

# Uma ferramenta para detectar anomalias de produção utilizando aprendizagem profunda e árvore de decisão

A framework to detect production anomalies using deep learning and decision tree

**Eduardo da Silva Paranhos Sobrinho** <sup>1</sup> | **Felipe Lima de Oliveira** <sup>2</sup> | **Jorel Lopes Rodrigues dos Anjos** <sup>3</sup> | **Clemente Gonçalves** <sup>4</sup> | **Marcus Vinicius Duarte Ferreira** <sup>5</sup> | **Lucas Gouveia Omena Lopes** <sup>6</sup> | **William Wagner Matos Lira** <sup>7</sup> | **João Paulo Nogueira de Araújo** <sup>8</sup> | **Thiago Barbosa da Silva** <sup>9</sup> | **Lucas Pereira de Gouveia** <sup>10</sup>.

1. UNIVERSIDADE FEDERAL DE ALAGOAS, CTEC, . MACEIÓ - AL - BRASIL, eduardo.paranhos@lccv.ufal.br 2. CENPES/PDDP/PCP, PDDP/PCP, . RIO DE JANEIRO - RJ - BRASIL, flimao@petrobras.com.br 3. POCOS/CTPS/EO, CTPS/EO, . RIO DE JANEIRO - RJ - BRASIL, jorel.anjos@petrobras.com.br 4. CENPES/PDDP/PCP, PDDP/PCP, . RIO DE JANEIRO - RJ - BRASIL, clemente@petrobras.com.br 5. POCOS/CTPS/QC, CTPS/QC, . RIO DE JANEIRO - RJ - BRASIL, mvdferrera@petrobras.com.br 6. UNIVERSIDADE FEDERAL DE ALAGOAS, CTEC, . MACEIÓ - AL - BRASIL, lucasomena@lccv.ufal.br 7. UNIVERSIDADE FEDERAL DE ALAGOAS, CTEC, . MACEIÓ - AL - BRASIL, william@lccv.ufal.br 8. UNIVERSIDADE FEDERAL DE ALAGOAS, CTEC, . MACEIÓ - AL - BRASIL, joaopna@lccv.ufal.br 9. UNIVERSIDADE FEDERAL DE ALAGOAS, CTEC, . MACEIÓ - AL - BRASIL, thiago.barbosa@lccv.ufal.br 10. UNIVERSIDADE FEDERAL DE ALAGOAS, CTEC, . MACEIÓ - AL - BRASIL, lucasgouveia@lccv.ufal.br

## Resumo

Este trabalho apresenta um sistema para o monitoramento em tempo real de comportamentos anômalos de poços de petróleo utilizando dados de produção de pressão e temperatura. A detecção de eventos anômalos e inesperados é um campo de interesse na indústria petrolífera para melhorar a segurança das operações. A aplicação utiliza uma combinação de aprendizagem profunda e abordagens analíticas baseadas em regras para realizar a detecção de dados anômalos de sensores. Estudos iniciais são realizados para determinar os comportamentos e correlações a partir de dados de pressão e temperatura para as combinações mais comuns de estados de válvulas. A resposta esperada do sistema é aplicada ao treinamento de redes neurais dedicadas à detecção de anomalias, e dados de diferentes bacias petrolíferas são usados para treinar e validar a ferramenta proposta. Uma análise analítica baseada nas posições das válvulas e sensores também é usada para indicar possíveis cenários que possam causar as anomalias detectadas. A ferramenta obtida como resultado deste trabalho conduz a um sistema inovador de monitoramento em tempo real. Portanto, ele fornece um sistema para detecção de anomalias de resposta imediata, identificando ou antecipando eventos de falha, melhorando assim a segurança operacional e reduzindo custos.

**Palavras-chave:** Digital-Twin. Monitoramento de poços de petróleo. Detecção de anomalias. Aprendizagem Profunda

## Abstract

This work presents a framework for real-time monitoring of oil well anomalous behavior using pressure and temperature production data. The detection of abnormal and unexpected events is a field of interest in the oil industry to improve the safety of operations. The application uses a combination of deep learning and rule based analytic approaches to perform the detection of anomalous sensor data. Initial studies are conducted to determine the behaviors and correlations from pressure and temperature data for the most common combinations of valve states. The expected response of the system is applied to the training of neural networks devoted to anomaly detection, and data from different oil basins are used to train and validate the proposed tool. An analytical analysis based on the valve and sensor positions is also used to indicate possible scenarios that might cause the detected anomalies. The framework obtained as a result of this work lead to an innovative real time monitoring system. Therefore, it provides a detection tool for anomalies in a timely response, identifying or anticipating failure events, so improving operational safety and reducing costs.

