



# APLICAÇÃO DE APRENDIZADO DE MÁQUINA PARA DETECÇÃO DE FALHAS EM SISTEMAS DE ELEVAÇÃO DE PETRÓLEO

**Eduardo N. de OLIVEIRA<sup>1</sup> ; Kaio H. F. ROCHA<sup>2</sup> ; Rejane B. SANTOS<sup>3</sup>**

## RESUMO

Este trabalho apresenta o desenvolvimento inicial de um modelo de rede neural para detecção de anomalias do tipo restrição no CKP em poços de petróleo, utilizando dados reais do *3Wdataset*. Foram aplicadas etapas de tratamento dos dados, seleção de variáveis por PCA e normalização. O modelo foi treinado com três variáveis principais e obteve acurácia superior a 99 %, com ótimo desempenho validado por matriz de confusão. O projeto encontra-se em andamento.

**Palavras-chave:** Rede Neural; CKP; Python.

## 1. INTRODUÇÃO

A produção de petróleo depende de complexos sistemas de elevação artificial que garantem o fluxo contínuo do reservatório. Esses sistemas são suscetíveis a anomalias operacionais – como incrustações de sólidos, corrosão de equipamentos ou obstruções em válvulas *choke* de produção (CKP) – que podem interromper abruptamente a produção. Estima-se que a paralisação de um poço devido a uma anomalia pode causar perdas da ordem de US\$ 1 milhão por dia na receita, por isso o monitoramento constante das variáveis de operação e a detecção precoce de desvios de comportamento são fundamentais (Andreolli, 2016). Dados de referência como o conjunto *3Wdataset* ilustram bem esse cenário: por exemplo, ele registra 221 ocorrências da classe “Restrição Rápida em CKP”, destacando a relevância dessa falha.

Diante desse contexto, o aprendizado de máquina (ML) desponta como uma solução promissora para a detecção de falhas em poços petrolíferos. Fernandes Júnior (2022) aponta que técnicas de ML permitem detectar e prevenir anomalias, minimizando as interrupções operacionais. Esses métodos conseguem processar grandes volumes de dados de sensores, aprendendo padrões normais de operação e assinalando automaticamente comportamentos anômalos sem intervenção humana. Assim, a detecção automática baseada em ML habilita ações de manutenção antes que as falhas se agravem, reduzindo custos operacionais.

Para este trabalho, emprega-se o *3Wdataset*, banco de dados público construído por Vargas *et*

<sup>1</sup>Bolsista PIBIC/CNPq, IFSULDEMINAS – Campus Pouso Alegre. E-mail: eduardo.nunes@alunos.ifsuldeminas.edu.br

<sup>2</sup>Voluntário PIBIC/CNPq, IFSULDEMINAS – Campus Pouso Alegre. E-mail: kaio.rocha@alunos.ifsuldeminas.edu.br

<sup>3</sup>Orientadora PIBIC/CNPq, IFSULDEMINAS – Campus Pouso Alegre. E-mail: rejane.santos@ifsuldeminas.edu.br

*al.*, (2019) com medições históricas de poços em operação normal e em diferentes estados de falha. Esse conjunto possui 1.984 instâncias multivariadas (597 normais e 1.387 anômalas) de três origens distintas – dados reais de planta, simulações computacionais e instâncias desenhadas por especialistas – cobrindo oito classes de anomalia. Em particular, a classe “Restrição Rápida em CKP” apresenta 221 ocorrências no *dataset*. Com esse material rotulado, o objetivo é desenvolver um modelo de rede neural que aprenda a distinguir o comportamento normal do poço e identifique precocemente a transição para um estado anômalo de restrição no *choke* de produção, contribuindo assim para a manutenção preditiva em sistemas de elevação de petróleo.

