

UNIVERSIDADE FEDERAL DE ALAGOAS
CENTRO DE TECNOLOGIA
GRADUAÇÃO EM ENGENHARIA DE PETRÓLEO

RICARDO NÍCOLAS SANTOS OLIVEIRA

APLICAÇÃO DE TÉCNICAS DE INTELIGÊNCIA ARTIFICIAL PARA
MODELAGEM DE FALHAS DE INTEGRIDADE NOS SISTEMAS DE
REVESTIMENTO DE POÇO.

MACEIÓ

2023

RICARDO NÍCOLAS SANTOS OLIVEIRA

**APLICAÇÃO DE TÉCNICAS DE INTELIGÊNCIA ARTIFICIAL
PARA MODELAGEM DE FALHAS DE INTEGRIDADE NOS
SISTEMAS DE REVESTIMENTO DE POÇO.**

Trabalho de Conclusão de Curso apresentado ao curso de Engenharia de Petróleo, da Universidade Federal de Alagoas, realizado sob orientação do Prof. Dr. João Paulo Lima Santos e coorientação do Prof. Dr. Eduardo Toledo de Lima Junior.

MACEIÓ

2023

Catlogação na fonte
Universidade Federal de Alagoas
Biblioteca Central
Divisão de Tratamento Técnico

Bibliotecária Responsável: Lívia Silva dos Santos - CRB 1670

O48a Oliveira, Ricardo Nícolas Santos.

Aplicação de técnicas de inteligência artificial para modelagem de falhas de integridade nos sistemas de revestimento de poço / Ricardo Nícolas Santos Oliveira. – 2023.
34f. : il.

Orientador: João Paulo Lima Santos.

Coorientador: Eduardo Toledo de Lima Júnior.

Monografia (Trabalho de Conclusão de Curso em Engenharia de Petróleo) – Universidade Federal de Alagoas. Centro de tecnologia. Maceió, 2023.

Bibliografia: f. 33-34

1. Hidrocarbonetos - Exploração. 2. Poços de petróleo - Petróleo. 3. Inteligência artificial - Petróleo. I. Título.

CDU: 665.6



ATA DE DEFESA DE TRABALHO DE CONCLUSÃO DE CURSO

Às 16:30 horas do dia 17 de outubro de 2023 na sala virtual do Centro de Tecnologia, reuniu-se a banca examinadora para avaliação e defesa do Trabalho de Conclusão de Curso (TCC) do(a) aluno(a) RICARDO NÍCOLAS SANTOS OLIVEIRA, sob o título “APLICAÇÃO DE TÉCNICAS DE INTELIGÊNCIA ARTIFICIAL PARA MODELAGEM DE FALHAS DE INTEGRIDADE NOS SISTEMAS DE REVESTIMENTO DE POÇO, composta pelos seguintes participantes:

Orientador: Prof(a) (A₁) João Paulo Lima Santos,

Co-orientador Prof(a) Eduardo Toledo de Lima Júnior,

Prof(a) Avaliador(a) (A₂) Lucas Pereira de Gouveia,

Prof(a) Avaliador(a) (A₃) Christiano Augusto Ferrario Várady Filho.

A apresentação oral do aluno foi realizada em 40 minutos, findos os quais foi iniciado o debate, perfazendo um tempo total de defesa de 2 hora (s) e 10 minutos.

Abaixo, assinatura do orientador, seguida dos membros da banca, de acordo com a ordem listada acima, todos seguidos da respectiva nota, a assinatura do aluno e a assinatura do Coordenador de TCC.

A nota final de cada avaliador, denominada NFA_n , será calculada da seguinte forma:

$$NFA_n = (0,6 \times \text{Monografia} + 0,2 \times \text{Apresentação oral do trabalho} + 0,2 \times \text{Arguição})$$

$$A_1: NFA_1 = (0,6 \times 9,8 + 0,2 \times 9,8 + 0,2 \times 9,5) = 9,80$$

$$A_2: NFA_2 = (0,6 \times 8,5 + 0,2 \times 8,5 + 0,2 \times 8,5) = 8,50$$

$$A_3: NFA_3 = (0,6 \times 8,00 + 0,2 \times 8,00 + 0,2 \times 7,00) = 8,00$$

Documento assinado digitalmente



JOAO PAULO LIMA SANTOS

Data: 18/10/2023 08:23:53-0300

Verifique em <https://validar.iti.gov.br>

Prof(a) (A₁) _____, nota final: $9,80 \times (0,9) + 8,55 \times (0,1) = 9,67$

Prof(a) (A₂) _____, nota final: $8,50 \times (0,9) + 8,20 \times (0,1) = 8,47$

Prof(a) (A₃) _____, nota final : $8,00 \times (0,9) + 8,90 \times (0,1) = 8,09$

*Para a nota final, deve-se ponderar a média desta etapa com 10% da nota que o aluno obteve no Plano de TCC.

O aluno obteve média final de Trabalho de Conclusão de Curso de: 8,74 (Oito e setenta e quatro décimos)

Documento assinado digitalmente



RICARDO NICOLAS SANTOS OLIVEIRA

Data: 24/10/2023 20:42:21-0300

Verifique em <https://validar.iti.gov.br>

Aluno: _____

Coordenador de TCC: _____

Documento assinado digitalmente



CHRISTIANO AUGUSTO FERRARIO VARADY FILH

Data: 18/10/2023 09:11:49-0300

Verifique em <https://validar.iti.gov.br>

Documento assinado digitalmente



LUCAS PEREIRA DE GOUVEIA

Data: 24/10/2023 17:36:02-0300

Verifique em <https://validar.iti.gov.br>

AGRADECIMENTO

À minha mãe Roseli Oliveira, que nessa jornada foi pai e mãe na maioria do tempo, sempre se sacrificou pelos filhos, sendo o alicerce da minha família, sem ela nada disso seria possível.

À minha namorada Tatiana Costa, que prestou todo apoio e empatia nesses anos que minha acompanha.

A meu pai Ricardo Oliveira, que o pouco tempo que convivemos formou parte do meu caráter, que esteja em um lugar melhor que nós.

A meu irmão Rafael Santos, que sempre me incentivou a sonhar e lutar pelo melhor.

A meus amigos de turma, sem eles a jornada seria mais árdua, e com certeza tem uma influência positiva na trilha da minha graduação.

Aos orientadores, à banca examinadora e aos professores do curso de Engenharia de Petróleo da UFAL que dedicaram tempo e compartilharam o conhecimento técnico necessário para que este trabalho e a graduação fossem possíveis.

RESUMO

Atualmente, o petróleo é uma das fontes de energia mais consumidas no mundo, e essa demanda está em constante crescimento. No entanto, a produção de hidrocarbonetos é repleta de desafios. Pode-se citar a natureza complexa de suas operações, bem como o ambiente severo ao qual suas instalações estão expostas. É o caso, por exemplo, de poços offshore sujeitos a condições extremas de carregamentos ambientais, sob elevados níveis de pressão e temperatura. Este trabalho aborda anomalias ocorridas ao longo da vida em produções dos poços, que podem levar a situações de perda de integridade e, conseqüentemente, a danos ambientais, riscos a vidas humanas e prejuízo financeiro. Dessa forma, a classificação de eventos indesejados durante a produção de hidrocarbonetos beneficia a indústria como um todo, corroborando com técnicas capazes de diminuir custos associados à intervenção em poços, reduzindo o erro humano, o retrabalho e garantindo a segurança operacional. A literatura e a prática mostram que os métodos de inteligência artificial (IA) são capazes de contribuir de diversas formas nesse âmbito. Nesse sentido, o objetivo do presente trabalho é desenvolver um modelo automatizado com técnicas de IA para classificar anomalias com base em séries de dados públicos disponibilizados por Vargas et al. (2019). Este trabalho utiliza as técnicas de *K-nearest neighbors* (KNN), *Support Vector Machine* (SVM) e *Non-Myopic Early Classifier*, e por meio da biblioteca de aprendizado de máquina *sklearn* como ferramenta de análise de dados e *Python* para o desenvolvimento dos algoritmos. Nesse contexto, com a aplicação das técnicas, alcançou-se um desempenho notável. O método KNN, em particular, obteve uma acurácia de mais de 90%. Além disso, a análise inclui métricas essenciais como acurácia balanceada, precisão, revocação e *F1 score*.

Palavras-chave: Inteligência Artificial, Classificação de Anomalias, Monitoramento de Poços de Petróleo, Classificação de Séries Temporais.

