Avoidance: A Convolutional Neural Network Approach. *IADC/SPE Drilling Conference*. Society of Petroleum Engineers. <https://doi.org/10.2118/98378-MS>
- Smith, E. (2020). *Python, the Fundamentals*. [https://doi.org/10.1007/978-3-030-60808-8\\_3](https://doi.org/10.1007/978-3-030-60808-8_3)
- Thomas, J. E. (2001). *Fundamentos de engenharia de petróleo*. Interciência.
- Toreifi, H., Rostami, H., & Manshad, A. K. (2014). New method for prediction and solving the problem of drilling fluid loss using modular neural network and particle swarm optimization algorithm. *Journal of Petroleum Exploration and Production Technology*, 4(4), 371–379. <https://doi.org/10.1007/s13202-014-0102-5>
- Wang, G., Pu, X.-L., & Tao, H.-Z. (2012). A Support Vector Machine Approach for the Prediction of Drilling Fluid Density at High Temperature and High Pressure. *Petroleum Science and Technology*, 30(5), 435–442. <https://doi.org/10.1080/10916466.2011.578095>
- Zhu, D., Liu, G. X., & Zhang, Q. Z. (2013). Research of Prewarning Pipe-Sticking Based on Neural Network. *Applied Mechanics and Materials*, 325–326, 1734–1737. <https://doi.org/10.4028/www.scientific.net/AMM.325-326.1734>