**Keywords:** Digital-Twin. Oil well monitoring. Anomaly detection. Deep learning

**Received:** March 15, 2020 | **Accepted:** Jun 06, 2020 | **Available online:** Dec. 01, 2020.

**Article n°:** 437

**Cite as:** Rio Oil & Gas Expo and Conference, Rio de Janeiro, RJ, Brazil, 2020 (20)

**DOI:** <https://doi.org/10.48072/2525-7579.rog.2020.437>

## 1. Introdução

A indústria de óleo e gás figura entre as mais importantes do mundo, visto que seus derivados estão distribuídos em diversos setores da economia. A complexidade envolvida no processo de extração de óleo e gás, desde a fase inicial de exploração até a final com a distribuição, exige bastante cuidado na execução dos processos realizados ao longo de toda a cadeia produtiva do petróleo. Essa característica leva a indústria a buscar cada vez mais eficiência aliada a segurança.

Durante a sua fase de produção, um poço requer atenção constante para garantir que o fluido oriundo do reservatório chegue a superfície com eficiência e segurança. Uma vez que o poço está sujeito a diferentes condições de temperatura e pressão durante toda a sua vida útil, diversos problemas podem ocorrer, tais como a falha de equipamentos, operações mal executadas ou até mesmo situações não previstas no projeto.

Dada a alta complexidade da construção de um poço e a sua posterior manutenção, durante a etapa de produção, são instalados equipamentos de monitoramento para acompanhar o seu comportamento. Para esse acompanhamento, é comum o uso de sensores de pressão e temperatura instalados na região dos canhoneios e na árvore de natal. A partir dos dados transmitidos é possível verificar se o poço se encontra dentro da sua normalidade.

O monitoramento através de sensores é uma prática comum na avaliação de valores de cargas, vibração, deformações, efeitos dinâmicos ou outros dados de interesse a fim de detectar e reparar possíveis patologias que possam surgir nas estruturas. A previsão e detecção de eventos indesejados ou inesperados é um campo de interesse da indústria do petróleo para melhorar a segurança das operações e ajudar na tomada de decisões.

Vargas et al. (2019) fala em seu trabalho que a detecção e classificação de eventos indesejáveis raros são tarefas relevantes em diversas atividades realizadas e/ou monitoradas por seres humanos. A tarefa de responder a eventos inesperados em um processo envolve a detecção em tempo hábil de comportamentos anômalos, diagnosticando suas causas principais e, em seguida, tomando as decisões e ações de controle adequadas para trazer a operação de volta a um estado normal, seguro e operacional (VENKATASU-BRAMANIAN, 2003).

Dentre as principais ferramentas utilizadas para a realização de monitoramentos autônomos, pode-se citar árvores de decisão baseadas em regras, metodologias de aprendizagem de máquina e profunda. Uma importante vertente de metodologias que aprendizagem profunda é a interpretação dos dados coletados por sensores, identificando possíveis anomalias na operação. Como exemplo, os trabalhos Marchi et al. (2015) e Malhotra et al. (2015) apresentam metodologias que utilizam da aprendizagem profunda para detecção de comportamentos anormais em séries de dados temporais. Nesse contexto, o presente trabalho apresenta uma aplicação que utiliza a combinação de aprendizagem profunda e abordagens analíticas baseadas em regras para realizar a detecção de dados anômalos de sensores.

## 2. Sistema de Detecção de Anomalias

---

Alguns dos sensores instalados em poços emitem sinais regularmente contendo informações de pressão, temperatura e vazão, esses sensores se encontram em pontos estratégicos para monitorar o comportamento de um poço durante a sua vida produtiva. Com um tratamento adequado desses dados seguido da aplicação de técnicas de identificação de padrões, torna-se possível avaliar se o poço opera dentro da normalidade.