## 2. MATERIAL E MÉTODOS

Para a construção do classificador de falhas do tipo restrição rápida em *choke* de produção (CKP), utilizou-se o *3WDataset*, disponibilizado publicamente por Vargas *et al.*, (2019) como referência. Essa base de dados contém séries temporais multivariáveis extraídas de poços marítimos operados por elevação natural, com instâncias rotuladas em duas classes — normal (0) e anomalia de restrição em CKP (6) —, e originalmente composta por cinco variáveis de processo: pressão e temperatura no *Temperature and Pressure Transducer* (P-TPT, T-TPT), pressão montante e temperatura jusante ao *choke* de produção (P-MON-CKP, T-JUS-CKP) e pressão jusante ao *choke* de gás-líquido (P-JUS-CKGL). Todo o desenvolvimento foi implementado em ambiente *Python*, com as bibliotecas *pandas* e *NumPy* para manipulação de dados, *scikit-learn* para pré-processamento e PCA, *TensorFlow/Keras* para construção e treinamento da rede neural, e *Matplotlib* e *Seaborn* para visualização gráfica, executando-se no Google Colab.

O pré-processamento iniciou-se removendo-se as colunas cuja proporção de valores ausentes ou congelados excedia 90 % — especificamente P-PDG, T-JUS-CKGL e QGL — e, em seguida, eliminando totalmente as observações ainda contendo *NaN*. Para identificar variáveis “congeladas” (valores constantes por longos trechos), aplicou-se uma simples lógica de detecção de repetição de valor (mais de 300 amostras seguidas iguais), que permitiu descartar registros sem variabilidade relevante. Como as classes estavam desequilibradas, fez-se a fusão das classes 6.0 (estado estacionário de falha) e 106.0 (estado transiente de falha) em uma única classe positiva, seguida de *undersampling* aleatório na classe negativa (0.0) até igualar o seu tamanho ao do conjunto positivo, de modo a evitar viés de maioria durante o treinamento.

Para reduzir ruído e investigar a relevância de cada variável, aplicou-se Análise de Componentes Principais (PCA) sobre o conjunto limpo. Os cinco componentes explicaram 98 % da variância total e os *loadings* indicaram que P-TPT, T-TPT e P-MON-CKP eram as variáveis que mais contribuíam para a dispersão dos dados. Assim, manteve-se apenas esse trio de atributos na

modelagem. Antes de entrarem na rede neural, esses três sinais foram normalizados via *MinMaxScaler* do *scikit-learn*, mapeando seus valores para a faixa [0, 1].

Além da rede neural artificial, aplicou-se também o método de Máquina de Vetores de Suporte (SVM) como alternativa para classificação das anomalias. O objetivo foi comparar os resultados entre os dois métodos, avaliando a simplicidade de implementação, o tempo de treinamento e o desempenho final de classificação.

### 3. RESULTADOS E DISCUSSÃO

Durante os testes iniciais, foi implementado um modelo de rede neural com duas camadas ocultas, porém os resultados obtidos apresentaram baixa acurácia e desempenho insatisfatório na detecção de anomalias. Após essa avaliação, a arquitetura foi ajustada para um modelo sequencial de *perceptron* multicamada com três camadas ocultas, contendo respectivamente 64, 32 e 16 neurônios, todas com função de ativação ReLU. A camada de saída permaneceu com um neurônio e ativação sigmoide, adequada para classificação binária. O conjunto de dados foi dividido de forma estratificada em 70 % para treinamento e 30 % para validação, com *batch size* de 128 e 100 épocas de treinamento.