ABSTRACT

Currently, oil is one of the most consumed sources of energy in the world, and this demand is constantly growing. However, hydrocarbon production is filled with challenges. These challenges include the complex nature of its operations, as well as the harsh environment to which its facilities are exposed. This is the case, for example, with offshore wells subject to extreme environmental conditions, under high levels of pressure and temperature. This work addresses anomalies that occur throughout the lifespan of well productions, which can lead to situations involving loss of integrity, resulting in environmental damage, risks to human lives, and financial loss. Thus, the classification of unwanted events during hydrocarbon production benefits the industry as a whole, supporting techniques capable of reducing costs associated with well interventions, minimizing human error, avoiding rework, and ensuring operational safety. Literature and practice demonstrate that artificial intelligence (AI) methods can contribute in various ways in this context. In this sense, the objective of this work is to develop an automated model using AI techniques to classify anomalies based on publicly available data series provided by Vargas et al. (2019). This work utilizes the K-nearest neighbors (KNN), Support Vector Machine (SVM), and Non-Myopic Early Classifier techniques, along with the sklearn machine learning library as a data analysis tool and Python for algorithm development. In this context, the techniques applied achieved the best performance in terms of accuracy with KNN, with a value of 97.56%. Additionally, the analysis is conducted using metrics such as balanced accuracy, precision, recall, and F1 score. The results obtained demonstrated a significant improvement compared to the public data.

Keywords: Artificial Intelligence, Anomaly Classification, Oil Well Monitoring, Time Series Classification.

LISTA DE ILUSTRAÇÕES

Figura 1 - Esquemático de poço.	13
Figura 2 – Esquema simplificado de um sistema de produção de petróleo. ...	15
Figura 3 - Ilustração do método KNN.	17
Figura 4 - Fluxograma do trabalho.	20
Figura 5 - Matriz de Correlação.	22
Figura 6 –Janela operacional encurtada de cada sensor. Erro! Indicador não definido.	
Figura 6 - Janela operacional completa 'P-TPT'.....	23
Figura 7 - Janela operacional completa 'T-TPT'	24
Figura 8 - Janela operacional completa 'P-MON-CKP'.....	24
Figura 9 - Janela operacional completa 'T-JUS-CKP'	25

LISTA DE TABELAS

Tabela 1 - Tamanho da amostragem de dados	21
Tabela 2 - informações sobre sensores.....	21
Tabela 3 - hiperparâmetros utilizados nos modelos.....	26
Tabela 4 - Métricas para um poço normalizado pela média.....	28
Tabela 5 - Métricas do treinamento com todos os poços normalizados pela média.....	29
Tabela 6 - Métrica para um poço normalizado pelo desvio padrão.....	30
Tabela 7 - Métricas do treinamento com todos os poços normalizados pelo desvio padrão.....	30

SUMÁRIO

1	INTRODUÇÃO	9
1.1	Objetivos	Erro! Indicador não definido.
1.2	Objetivos Específicos.....	Erro! Indicador não definido.
1.3	Justificativa	Erro! Indicador não definido.
2	REFERENCIAL TEÓRICO	12
2.1	Produção em Poços de Petróleo.....	12
2.2	Completação Inteligente	13
2.3	Inteligência Artificial	15
2.3.1.	K-Nearest Neighbors (KNN).....	16
2.3.2.	Support Vector Machine (SVM)	17
2.3.3.	Non-Myopic Early Classifier.....	17
2.3.4.	Anomalias de Produção	18
3	METODOLOGIA	20
3.1	Fonte dos Dados.....	21
3.2	Visualização do Problema.....	23
3.3	Treinamento das Redes	25
3.4	Avaliação de Desempenho.....	26
4	RESULTADOS E DISCUSSÕES	28
5	CONCLUSÕES	31
	REFERÊNCIAS	33

1 INTRODUÇÃO

De acordo com projeções da Organizações dos Países Exportadores de Petróleo (OPEP) (2022), a demanda global por petróleo deverá apresentar um aumento significativo ao longo do tempo. Essas projeções indicam que a demanda, na referência da OPEP, o consumo de petróleo continuará aumentando até atingir um pico de 110 milhões de barris por dia em 2035.

A indústria de óleo e gás tem importância reconhecida dentro da economia global, sendo responsável pela grande parcela da oferta energética mundial. Além disso, a exploração de hidrocarbonetos envolve grandes desafios e riscos, como os adversos ambientes instaurados para a exploração, como águas profundas e regiões árticas e a necessidade de lidar com problemas ambientais e de segurança.

A utilização de ferramentas de inteligência artificial com aprendizado de máquina vem se tornando cada vez mais comum em diversas áreas, inclusive na indústria de exploração de hidrocarbonetos. A classificação de eventos de falhas de integridade em sistemas de revestimentos de poços pode ser uma solução interessante para a indústria de óleo e gás, trazendo um melhor entendimento do problema, resultando na prevenção de acidentes e redução do prejuízo financeiro.

O trabalho de Vargas et al. (2019), disponibiliza dados coletados ao longo de 10 anos em 45 poços de petróleo, com mais de 4.500 sensores registrando informações em tempo real. O conjunto de dados pode ser utilizado para desenvolver algoritmos de detecção ou classificação automática de eventos indesejados, contribuindo para a redução de custos e aprimoramento da segurança nas operações de produção de petróleo.

Além disso, o trabalho tem como objetivo classificar eventos anômalos de produção reais, simulados e desenhados por especialistas, disponíveis no repositório [3W Dataset](#) (VARGAS et al, 2019), utilizando técnicas de inteligência artificial como *K-nearest neighbors* (KNN), *Support Vector Machine* (SVM) e *Non-Myopic Early Classifier* para as classificações das anomalias em oito diferentes classes, sendo elas: aumento abrupto do *Basic Sediments and Water* (BSW), fechamento espúrio da *downhole safety valve* (DHSV), *Severe Slugging*, instabilidade do fluxo, queda rápida de produtividade, restrição rápida na *Production Choke* (PCK), *Scaling in PCK* e formação de hidratos nas linhas de produção.

1.1 Objetivos

O objetivo do presente trabalho é desenvolver um modelo automatizado com técnicas de IA para classificar anomalias durante a produção de hidrocarbonetos utilizando as técnicas SVM, KNN e *Non-Myopic Early Classifier*.

1.2 Objetivos específicos

- Treinar redes de IA com as técnicas SVM, KNN e *Non-Myopic Early Classifier* para classificar anomalias em dados de produção de poços de hidrocarbonetos;
- Comparar as métricas das técnicas em relação a outros trabalhos similares para validação;
- Analisar as características dos dados de produção de poços de hidrocarbonetos que possam afetar a precisão dos modelos e propor soluções para esses problemas.

1.3 Justificativa

O treinamento de uma IA para atuar no contexto analisado é uma necessidade emergente devido à crescente demanda por sistemas mais autônomos, menos dependentes da análise humana. Esses sistemas são capazes de reduzir significativamente o tempo de detecção de problemas, garantindo a segurança da operação de intervenção de poços e contribuindo para a minimização dos custos associados.

A detecção rápida e precisa de anomalias é um requisito crítico para a operação de intervenção de poços. Isso requer a análise de grandes quantidades de dados gerados por sensores, o que pode ser uma tarefa demorada e suscetível a erros humanos. Ao utilizar uma IA para o treinamento de classificação de anomalias em dados de produção de poços de hidrocarbonetos, é possível garantir a precisão e rapidez na detecção de anomalias, evitando possíveis problemas operacionais.

Além disso, a aplicação de uma IA treinada nesse contexto pode contribuir significativamente para a redução dos custos associados à operação de intervenção de poços. Ao detectar e classificar rapidamente as anomalias, é possível garantir uma intervenção mais assertiva, evitando retrabalhos e prejuízos financeiros. Por fim, é importante salientar que a utilização de uma IA treinada para esse propósito pode ser uma contribuição significativa para a comunidade acadêmica e industrial.

2 REFERENCIAL TEÓRICO

Este capítulo consiste numa revisão bibliográfica inicial acerca dos temas tratados no trabalho, envolvendo sistemas de poços de petróleo, sistemas de completação inteligente e anomalias associadas, bem como a utilização de inteligência artificial para a classificação das falhas.