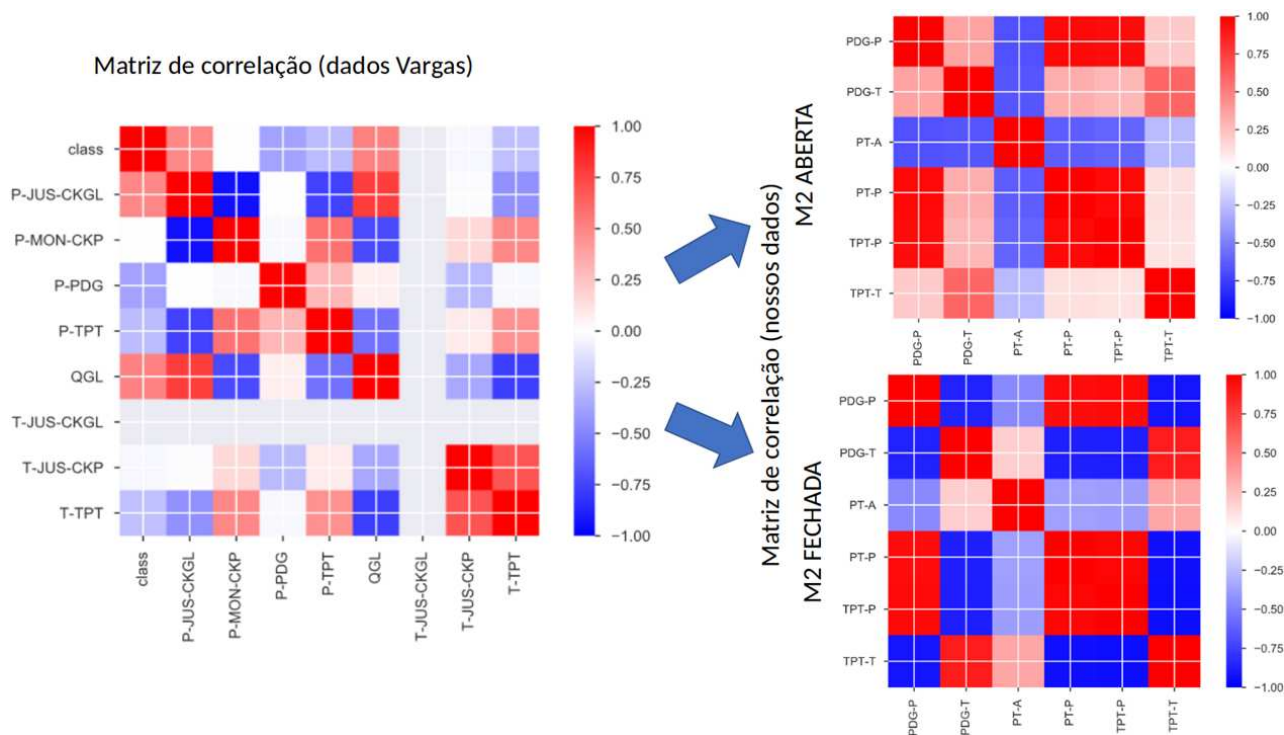
No sistema de detecção de anomalias existe um ponto de entrada dos dados oriundos dos sensores, a partir dos quais são realizadas verificações com a associação de uma árvore de decisão e uma rede de aprendizagem profunda. O sistema é capaz de aplicar essas metodologias a múltiplos poços simultaneamente, além disso, o mesmo está apto a mapear diversas configurações de válvulas e ajustar seus mecanismos de busca de anomalias para atender a que está vigente.

As anomalias detectadas pelo sistema podem ser classificadas em mapeadas e não mapeadas. Para o primeiro caso, a ferramenta conta com um portfólio de anomalias que contém o atributos de cenários indesejados em termos de pressão, temperatura e vazão. Quanto ao segundo grupo de anomalias, trata-se de casos que ainda não foram implementados no sistema, mas que configuram um cenário que não corresponde à normalidade do poço em questão, mas que pode ser detectado devido aos padrões de comportamento que sistema atribui ao poço.

## **2.1. Análise dos dados dos sensores de pressão e temperatura**

Uma das etapas do tratamento de dados consiste no agrupamento e sincronização dos dados dos sensores de pressão e temperatura fornecidos pela indústria. Os dados de cada sensor encontram-se armazenados em arquivos, contendo informações captadas em diferentes instantes de tempo. Os dados foram sincronizados e agrupados baseando-se nas combinações dos estados de válvula em produção. Uma primeira análise levou em conta as duas principais combinações de válvula de poços em produção da indústria, que diferem entre si pelo estado aberto ou fechado da válvula Master-2 (M2). Geram-se gráficos e estudos de correlação utilizados para o desenvolvimento de metodologias analíticas capazes de identificar comportamentos que fujam da normalidade. Fig. 1 mostra observações da correlação ente de valores de pressão e temperatura (PT, TPT e PDG) e vazão flow tanto para os dados da literatura disponibilizados por Vargas et al. (2019), quanto os dados disponibilizados pela indústria para a válvula M2 aberta e fechada.

Figura 1: Correlação entre os dados dos sensores.



Fonte: Autor (2020)

A cor vermelha indica correlação positiva (variáveis diretamente relacionadas), e a cor azul correlação negativa (variáveis inversamente proporcionais). Esse estudo indica que diferentes configurações de estado de válvula devem ser averiguadas e levadas em consideração durante o desenvolvimento das metodologias para detecção de anomalias. Trata-se, portanto, de um estudo qualitativo que indica a importância de avaliar metodologias de detecção que levam em consideração estados de válvula diferentes, pois eles influenciam na correlação das variáveis.