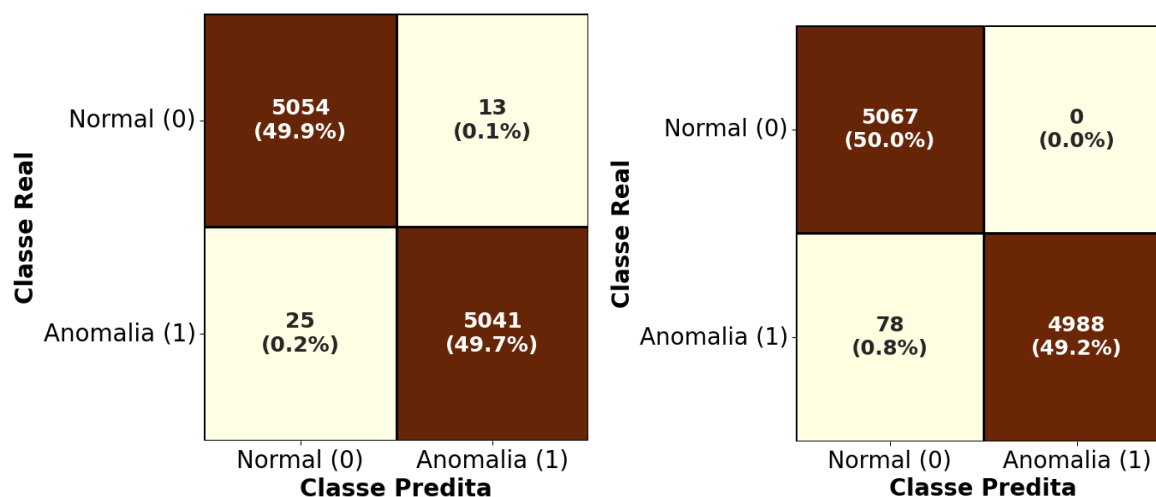
Com essa nova estrutura, o modelo apresentou desempenho significativamente superior, alcançando acurácia acima de 99 % e excelente separação entre classes, confirmada por meio da matriz de confusão. Esse ajuste mostrou-se essencial para que a rede aprendesse os padrões relacionados à anomalia de forma mais precisa.

A matriz de confusão gerada a partir do conjunto de teste mostrou que o modelo foi capaz de identificar corretamente a grande maioria dos casos, tanto normais quanto anômalos. Foram registrados apenas alguns poucos erros, o que é esperado em qualquer sistema de aprendizado de máquina, principalmente quando se trata de um fenômeno industrial complexo como o comportamento de um poço de petróleo. A utilização da matriz de confusão foi fundamental para visualizar esses acertos e erros de forma clara, ajudando a confirmar que o modelo não está simplesmente "chutando", mas sim fazendo previsões consistentes.

O modelo SVM apresentou desempenho equivalente ao da rede neural, com acurácia de aproximadamente 99 % e matriz de confusão bem equilibrada. Apesar de mais simples, o SVM mostrou-se eficiente para este problema, oferecendo tempo de processamento menor e implementação mais direta, conforme Figura 1.

Por fim, esses resultados sugerem que a abordagem utilizada, com base em redes neurais e dados reais do *3WDataset*, é eficaz para auxiliar na detecção precoce de falhas operacionais. Mesmo sendo uma primeira versão do modelo, os indicadores iniciais já apontam para um sistema promissor, com potencial de aplicação prática em monitoramento contínuo de poços.

Figura 1: Matriz confusão da rede neural vs matriz confusão do SVM



Fonte: Elaborado pelo autor.

#### 4. CONCLUSÃO

Neste trabalho, desenvolveu-se um modelo inicial de rede neural supervisionada com o objetivo de detectar anomalias do tipo restrição rápida no CKP, utilizando dados reais provenientes do *3WDataset*. Foram aplicadas etapas de pré-processamento, seleção de variáveis e balanceamento de classes, seguidas do treinamento de um classificador binário com bom desempenho. Os resultados obtidos, especialmente a matriz de confusão e a alta acurácia no conjunto de teste, indicam que a abordagem adotada é promissora para o diagnóstico automático de falhas em sistemas de elevação de petróleo.

Conclui-se que ambos os métodos, RNA e SVM, são capazes de identificar a anomalia de restrição no CKP com alta precisão. O SVM, por sua simplicidade e desempenho comparável, destaca-se como uma alternativa interessante para aplicações que exigem menor custo computacional.

#### REFERÊNCIAS

- ANDREOLLI, Ivanildo. Introdução à elevação e escoamento monofásico e multifásico de petróleo. Rio de Janeiro: Interciência, 2016.
- FERNANDES JÚNIOR, Wander. Comparação de classificadores para detecção de anomalias em poços produtores de petróleo. 2022. 64 f. Dissertação (Mestrado em Computação Aplicada) - Instituto Federal do Espírito Santo, Campus Serra, Serra, 2022.
- VARGAS, Ricardo Emanuel Vaz et al. A realistic and public dataset with rare undesirable real events in oil wells. *Journal of Petroleum Science and Engineering*, v. 181, p. 106223, 2019. ISSN 0920-4105.