2.1 Produção em Poços de Petróleo

Para o melhor entendimento do sistema de produção, este tópico está focado em destrinchar o sistema de produção em poços de petróleo. De acordo com Filho (2013):

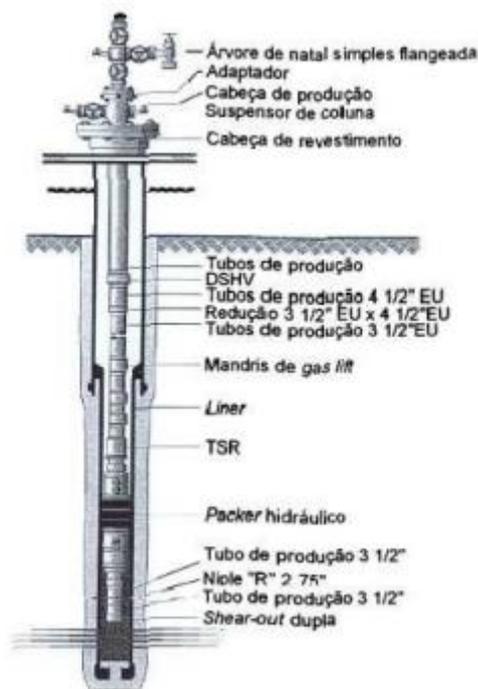
- **Árvore de Natal:** é um conjunto de equipamentos utilizados para o controle e monitoramento da produção. É composta por válvulas, tubos e componentes para regular fluxos, bem como permitir a injeção de fluidos. Ela é extremamente necessária em um sistema de produção.
- **Cabeça do poço:** é uma estrutura instalada no topo de um poço de óleo e/ou gás utilizada para permitir conexões de equipamentos para produção/injeção de fluidos. Por meio dela, realiza-se a vedação do poço, evitando fugas de fluidos para a superfície.
- **Revestimentos:** são tubos de aço instalados no interior do poço perfurado com o objetivo de evitar o desmoronamento das paredes dos poços e isolar zonas de produção. Existem alguns tipos de revestimento, sendo eles: condutor, superfície, intermediário e produção. Em geral, cada camada de revestimento instalada deve ser cimentada completamente ou parcialmente para garantir a integridade do poço.
- **Coluna de produção:** é uma coluna de aço por onde o fluxo de hidrocarbonetos é produzido. Ela faz a conexão da cabeça do poço à produção de óleo.
- **Anulares:** o espaço que dista a coluna de produção ao revestimento de produção é denominado de anular A, sendo o anular B o espaço entre o revestimento de produção ao revestimento seguinte. Esses anulares são importantes para observar possíveis anomalias relacionadas ao

revestimento ou à coluna, além de serem utilizados na injeção de fluidos.

- *Blowout Preventer* (BOP): é um equipamento de segurança crítico utilizado nos sistemas de produção utilizado durante a perfuração. Tem como função proteger o poço contra explosões causadas pelo descontrole da pressão interna. São projetados para suportar altas pressões e temperaturas e são acionados remotamente interrompendo o fluxo de fluidos.

É possível ver na Figura 1 com detalhamento o esquemático simples de um poço petróleo contendo equipamentos essenciais para a exploração e segurança.

Figura 1 - Esquemático de poço.



Fonte: Thomas (2001).

2.2 Completação Inteligente

A completção inteligente de poços consiste em um sistema avançado que tem a capacidade de coletar, transmitir e analisar informações cruciais relacionadas à completção, produção e reservatório. Além disso, a tecnologia permite a tomada de medidas necessárias para controlar efetivamente o poço e otimizar os processos de produção. A grande vantagem dessas tecnologias é a possibilidade de modificar ativamente o dinamismo do poço por meio do controle de fluxo. Sendo assim, é

possível monitorar a resposta e o desempenho das zonas de produção em tempo real por meio da aquisição de dados do fundo do poço. Dessa forma, permite-se maximizar o valor do ativo por meio de uma gestão mais eficiente e inteligente da produção de hidrocarbonetos. (MOORE et al, 2002).

Segundo Garcia (1997), após o término da perfuração o poço é abandonado temporariamente com tampões de cimento para então ser completado, sendo executadas as seguintes fases de completação:

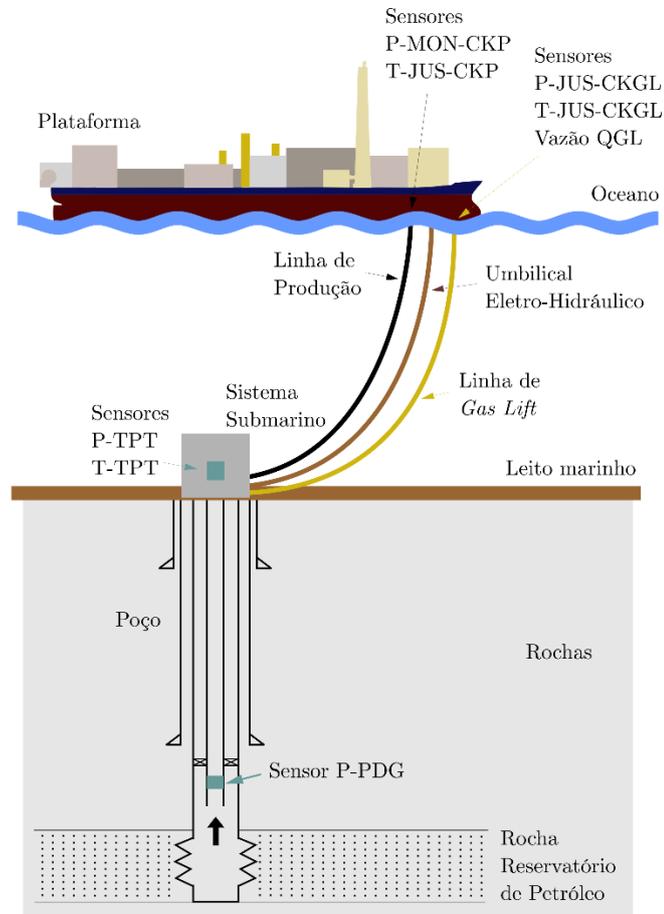
- Instalação dos equipamentos de segurança do poço;
- Condicionamento do revestimento de produção;
- Condicionamento do *liner* (se tiver) e substituição do fluido nele contido por fluido de completação;
- Verificação da qualidade de cimentação primária realizada pela perfuração, e correção, se necessário;
- Canhoneio das regiões de interesse, para comunicação do reservatório interior com o revestimento de produção;
- Descida da coluna de produção;
- Instalação dos equipamentos do interior de poço, para garantir a produção de forma segura e eficiente;
- Instalação da árvore de natal, molhada ou convencional;
- Indução à surgência.

Desse modo, o sensoriamento do poço durante sua vida em operação é de extrema importância. Os sensores são capazes de monitorar temperatura e pressão e controlar os diversos processos de produção de hidrocarbonetos, a cada instante e em diferentes locais. Tais locais são planejados previamente e são comumente utilizados nas válvulas do sistema para garantir segurança e confiabilidade da operação, permitindo a visualização da assinatura de produção normal e assinatura de produção anômala, correlacionando-as com o tempo.

Os sensores de fundo do poço *Permanent Downhole Gauge* (PDG), localizado na coluna de produção; Transmissor de Pressão e Temperatura (TPT), localizado na Árvore de Natal e os sensores da válvula *Choke* (CKP) são responsáveis pelo registro contínuo dos parâmetros de temperatura e pressão. A análise desses dados permite que o operador tome decisões importantes para otimizar a produção e contornar

problemas associados, a Figura 2 exemplifica os sensores e apresenta um esquemático de produção offshore.

Figura 2 – Esquema simplificado de um sistema de produção de petróleo.



Fonte: Fernandes Júnior, 2020.

2.3 Inteligência Artificial

A inteligência artificial pode ser definida como um conjunto de técnicas e algoritmos que permitem às máquinas aprender e realizar tarefas que antes eram exclusivas dos seres humanos. Na prática, a indústria de petróleo utiliza essas técnicas para redução do tempo de resolução do problema e minimização do custo dele. Por meio de classificação e previsão de comportamentos a partir de dados que são extraídos, por exemplo, do sensoriamento dos poços. Isso implica diretamente na segurança e fluidez da operação, beneficiando as empresas e seus colaboradores.

Nos últimos tempos, a complexidade dos problemas computacionais tem-se mostrado mais intensa. A quantidade de dados gerados pelos diversos setores é maior a cada dia, fazendo-se necessário o uso de ferramentas computacionais mais

sofisticadas e robustas, reduzindo a intervenção humana e dependência de especialistas. Para isso, essas técnicas deveriam ser capazes de criar por si próprias, a partir da experiência passada, uma hipótese capaz de resolver o problema que se deseja tratar. A esse processo de indução a uma hipótese a partir de uma experiência anterior é denominado de Aprendizado de Máquina (AM) (FACELI et al, 2011).

No campo do aprendizado de máquina, os métodos podem ser divididos em duas categorias principais: o aprendizado supervisionado e o aprendizado não supervisionado. No aprendizado supervisionado, fornecemos ao algoritmo exemplos de dados com resultados esperados, permitindo que ele aprenda a fazer previsões ou classificações com base nesses exemplos. Por outro lado, no aprendizado não supervisionado, não fornecemos ao algoritmo resultados esperados. O algoritmo precisa identificar padrões ou estruturas nos dados por conta própria. Isso é útil quando não temos rótulos claros ou quando queremos explorar a estrutura oculta nos dados. Dentre as aplicações mais difundidas dos modelos do AM destacam-se as de previsão e classificação. O presente trabalho tratará de problemas de classificação de dados e AM de dados supervisionados.