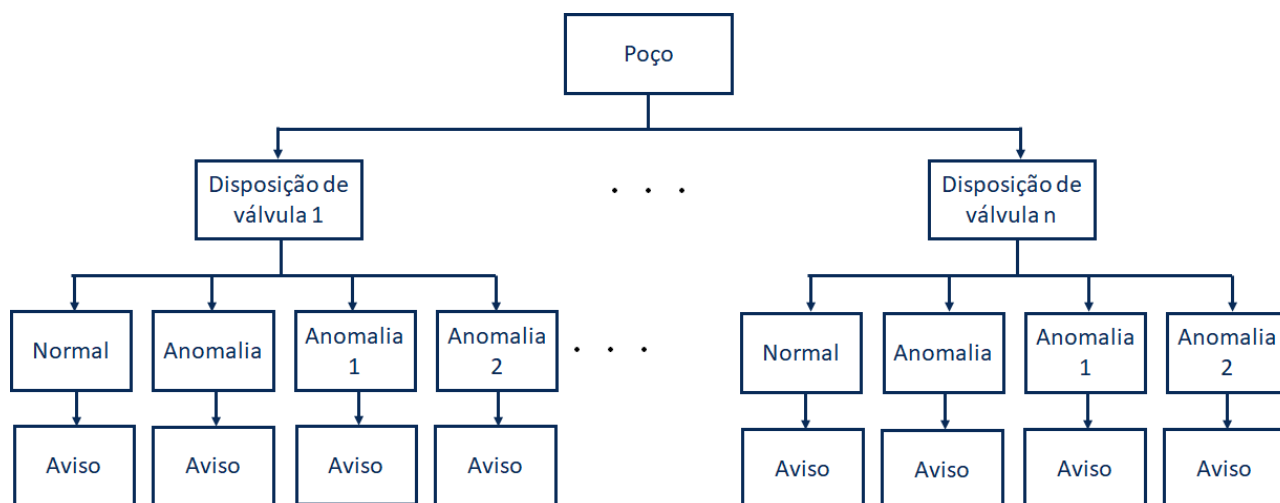
#### Detecção de anomalias a partir da árvore de decisão

A metodologia implantada que define a atuação da árvore de decisão consiste na estruturação de um conjunto de regras com base em camadas, as quais determinam o tipo de informação que é verificado. A medida que as decisões avançam por meio das camadas, os dados tornam-se cada vez mais específicos, dessa forma a árvore de decisão consegue gerenciar as informações de modo a extrair o maior proveito das mesmas para indicar a ocorrência de anomalias.

A primeira camada da árvore de decisão é referente ao poço que está sendo monitorado, dada a capacidade da ferramenta de monitorar múltiplos poços simultaneamente. Na camadas seguinte, é

feita a verificação de qual o estado do conjunto de válvulas atuante no referido poço, seguida da análise dos dados de pressão, temperatura e vazão para identificar qual o comportamento do poço, seja ele normal, anômalo não mapeado ou anômalo mapeado, em seguida é emitido um aviso da condição vigente no poço. A Figura X exibe a estrutura da árvore de decisão.

Figura 2: Estrutura da árvore de decisão analítica



Fonte: Autor (2020)

Para o desenvolvimento da metodologia, em um primeiro momento, analisa-se os dados referentes a sensores de pressão e temperatura (PDG, TPT e PT) em conjunto com a combinação dos estados de válvula (DHSV, M1, M2, W1, W2, XO, PXO). O histórico de dados é recolhido de poços e plataformas, em diferentes bacias petrolíferas. Estudos iniciais são realizados para determinar os comportamentos e correlações a partir dos dados de pressão e temperatura para as combinações mais comuns de estados de válvulas. É desenvolvida então uma abordagem analítica baseada em regras para detecção de variações anormais dos dados, servindo de base para indicar a natureza da anomalia com base em um portfólio conhecido.

A resposta desse sistema analítico é aplicada ao treinamento de redes neurais também dedicadas à detecção de anomalias. Os modelos treinados recebem dados em tempo real e procuram monitorar o comportamento das válvulas usando condições de carga que possam indicar eventos anormais durante a produção de petróleo com base no esquema de válvulas vigente. Um conjunto de dados presentes na literatura são usados para treinar e validar a metodologia proposta.

Com a validação da aplicação conjunta das metodologias analítica e de aprendizagem de máquina, é então desenvolvida uma interface utilizando a plataforma Node-Red. Essa ferramenta é capaz de receber dados de diferentes poços e fazer o monitoramento simultâneo utilizando as metodologias desenvolvidas.

## 2.2. Detecção de Anomalias a partir de Aprendizagem de Máquina

São desenvolvidas redes neurais profundas para detecção de anomalias, cujo treinamento é realizado utilizando séries recentes de dados de pressão e temperatura adquiridos por sensores. São exploradas diferentes arquiteturas de Redes Neurais artificiais (RNAs) (Haykin, 2001). Estudos preliminares de análise de arquiteturas mostram que modelos baseados em Suport Vector Machines (SVMs), Long Short-Term Memory (LSTM) e Redes Generativas Adversariais (GAN) podem reconhecer os eventos indesejáveis nas válvulas com base nas respostas de curto e longo prazo utilizadas para o treinamento. As redes propostas, com treinamento contínuo e robusto, podem ser usadas para detecção de falhas em tempo real, aliadas aos processos analíticos que podem indicar suas possíveis causas para os operadores.

A detecção de novidade da série temporal, ou detecção de anomalias, refere-se à identificação automática de eventos novos ou anormais incorporados em pontos normais da série temporal. Embora seja um tema desafiador na mineração de dados, tem vindo a adquirir uma atenção crescente devido ao seu enorme potencial para aplicações imediatas. A utilização de SVMs de uma classe configura-se como um método consagrado na literatura, como exemplo dos trabalhos Vargas et al (2019), Ma e Perkins (2003) e Taylor e Slicar (2015).

Com o intuito de criar redes neurais recorrentes (RNNs) mais robustas que tentam evitar a ocorrência de vanishing gradients, foram criadas as redes recorrentes Long Short-Term Memory (LSTM). As redes LSTM tem uma formulação mais sofisticada que possibilita configurações para "lembrar e esquecer" de maneira seletiva. Dessa forma, pode-se determinar uma quantidade de dados anteriores a serem considerados para previsão do próximo dado. Utilizando apenas dados de sistemas em funcionamento normal, os trabalhos Jan et al. (2019) , March et al. (2015) e Melhotra et al. (2015) apresentam metodologias para detecção de comportamentos anormais em séries de dados temporais.

As redes Generative Adversarial Networks (GAN) utilizam dados de instâncias normais para treinar duas redes neurais interligadas. Essas redes são conhecidas como "Gerador" e "Discriminador". Enquanto o gerador é responsável por tentar reproduzir cenários de funcionamento normal, o discriminador detecta anomalias nos dados de entrada de instâncias anormais, e as duas redes são treinadas em conjunto. Os trabalhos recentes (Zenati et al, 2018, Li et al, 2018, Sabokrou et al., 2018, Mattia et al., 2019, Schlegl et al, 2019) mostram que a utilização do discriminador das GANs na detecção de comportamento anômalo de sistemas trazem bons resultados.

A partir de testes iniciais, constata-se que as três metodologias se sobressaem por meio de determinadas características. Metodologias que utilizam SVMs dispõem de configurações pré-estabelecidas e consagradas que facilitam sua implementação. Porém, devido a sua simplicidade em relação aos outros dois métodos, apresenta um percentual de acerto aquém do esperado. A detecção de estados anômalo alcançou uma acurácia de aproximadamente 70%, condizente com os resultados apresentados por Vargas et al.(2019). Além disso, o ajuste no grau de tolerância é vinculado à uma nova configuração dos parâmetros da SVM, o que torna sua utilização pouco versátil.

As redes baseadas em GAN são as mais sofisticadas dentre as arquiteturas citadas. Devido ao seu grau de robustez, possui configuração mais complexa e exige uma considerável massa de dados. A preparação dos dados para o treinamento das GANs e a separação entre gerador e discriminador

promovem uma rede ao mesmo tempo extremamente robusta (capaz de reproduzir sinais complexos de maneira precisa) e não prática.

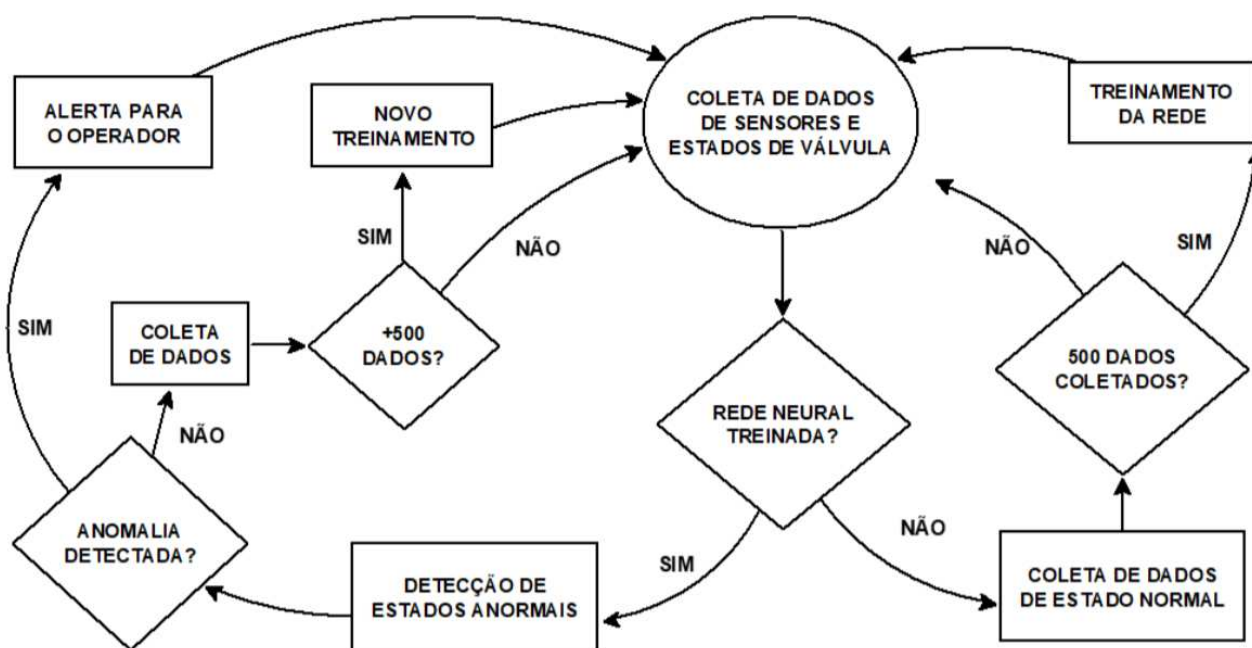
As redes LSTM por sua vez são simples o suficiente para serem automaticamente configuradas para cada poço e estado de válvula, versáteis o suficiente para ajuste de sua configuração e tolerância na detecção de anomalias após o seu treinamento, e precisas o suficiente para detectarem de maneira acurada trechos de dados anômalos. Seus resultados nos testes iniciais mostraram um percentual de acerto entre 90 a 98%, validando o método nos casos em estudo utilizando tanto dados da literatura quanto fornecidos pela indústria.