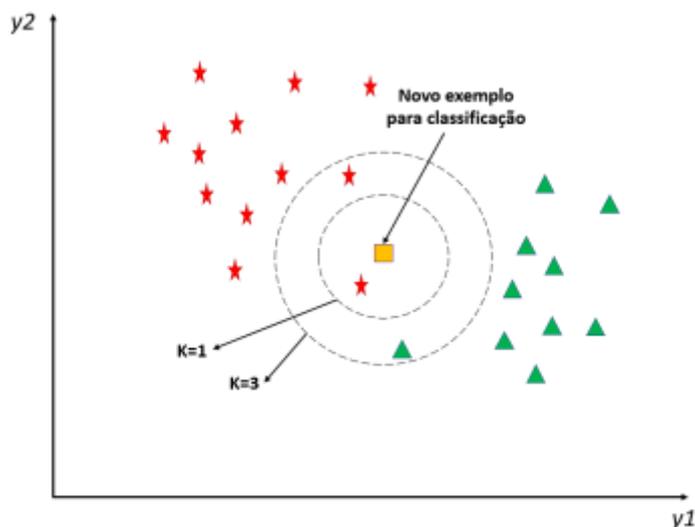
Um requisito para uma boa funcionalidade do AM é que ele seja capaz de tratar dados com presença de ruídos, ausência de dados, dados inconsistentes. Entretanto, em casos em que as imperfeições dos dados são extensas, pode haver prejuízos aos resultados esperados, atrapalhando o processo indutivo. Nesses casos, é necessário utilizar técnicas de pré-processamento dos dados para diminuir a ocorrência desses problemas (FACELI et al, 2011).

2.3.1. K-Nearest Neighbors (KNN)

Este algoritmo apresenta a técnica de aprendizado de máquina supervisionado que pode ser usado tanto para classificação quanto para regressão. A ideia é classificar ou prever um ponto dentro de um conjunto de dados com base na maioria dos “vizinhos” mais próximos, sendo essa proximidade calculada normalmente pela distância Euclidiana. O "K" em KNN denota o número de vizinhos mais próximos que o algoritmo deve considerar ao tomar uma decisão. Por exemplo, se K for definido como 3, o algoritmo considerará os três pontos mais próximos para tomar sua decisão. A escolha do valor de K pode afetar significativamente os resultados. Um K muito pequeno pode levar a um modelo instável e sensível ao ruído, enquanto um K muito grande pode suavizar os limites de decisão. A escolha do KNN para ser

utilizada neste trabalho parte da imensa aplicabilidade e recorrência do uso técnica que é robusta, mas com ótima capacidade de resolução de problemas. Pode-se ilustrar com a representação da Figura 3 abaixo.

Figura 3 - Ilustração do método KNN.



Fonte: Nascimento (2022).

2.3.2. Support Vector Machine (SVM)

O SVM é um poderoso algoritmo de aprendizado de máquina para dados supervisionados usado para classificação, regressão e detecção de outliers. Uma das características é sua eficácia em espaços de grandes dimensões, tornando-as ideais para lidar com problemas complexos e conjuntos de dados com muitos recursos. A técnica é eficiente em termos de memória, uma vez que utilizam apenas um subconjunto dos pontos de treinamento, conhecidos como vetores de suporte, na função de decisão. A versatilidade do SVM é realçada pela possibilidade de especificar diferentes funções de kernel para a função de decisão, permitindo a adaptação a uma variedade de problemas, em específico, no presente trabalho é utilizado SVM com o *kernel* linear. A escolha para esse algoritmo foi baseada em outros trabalhos com o mesmo tipo de proposta, classificação de anomalias.

2.3.3. Non-Myopic Early Classifier

Destaca a importância sequencial dos dados para determinados problemas e busca a otimização do equilíbrio entre tempo e precisão que geralmente é um problema complexo. No momento do teste, a previsão é feita em um carimbo de data/hora de modo que a precisão de antecipação esperada seja otimizada, o que pode, portanto, variar entre as séries temporais. A escolha deste sistema de

aprendizado de máquina se deu na sua capacidade de tomar decisões a longo prazo levando em consideração não apenas as informações imediatas disponíveis, mas também considerando o impacto de suas decisões no futuro.

2.3.4. Anomalias de Produção

As anomalias de produção de petróleo referem-se a desvios não planejados ou indesejados nos padrões de produção de petróleo em campos petrolíferos. Entre as várias anomalias de produção, Vargas et al. (2019) traz a perspectiva de oito anomalias em seu trabalho. São elas:

- **Aumento abrupto do BSW (*Basic Sediment and Water*):** definido como a proporção entre a taxa de fluxo de água e sedimento e a taxa de fluxo de líquido, ambos medidos sob condições normais de temperatura e pressão (CNTP) (Andreolli, 2016 e Abass and Bass, 1988). Ao longo do ciclo de vida de um poço de petróleo, espera-se que seu índice de BSW (Sedimento Básico e Água) aumente devido ao crescimento na produção de água, do aquífero natural do reservatório ou de injeções artificiais para evitar a diminuição da produção. No entanto, um aumento repentino do BSW pode acarretar diversos problemas relacionados à garantia do fluxo, à redução na produção de petróleo, ao levantamento do óleo, à formação de incrustações, ao processamento nas instalações industriais e ao fator de recuperação. A identificação automática de eventos indesejados desse tipo pode permitir a adoção de medidas como o ajuste na produção ou a injeção controlada para prevenir esse tipo de problema (VARGAS et al,2019).
- **Fechamento espúrio da DHSV:** as vezes, o fechamento dessa válvula podem acontecer inesperadamente e podendo ocorre sem sinais visíveis na superfície. Detectar automaticamente esse fechamento inesperado da válvula de forma rápida possibilita sua reabertura por meio de ações operacionais de correção. Isso ajuda a evitar perdas na produção e custos adicionais.
- **Severe Slugging (intermitência severa):** Trata-se de uma forma significativa de instabilidade, destacando-se por duas características distintas. A primeira é a presença de um padrão regular e previsível, geralmente ocorrendo a cada 30, 45 ou 60 minutos. A segunda

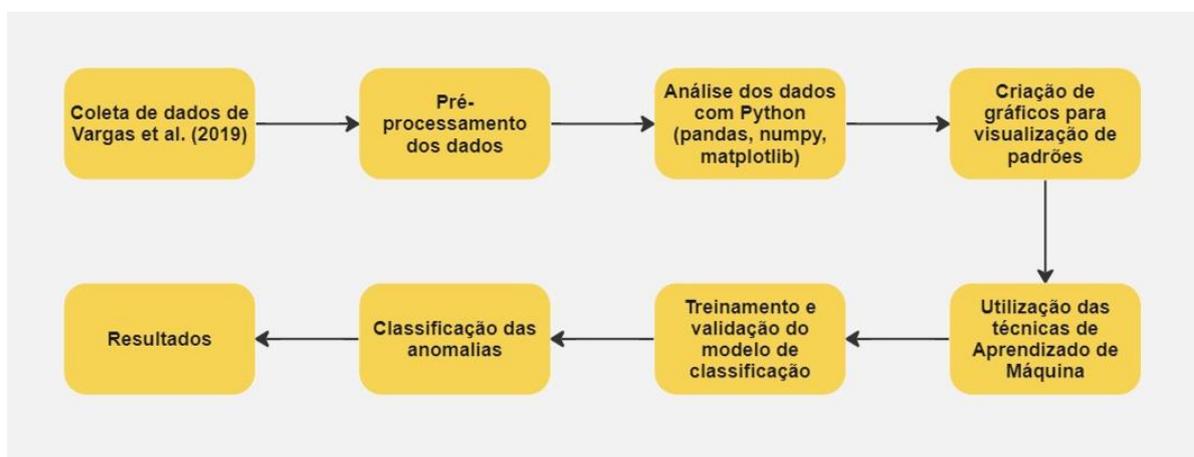
característica notável é a sua intensidade, que é frequentemente tão pronunciada que pode ser detectada por sensores em toda a linha de produção (Meglio et al., 2012), (Schmidt et al., 1985).