### 3. Resultados e Análises

A proposta para a rota de detecção de anomalias utilizando Inteligência Artificial consiste em uma metodologia capaz de aprender de maneira automática o que se enquadra como **estado normal**, e indicar ao operador do sistema quando os dados recebidos representam um **estado anômalo**. Dessa forma, para cada poço, a rede deve coletar dados e aprender sozinha sem interferência do operador.

Nesse sentido, dois módulos são desenvolvidos: módulo de Aprendizado e módulo de Detecção. O módulo de aprendizado armazena dados temporais sequenciais catalogados como **estado normal** pela formulação analítica de árvore de decisões. A partir da aquisição de aproximadamente 1000 dados normais (número de dados determinado através de testes acrescido de fator de segurança), o módulo entra na fase de treinamento e gera três saídas: 1) parâmetros de normalização dos dados; 2) pesos das redes neurais e 3) tolerância de anomalias calculada durante o treinamento. Quando essas três saídas são geradas e salvas, o módulo de detecção começa a operar, detectando comportamentos anormais em novos dados adquiridos. Fig. 3 ilustra o fluxograma de processos para a detecção de anomalias utilizando a rede LSTM desenvolvida.

Figura 3: Metodologia geral de detecção de anomalias utilizando IA

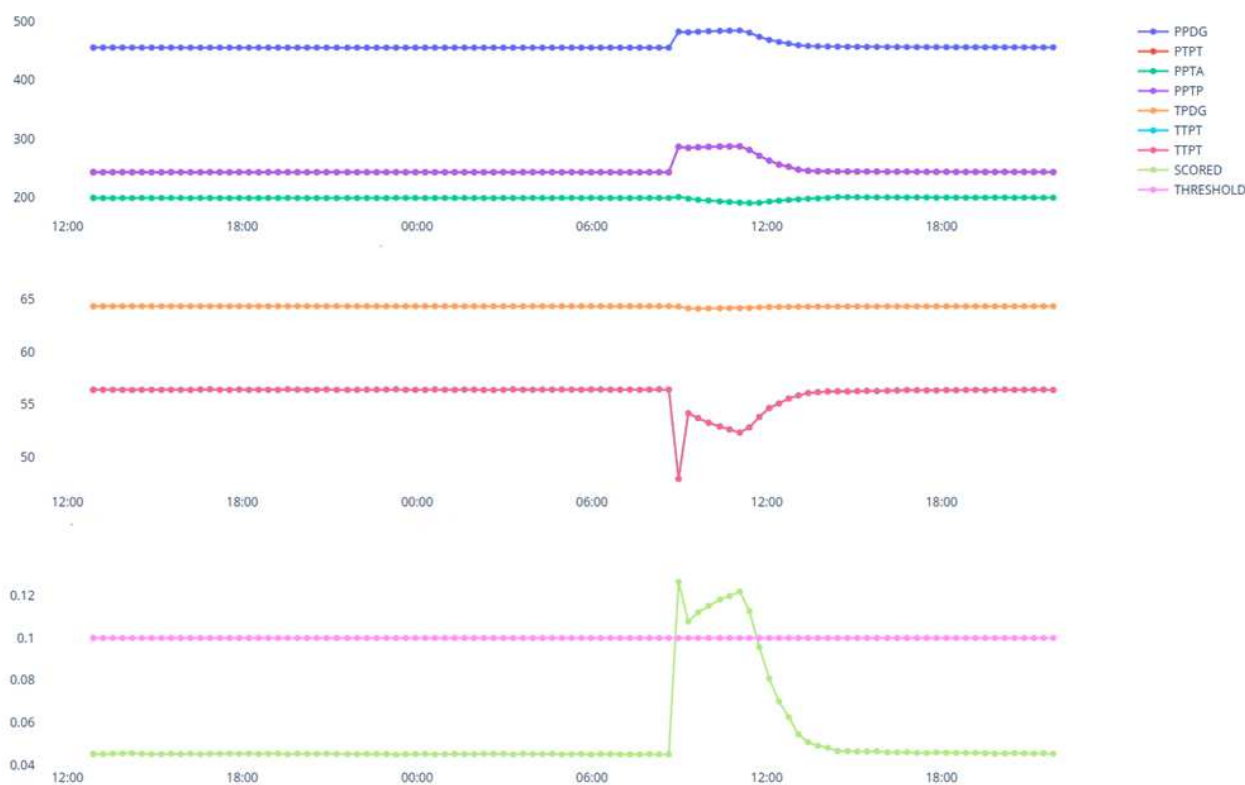


Fonte: Autor (2020)

Fig. 4 mostra um **estado anômalo**. Nessas figuras, observamos os valores de pressão nos sensores localizados em PDG, PT e TPT no gráfico superior, os valores de temperatura do PDG e TPT no gráfico central, e o valor de score de anomalia calculado pela rede neural no gráfico inferior, juntamente com uma linha de tolerância. Caso o score de anomalia ultrapasse a tolerância, a rede acusa a ocorrência para o operador.