- **Instabilidade de fluxo:** é um fenômeno significativo para a indústria de petróleo e tem ocorrência elevada, conhecido também por golfadas. Ocorre quando uma ou mais variáveis monitoradas sofre mudanças substanciais, porém dentro de uma amplitude tolerável. Se caracteriza por ausência de periodicidade e falta de padrão, ao identificar uma instabilidade de fluxo evita-se que se torne um *Severe Slugging* podendo economizar recursos, tempo e riscos associados a interrupções não planejadas na produção.
- **Perda rápida de produção:** A produtividade de um poço de fluxo natural depende de várias propriedades, como a pressão estática do reservatório, a porcentagem de sedimento básico e água, a viscosidade do fluido produzido, o diâmetro da linha de produção, entre outros (Hausler et al., 2015). Quando essas características são modificadas a ponto de não mais fornecerem energia adequada para compensar as perdas, o fluxo diminui de velocidade ou pode até parar completamente.
- **Restrição rápida na *Choke* de produção:** situação em que essa válvula é fechada abruptamente em um curto intervalo de tempo, e isso pode ocorrer manualmente devido a problemas operacionais. Detectar esse fechamento indesejado de forma automática é importante, pois permite uma reversão mais rápida da situação.
- **Depósitos na *choke* de produção:** a deposição de material na *choke* reduz significativamente a produção pela restrição física, sendo assim é necessário monitoramento contínuo.
- **Hidratos na linha de produção:** a partir de água e gás natural somado às condições ambientais, ocorre a deposição de hidratos nas linhas de produção, além de altas pressões e baixas temperaturas (Andreolli, 2016 e Ellison et al., 2000). A manutenção periódica e monitoramento é o maior aliado contra essa anomalia de produção.

3 METODOLOGIA

A metodologia deste trabalho consiste em utilizar os dados disponibilizados por Vargas et al. (2019) com a utilização de linguagem de programação *Python* somado ao Aprendizado de Máquina para classificar as falhas já destacadas, sendo o KNN, SVM e *Non Myopic* as técnicas aplicadas. A metodologia segue o fluxograma da Figura 4.

Figura 4 - Fluxograma do trabalho



Fonte: Autor (2023).

Antes de treinar as redes, na etapa de pré-processamento dos dados ocorrerá a retirada dos dados ausentes, em branco ou faltantes, normalização particionada no tempo, em intervalos de 10 minutos de dados para o processamento computacional menos custoso e a seleção de melhores variáveis para a classificação de anomalias determinada pelo estudo dos dados em relação a instabilidade de fluxo. Além disso o uso da padronização dos dados pela média e desvio padrão, separadamente, com o intuito de colocar todos os nossos dados em escalas próximas para que o modelo entenda cada um deles como igualmente importantes. A utilização da linguagem de programação Python com as bibliotecas mencionadas tem como funcionalidade a análise primária dos dados e a sua visualização. Os algoritmos foram escolhidos por serem técnicas de aprendizado de máquina que têm apresentado bons resultados em problemas de classificação, sendo capaz de separar dados em classes distintas com base em critérios de otimização matemática. Além disso, conseguem lidar com dados em alta dimensão, o que é comum em problemas da indústria de petróleo e gás, onde há muitos sensores coletando informações em tempo real. Portanto, as técnicas se mostraram uma escolha adequada para classificar as anomalias de produção de hidrocarbonetos presentes no conjunto de dados disponibilizado.

3.1 Fonte dos Dados

Os dados utilizados neste trabalho foram obtidos do repositório disponibilizado por Vargas et al. (2019), que contém dados simulados, desenhados por especialistas e dados reais coletados na bacia de Campos, localizada entre a costa norte do estado do Rio de Janeiro e o sul do Espírito Santo. As instâncias reais são aquelas que ocorreram nos poços reais da Petrobras durante a produção de óleo, os simulados pelo computador e projetadas a mão por especialistas com vasto conhecimento na área tem o objetivo de desequilibrar do conjunto inicial formado apenas por instâncias reais. Os dados estão disponíveis em formato *Comma-Separated Values* (CSV) e incluem informações como pressão, temperatura e outras variáveis relevantes para o monitoramento da produção de hidrocarbonetos. O intervalo de tempo é de dezembro de 2013 a abril de 2018, totalizando 16 meses. Todas as instâncias, sejam elas reais, simuladas ou desenhadas foram disponibilizadas em taxa de amostragem fixa (1 Hz) e tamanho de 4,78 GBytes, com as unidades em Pascal (Pa), metros cúbicos por segundo (m³/s) e grau Celsius (°C). Na Tabela 1, é possível dimensionar o volume de dados disponibilizados.

Tabela 1 - Tamanho da amostragem de dados

	Real	Simulado	Desenhado
Normal	9956791	0	0
Anômalo	4087211	33902803	2966410

Fonte: autor (2023).

Ademais, o repositório 3W Dataset conta com 8 variáveis que são atribuídas aos dados dos sensores do sistema de equipamentos de produção hidrocarbonetos, a Tabela 2 possui informações sobre estes.

Tabela 2 - Informações sobre sensores.

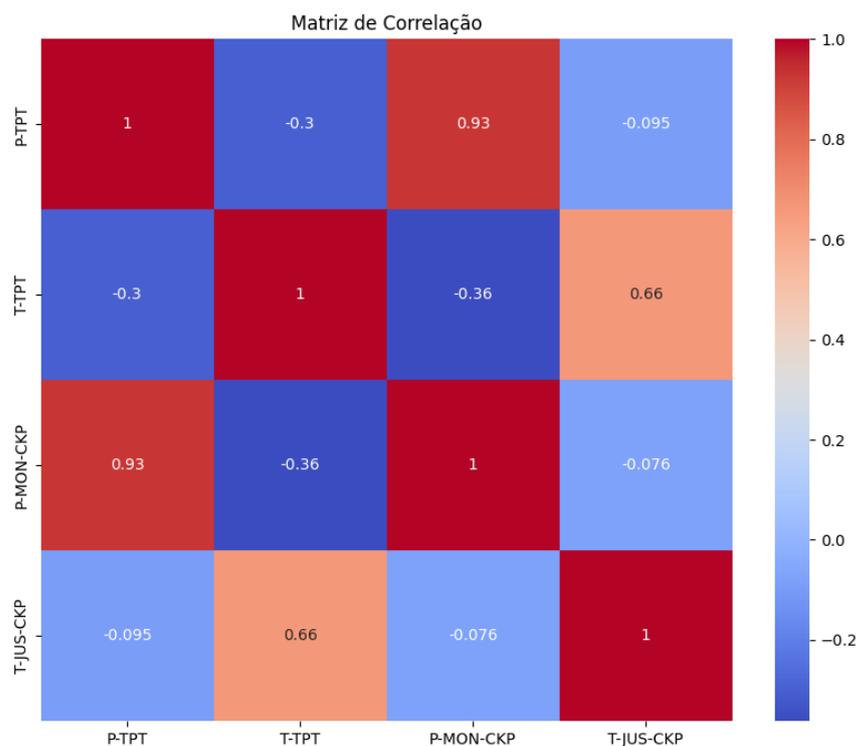
Sigla	Nome	Especificação
PDG	<i>Permanent Downhole Gauge</i>	Sensor de pressão localizado no interior do poço
P-TPT	<i>Pressure at the Temperature and Pressure Transducer</i>	Sensor de pressão localizado na árvore de natal
T-TPT	<i>Temperature as the Temperature and Pressure Transducer</i>	Sensor de pressão localizado na árvore de natal
P-MON-CKP	<i>Pressure upstream the choke and production valve</i>	Sensor de pressão à montante da válvula choque de produção

T-JUS-CKP	<i>Temperature downstream the choke production valve</i>	Sensor de temperatura à jusante da válvula choque de produção
P-JUS-CKGL	<i>Pressure downstream choke gas lift valve</i>	Sensor de pressão à jusante da válvula choque de gás lift
T-JUS-CKGL	<i>Temperature downstream choke gas lift valve</i>	Sensor de temperatura à jusante da válvula choque de gás lift
QGL	<i>Gas lift flow</i>	Vazão de gás lift

Fonte: Santos (2020).

Outro fator interessante dos dados é como eles se relacionam entre si. De forma que quando um cresce o outro também ou o inverso tratado negativamente. E, portanto, observar a forma com que a o problema de instabilidade de fluxo se relaciona com os sensores. Diante disso, faz sentido a análise da matriz de correlação (Figura 5) desses dados. Uma matriz de correlação é uma tabela de dados que ilustra as relações estatísticas entre variáveis em um conjunto de dados. Ela é usada para calcular e resumir como diferentes variáveis são relacionadas umas com as outras. Essas relações são expressas na forma de coeficientes de correlação, que indicam a força e a direção das associações entre variáveis. Para valores positivos existe uma correlação diretamente proporcional e o contrário para os valores negativos.

Figura 5 - Matriz de Correlação.



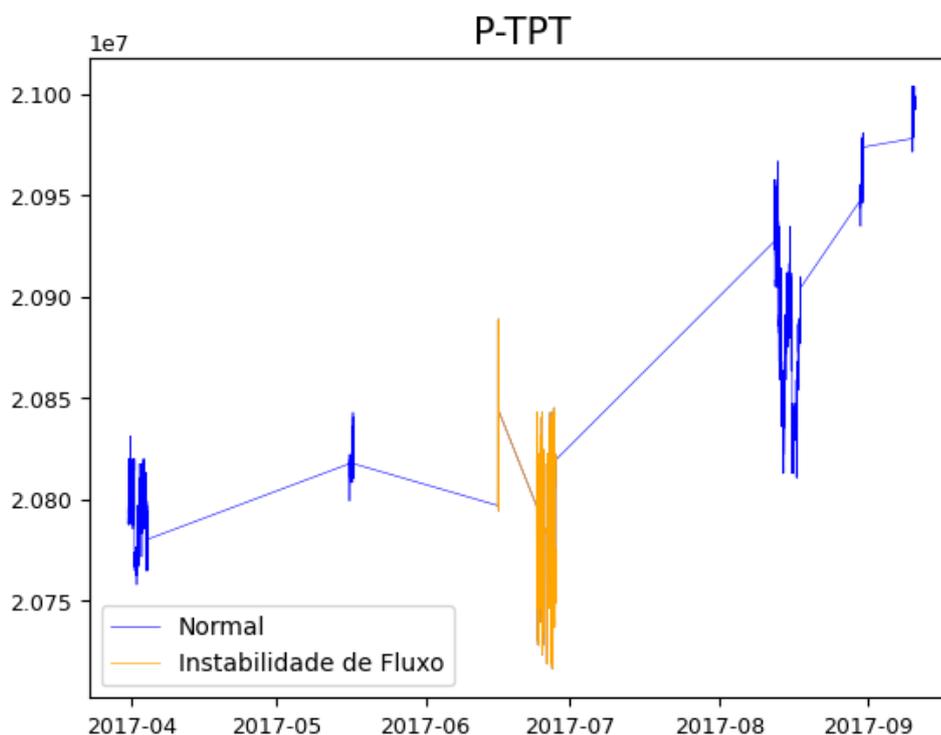
Fonte: autor (2023).

Podemos observar uma forte relação entre a pressão a montante da válvula choke e a pressão do sensor TPT, com um coeficiente de correlação de 0,93. Isso sugere que quando ocorrem variações na pressão do sistema, elas são sentidas de forma consistente por todos os sensores de pressão, independentemente de sua localização. Essas correlações indicam que as variáveis estão interconectadas de alguma forma, e mudanças em uma delas tendem a afetar as outras. Isso pode ser útil para compreender o comportamento do sistema e suas anomalias.

3.2 Visualização do Problema

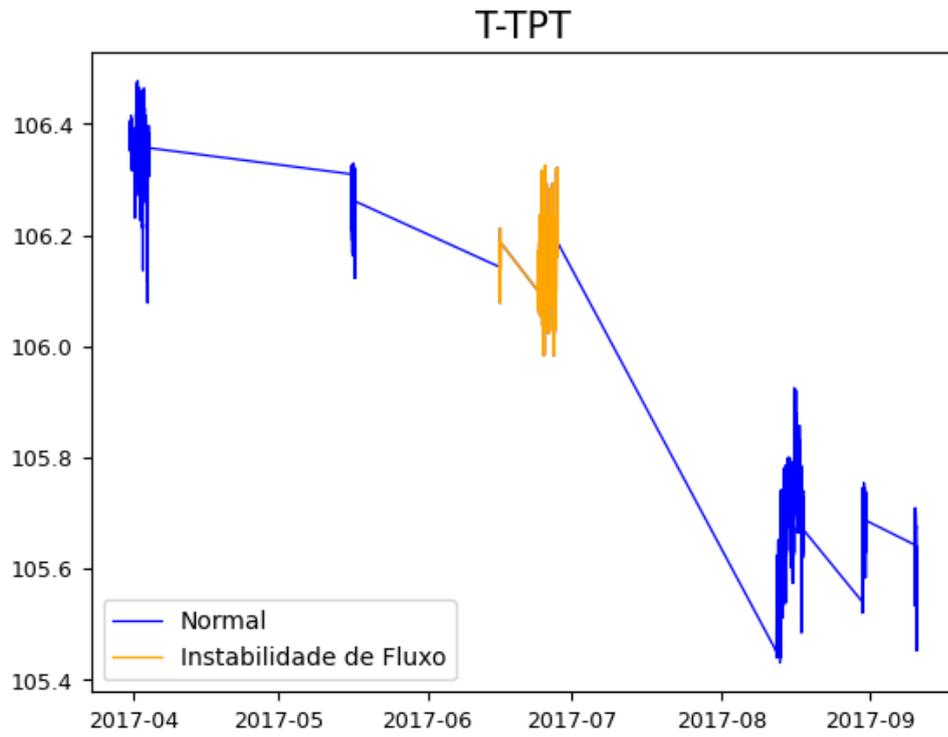
Uma análise relevante para o problema é a avaliação visual. Com isso, é possível colocar-se no lugar do operador, onde este terá que tomar decisões importantes para ajuste de produção caso algum erro seja identificado, outrossim a verificação quanto a coerência da solução encontrada pela rede treinada. Nas Figuras 6, 7, 8 e 9 observar-se a produção normal, destacada em azul, e a produção anômala de instabilidade de fluxo em laranja previamente supervisionados. Os dados crus são apresentados em na janela operacional de março a novembro de 2017 e mostra a assinatura da anomalia de instabilidade de fluxo.

Figura 6 - Janela operacional completa 'P-TPT'.



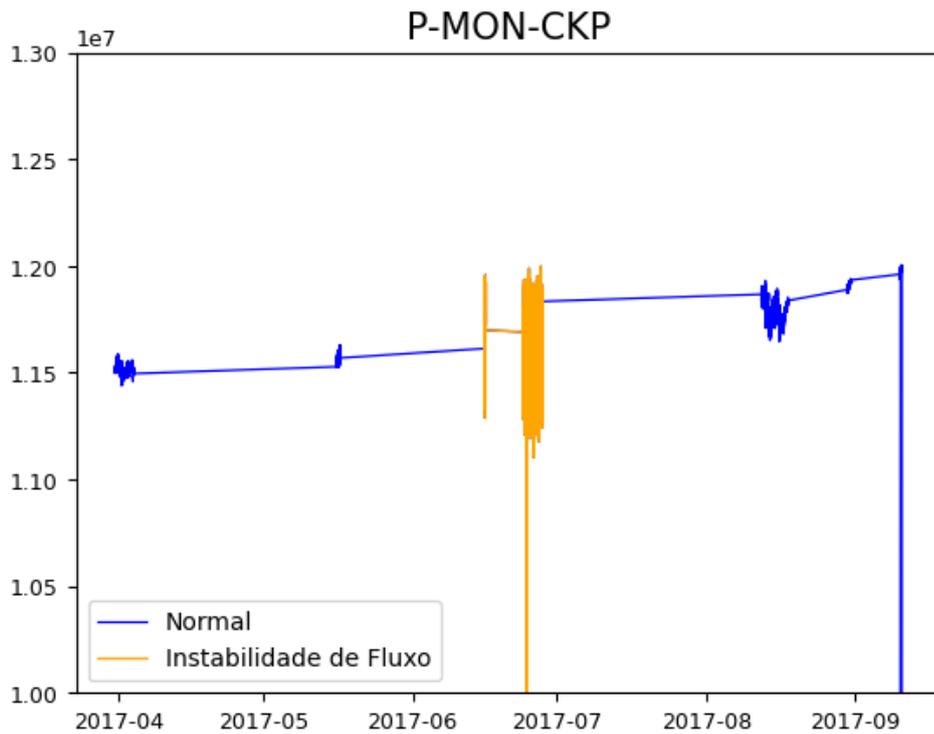
Fonte: Autor (2023).