Figura 4: Visualização de estado anormal durante a produção

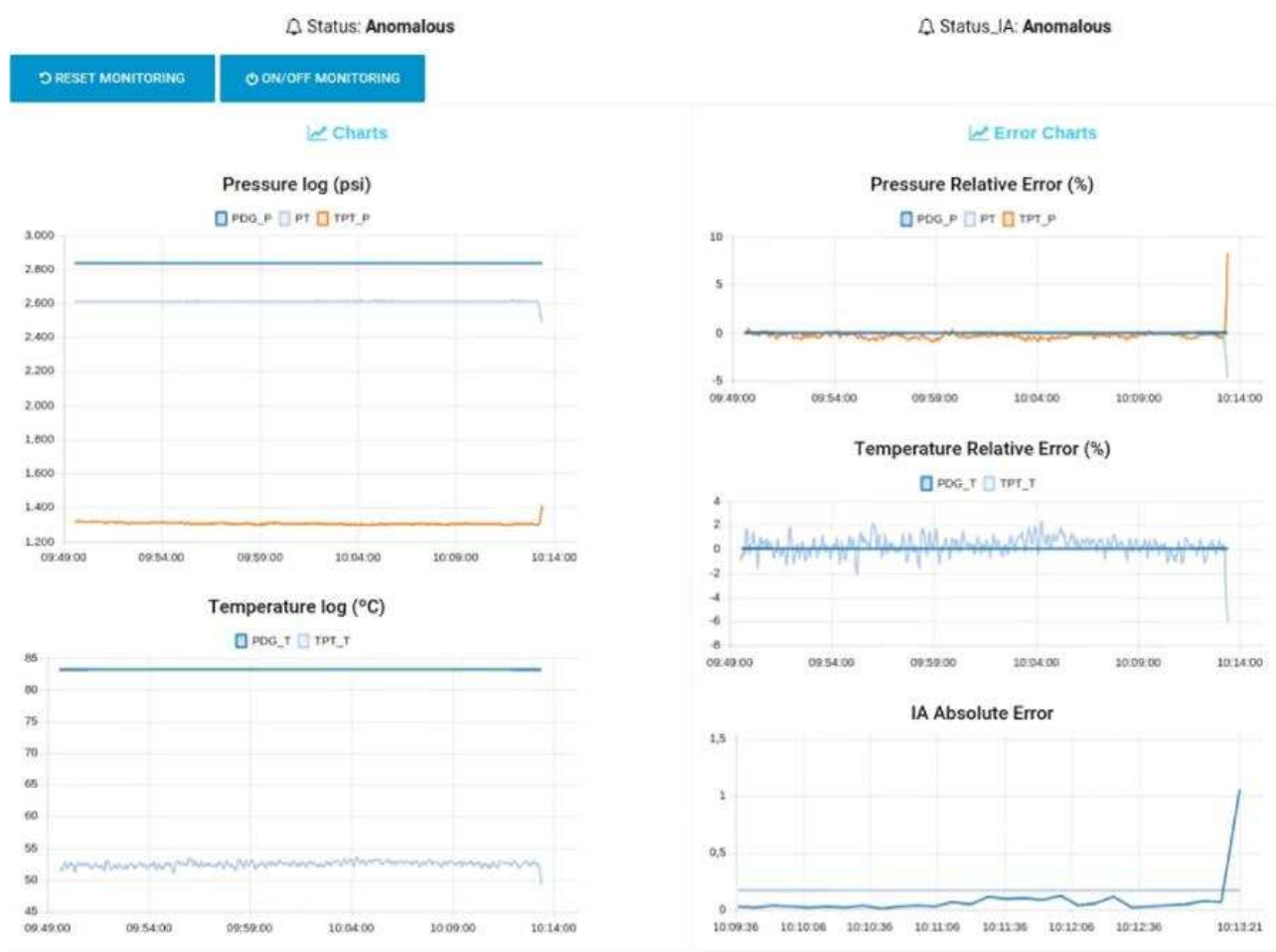




Fonte: Autor (2020)

Três parâmetros são obtidos após do treinamento da rede: 1) parâmetros de normalização dos dados; 2) pesos de redes neural e 3) tolerância de anomalias calculada durante o treinamento. Os dois primeiros são parâmetros de configuração pertinentes à transformação das variáveis de entrada e coeficientes da rede neural. A tolerância de anomalias é calculada com base do score de erro Médio Absoluto (MAE). A tolerância se baseia em amostras que tiverem um erro superior a média da distribuição de erros somado a três desvios padrões. Esse valor de tolerância engloba 99,73% dos valores aferidos para uma distribuição normal. Um dos casos de validação da ferramenta, assim como a Fig. 4, é apresentado em Fig. 5, onde é ilustrado um caso de monitoramento de dados reais através da interface da ferramenta desenvolvida, detectando de forma acurada o comportamento anômalo constatado em um poço real tanto pela abordagem analítica quanto pela rede neural.

Figura 5: Estado anormal detectado por árvore de decisões e inteligência artificial



Fonte: Autor (2020)

#### 4. Considerações finais

A partir dos resultados apresentados pelo sistema do monitoramento em tempo real e detecção de anomalias, considera-se os métodos apresentados efetivos e validados. A ferramenta apresenta abas de configuração e visualização dos resultados, além de possibilitar a adição de novas anomalias ao portfólio catalogado para a abordagem analítica baseada em regras. A aquisição contínua de dados por meio de sensores resulta numa evolução constante da precisão das redes neurais desenvolvidas, garantindo assim uma manutenção autônoma ao longo da vida útil do poço. Como consequência, a ferramenta proporciona a detecção de anomalias em tempo hábil de resposta, garantindo tanto maior segurança operacional quanto a redução de custos relacionados a possíveis perdas

#### 5. Agradecimentos

Os autores gostariam de agradecer a PETROBRAS pelo financiamento e incentivos de projetos que levaram aos resultados descritos.



## Referências

Haykin, S. (2007). *Redes Neurais: Princípios e Prática*. Brazil: Bookman Editora. Retrieved from

[https://books.google.com.br/books/about/Redes\\_Neurais.html?id=bhMwDwAAQBAJ&redir\\_esc=y](https://books.google.com.br/books/about/Redes_Neurais.html?id=bhMwDwAAQBAJ&redir_esc=y)

Jan, V., & Hromada, M. (2019). Optimization of the Novelty Detection Model Based on LSTM Autoencoder for ICS Environment (Vol. 1, pp. 306–319). Presented at the Computational Methods in Systems and Software, Zlin, Czech Republic. Retrieved from [https://link.springer.com/chapter/10.1007/978-3-030-30329-7\\_28](https://link.springer.com/chapter/10.1007/978-3-030-30329-7_28)

Li, D., Chen, D, Goh, J, & Ng, S. (2018). *Anomaly detection with generative adversarial networks for multivariate time series* Arxiv: Arxiv. Retrieved from <https://arxiv.org/abs/1809.04758>

Ma, J., & Perkins, S. (2003). . Time-series novelty detection using one-class support vector machines. *2003*, 3(8), 1741 – 1745. <https://doi.org/10.1109/IJCNN.2003.1223670>

Malhotra, P., Vig, L., Shroff, G., & Agarwal, P. (2015). Long Short Term Memory Networks for Anomaly Detection in Time Series (Vol. 6, pp. 22–24). Presented at the European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning., Bruges (Belgium),. Retrieved from

[https://www.researchgate.net/publication/304782562\\_Long\\_Short\\_Term\\_Memory\\_Networks\\_for\\_Anomaly\\_Detection\\_in\\_Time\\_Series](https://www.researchgate.net/publication/304782562_Long_Short_Term_Memory_Networks_for_Anomaly_Detection_in_Time_Series)

Marchi, E., Vesperini, F., Eyben, F., Squartini, S., & Schuller, B. (2015). A novel approach for automatic acoustic novelty detection using a denoising auto-encoder with bidirectional lstm neural networks. In *International Conference on Acoustics, Speech and Signal Processing* (Vol. 1, p. 1). Austrália. Retrieved from <https://ieeexplore.ieee.org/document/7178320>

Mattia, F. D., Galeone, P., Simoni, M. D., & Ghelfi, E. (2019). *A survey on gans for anomaly detection*. Arxiv: Arxiv. Retrieved from <https://arxiv.org/abs/1906.11632>

Sabokrou, M., Khalooei, M., Fathy, M., & Adeli, E. (2018). *Adversarially learned one-class classifier for novelty detection* Arxiv: Arxiv. Retrieved from <https://arxiv.org/abs/1802.09088>

Schlegl, T., Seeböck, P., Waldstein, S. M., Langs, G., & Schmidt, U. (2019). f-AnoGAN: Fast unsupervised anomaly detection with generative adversarial networks. *2019*, 54(13), 30–44. <https://doi.org/10.1016/j.media.2019.01.010>

Shawe-Taylor, J., & Žlićar, B. (2015). .Novelty Detection with One-Class Support Vector Machines. *Advances in Statistical Models for Data Analysis*, 1(8), 231–257. [https://doi.org/10.1007/978-3-319-17377-1\\_24](https://doi.org/10.1007/978-3-319-17377-1_24)

Vargas, R. E. V. et al. (2019). A realistic and public dataset with rare undesirable real events in oil wells *2019*, 181(8), 106223. <https://doi.org/10.1016/j.petrol.2019.106223>

Venkatasubramanian, V. (2003). A review of process fault detection and diagnosis: Part I: Quantitative model-based methods. *2003*, 27(3), 293–311. [https://doi.org/10.1016/S0098-1354\(02\)00160-6](https://doi.org/10.1016/S0098-1354(02)00160-6)