Figura 7 - Janela operacional completa 'T-TPT'



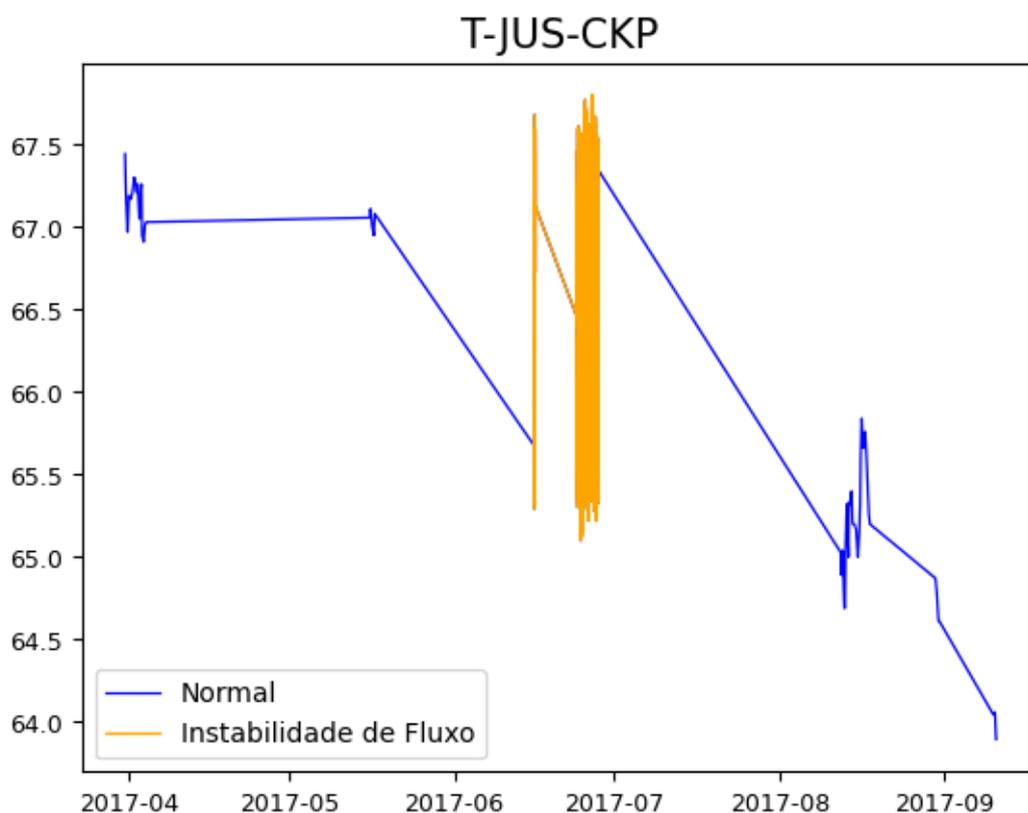
Fonte: Autor (2023).

Figura 8 - Janela operacional completa 'P-MON-CKP'



Fonte: Autor (2023).

Figura 9 - Janela operacional completa 'T-JUS-CKP'



Fonte: Autor (2023).

Nestes exemplos, seja em escala maior ou escala menor de tempo, é notável a dificuldade para identificar a assinatura da anomalia de produção caso não houvesse a supervisão para isso, outro ponto é como os sensores sofrem intensidade de variações em diferentes escalas, com esse conhecimento podemos entender como a influência do erro humano pode ser significativa para produção de hidrocarbonetos. Por isso, faz-se necessário uma análise que leve em conta a expertise humana somado a uma logica computacional eficaz para a identificação desse estado de produção anômalo.

3.3 Treinamento das redes

Inicialmente, no estágio embrionário, foi separado um único poço com anomalia de instabilidade de fluxo, com intuito de facilitar o entendimento do problema e gradualmente aumentar a dificuldade com os resultados que foram obtidos. Posteriormente, houve a expansão do problema para os demais poços que contém a anomalia. Foram utilizados hiperparâmetros de treinamentos idênticos para as redes

com o intuito de comparar as métricas de forma mais concisa sendo determinado, em todos os casos, o parâmetro de tamanho de amostra de teste de 20% e, por consequência, 80% para treinamento. Além disso, a semente do gerador de números aleatório (*random_state*) igual a 1, isso é útil para garantir que os resultados sejam consistentes e comparáveis entre diferentes execuções do código ou entre diferentes pessoas que trabalharem no mesmo projeto.

Na Tabela 3 pode-se ver os hiperparâmetros utilizados nos modelos, onde estes são modificados manualmente para obter a maior assertividade e melhores métricas.

Tabela 3 - Hiperparâmetros utilizados nos modelos.

Modelos	Hiperparâmetros
KNN	k = 1, p=2, weight= "uniform"
SVM	kernel='linear', C=1.0
Non-Myopic	n_clusters=2, cost_time_parameter=1e-3, lamb=1e2

Fonte: Autor (2023).

É fundamental ressaltar que o conjunto de dados em questão apresenta um desequilíbrio significativo entre as classes, o que representa um desafio adicional em comparação com dados balanceados, nos quais as quantidades de exemplos positivos e negativos são comparáveis. Mais especificamente, é observado que as instâncias de anomalias de produção são substancialmente menos frequentes em comparação com as instâncias de produção não anômala. Esse desequilíbrio pode criar dificuldades para os modelos de aprendizado de máquina, uma vez que esses modelos podem inclinar-se em direção à classe majoritária, negligenciando a classe minoritária e, como resultado, gerando previsões imprecisas para a classe minoritária.

3.4 Avaliação de Desempenho

Faz-se necessária a avaliação adequada do desempenho das redes treinadas por métricas usuais neste tipo de problema. De acordo com Kowsari et al. (2019), essa avaliação é feita por métricas, tais como acurácia, precisão e indicador *F1 score*. No presente trabalho foi utilizada algumas métricas para a análise, são elas:

- Acurácia: demonstra o quão frequente o classificador está correto e indica a porcentagem de amostras indicadas adequadamente em relação a todo conjunto de dados.

- Acurácia balanceada: leva em consideração o desequilíbrio de classes em um conjunto de dados. Ela calcula a média das taxas de verdadeiros positivos para cada classe.
- Precisão: proporção de exemplos classificados como positivos que são verdadeiros positivos. Em outras palavras, é a capacidade do modelo de evitar falsos positivos.
- Revocação: mede a proporção de exemplos positivos que foram corretamente identificados pelo modelo. Em outras palavras, é a capacidade do modelo de encontrar todos os verdadeiros positivos.
- *F1 score*: métrica que combina precisão e revocação em uma única medida. É especialmente útil quando se deseja encontrar um equilíbrio entre essas duas métricas.

Portanto, cada uma dessas métricas fornece uma perspectiva diferente sobre o desempenho do modelo de classificação, é ideal examinar várias métricas em conjunto para obter uma compreensão completa do desempenho do modelo.

4 RESULTADOS E DISCUSSÕES

A proposta para a classificação de anomalias de produção é dada inicialmente para obter uma maior assertividade e segurança nas tomadas de decisões durante a produção de um poço. Isso fica claro com o cenário da união de forças do humano com a inteligência artificial aumenta significativamente o potencial de acerto sobre a anomalia atuante ou o estado de produção normal. A consequência disso é a redução de custos com paradas de operação e manutenção de equipamentos.

Pode-se observar que a utilização de técnicas de pré-processamento e tratamento dos dados pode ter uma grande influência no resultados das métricas analisadas, sendo que inicialmente, no estágio embrionário do processo de treinamento da rede, os resultados obtidos sem a utilização adequada do tratamento dos dados e redimensionamento pela média ou/e desvio padrão foi abaixo do esperado, sendo necessário ajustes para adequar a uma resposta mais eficiente do algoritmo para assertividade do problema.

Desse modo, inicialmente com o treinamento em cima dos dados crus, utilizando o método *Non Myopic* e todos os sensores ao mesmo tempo foi obtido os valores de 74,68% para acurácia dos dados sem normalização.

Por outro lado, com a normalização a partir do *StandardScaler*, o redimensionamento pela média e pelo desvio padrão separadamente e análises de janela temporal com 600 dados, mas analisando ainda para apenas um poço, utilizando os algoritmos de aprendizado de máquina com todos os sensores ou pares de sensores diferentes para treinamento (apresentados na tabela), foi possível alcançar um resultado significativo quanto a classificação da anomalia, isso pode ser visto na Tabela 4. Os sensores foram separados de acordo com a matriz de correlação, sendo mais interessante ao problema a escolha de parâmetros que mudassem de forma diferente um com o outro, escolhendo então os sensores pela baixa correlação de dados.

Tabela 4 - Métricas para um poço normalizado pela média.

Métricas	KNN (Todos)	KNN (‘T-TPT’ E ‘P-TPT’)	KNN (‘P-MON- CKP’, ‘T-JUS- CKP’)	SVM (Todos)	SVM (‘T-TPT’ E ‘P-TPT’)	SVM (‘P-MON- CKP’, ‘T-JUS- CKP’)	NonMy (Todos)	NonMy (‘T-TPT’ E ‘P-TPT’)	NonMy (‘P-MON- CKP’, ‘T-JUS-CKP’)
Acurácia	1	1	95.2	86.21	95.1	73.57	91.54	1	91.54
Acurácia Balanceada	1	1	92.4	82.86	96.41	57.12	93.33	1	93.33

Fonte: autor (2023).

Nesse contexto, as redes demonstraram um desempenho notável devido à simplicidade do problema em questão. Elas se destacaram como os modelos mais eficazes, atingindo consistentemente taxas máximas de acurácia. A única exceção foi o SVM, que, embora não tenha alcançado o mesmo nível de desempenho que os outros, ainda obteve resultados interessantes. Especificamente, na configuração que utilizou apenas os sensores 'T-TPT' e 'P-TPT', o SVM alcançou uma taxa de acurácia de 95%, o que é um resultado significativo e relevante para o contexto do problema.

Já para uma análise mais completa, envolvendo a quantidade total de poços que apresentam a anomalia de instabilidade de fluxo, foi possível perceber diferenças consideráveis quanto as redes programadas. Em pontos específicos uma se sobressai em relação a outra e vice-versa. Isso fica evidenciado na Tabela 5 abaixo.

Tabela 5 - Métricas do treinamento com todos os poços normalizados pela média.

Métricas	KNN (Todos)	KNN ('T-TPT' E 'P-TPT')	KNN ('P-MON-CKP', 'T-JUS-CKP')	SVM (Todos)	SVM ('T-TPT' E 'P-TPT')	SVM ('P-MON-CKP', 'T-JUS-CKP')	NonMy (Todos)	NonMy ('T-TPT' E 'P-TPT')	NonMy ('P-MON-CKP', 'T-JUS-CKP')
Acurácia	95.4	94.38	96.93	87.24	84.18	84.18	94.89	93.36	93.87
Acurácia Balanceada	92.19	90.72	94.79	72.39	66.38	65.58	91.05	89.26	88.78
Precisão	92.68	90.24	95.25	95.23	88.23	93.33	92.5	87.8	92.1
Revocação	86.36	84.09	90.9	45.45	34.09	31.81	84.09	81.81	79.54
F1 score	89.41	87.05	93.02	61.53	49.18	47.45	88.09	84.7	85.36

Fonte: autor (2023).

Diante disso, em relação a acurácia para o treinamento com a normalização pela média, é possível observar que o KNN, utilizando os sensores 'P-MON-CKP' e 'T-JUS-CKP' para o treinamento, apresenta maior valor (96.93%) seguido pelo KNN utilizando todos os sensores (95.4%). O KNN também obteve bons resultados, com acurácia acima de 94% em todos os casos. Além disso, apresentou um bom equilíbrio entre precisão e revocação, indicando que é capaz de identificar corretamente tanto os verdadeiros positivos quanto os verdadeiros negativos. É importante considerar o equilíbrio entre precisão e revocação ao avaliar os resultados. O *F1 score* mede a média harmônica entre precisão e revocação. O KNN com 'P-MON-CKP' e 'T-JUS-CKP' tem o maior *F1 score* (93.02%), enquanto as configurações *Non-Myopic Early Classifier* têm valores próximos e o SVM com valores bem abaixo.

O classificador *Non-Myopic Early Classifier* apresentou uma alta precisão e acurácia, em quase todos os casos com valores acima de 90%. Isso indica que esse classificador é eficiente na identificação dos poços com instabilidade de fluxo. Também se destaca pelos resultados consistentes em todas as métricas. Isso sugere que é uma escolha promissora para futuras aplicações mantendo o *F1 score* considerável nesse contexto.

Por outro lado, para os dados que foram normalizados pelo desvio padrão têm suas métricas mostradas na Tabela 6.

Tabela 6 - Métrica para um poço normalizado pelo desvio padrão.

Métricas	KNN (Todos)	KNN ('T-TPT' E 'P-TPT')	KNN ('P-MON-CKP', 'T-JUS-CKP')	SVM (Todos)	SVM ('T-TPT' E 'P-TPT')	SVM ('P-MON-CKP', 'T-JUS-CKP')	NonMy (Todos)	NonMy ('T-TPT' E 'P-TPT')	NonMy ('P-MON-CKP', 'T-JUS-CKP')
Acurácia	1	1	0.95	0.86	0.95	0.73	1	0.91	1
Acurácia Balanceada	1	1	0.92	0.82	0.96	0.57	1	0.89	1

Fonte: autor (2023).

Nesse contexto, ocorre similaridade com os dados da Tabela 4, a mudança específica foi a alternância das porcentagens de acurácia e acurácia balanceada da coluna com todos os sensores e com os sensores 'T-TPT' E 'P-TPT' do algoritmo *Non-My Opic Early Classifier*. Os demais parâmetros foram idênticos.

Para análise mais completa com todos os poços nesse cenário de instabilidade de fluxo e com a normalização sendo feita pelo desvio padrão, foram obtidos os resultados das métricas na Tabela 7.

Tabela 7 - Métricas do treinamento com todos os poços normalizados pelo desvio padrão.

Métricas	KNN (Todos)	KNN ('T-TPT' E 'P-TPT')	KNN ('P-MON-CKP', 'T-JUS-CKP')	SVM (Todos)	SVM ('T-TPT' E 'P-TPT')	SVM ('P-MON-CKP', 'T-JUS-CKP')	NonMy (Todos)	NonMy ('T-TPT' E 'P-TPT')	NonMy ('P-MON-CKP', 'T-JUS-CKP')
Acurácia	96.95	86.58	97.56	84.75	81.09	84.75	95.12	86.58	96.34
Acurácia Balanceada	95.68	81.91	97.59	73.17	66.2	73.17	93.69	82.66	96.02
Precisão	95.23	75.6	93.33	87.5	83.33	87.5	90.69	74.41	91.11
Revocação	93.02	72.09	97.67	48.83	34.88	48.83	90.69	74.41	95.34
<i>F1 score</i>	94.11	73.8	95.45	62.68	49.18	62.68	90.69	74.41	93.18

Fonte: autor (2023).

Em relação aos dados classificados com o uso da média, a diferença entre os valores das porcentagens foram as grifadas em vermelho, onde essas possuem valor

superior a Tabela 5. Entretanto, a margem de variação desses resultados foi pequena, sendo as duas técnicas viáveis de utilizar para este fim.

É importante considerar que o desequilíbrio entre as classes afeta a interpretação das métricas. Quando se trata de classes não balanceadas, a acurácia pode não ser uma métrica informativa por si só, e métricas como a acurácia balanceada, precisão, revocação e *F1 score* são mais relevantes.

O desafio é encontrar um ponto de equilíbrio que permita identificar corretamente a maioria dos poços com instabilidade de fluxo, sem gerar muitos falsos positivos. Além das métricas, é importante analisar visualmente os resultados por meio de gráficos e matriz de confusão para entender melhor o desempenho do modelo em diferentes classes.

A comparação das métricas alcançadas foi realizada em relação aos resultados anteriores obtidos por Vargas (2019), utilizando a técnica SVM (*One-class*) para detecção de anomalias no mesmo conjunto de dados de referência. Os resultados de Vargas apresentaram uma média de 0,414 e um desvio padrão de 0,195 para as métricas *F1 score* e precisão, respectivamente, em ambas as métricas. É importante ressaltar que as outras metodologias apresentadas neste estudo, ou seja, o KNN e o *Non-Myopic Early Classifier*, não foram utilizadas na análise anterior de Vargas (2019).

5 CONCLUSÕES

Na análise das técnicas aplicadas à detecção de anomalias na produção de poços, tornou-se evidente que algumas estratégias se destacaram como mais eficazes. Em particular, o algoritmo K-Nearest Neighbors (KNN) com os parâmetros específicos e o *Non-Myopic Early Classifier* demonstraram um desempenho superior em termos de acurácia e precisão. O KNN, quando configurado com os sensores 'P-MON-CKP' e 'T-JUS-CKP' para o treinamento, atingiu a impressionante marca de 97,67% de acurácia balanceada. Além disso, o KNN manteve uma acurácia acima de 94% em todas as configurações testadas em relação à média. Esta consistência é um indicativo da robustez do algoritmo na identificação de anomalias com alta precisão.

O impacto do pré-processamento e tratamento adequado dos dados revelou-se um fator crítico na análise de desempenho das técnicas de detecção de anomalias na produção de poços. Em particular, a normalização dos dados pela média e pelo

desvio padrão destacou-se como um aspecto crucial que influenciou significativamente os resultados, contribuindo para a obtenção de resultados mais satisfatórios, desempenhando um papel fundamental na redução das variações nas escalas dos dados. Isso permitiu que os algoritmos fossem treinados de forma mais eficiente. Além disso, a normalização dos dados ajudou a garantir uma interpretação mais confiável das métricas de desempenho, especialmente em um cenário com classes desbalanceadas. Métricas como a acurácia balanceada, precisão, revocação e *F1 score* se tornaram mais informativas e relevantes quando os dados foram adequadamente pré-processados.

Este estudo fornece informações valiosas que têm o potencial de trazer vantagens significativas para as operações da indústria, tanto em termos de tomada de decisões estratégicas quanto na redução de despesas. Características, como a detecção ágil de qualquer desvio dos padrões de produção usuais, fortalecem a confiabilidade das operações e minimizam os riscos de falhas inesperadas.

REFERÊNCIAS

- ABASS, H.; BASS, D. **The Critical Production Rate in Water-Coning System**. Permian Basin Oil and Gas Recovery Conference. Society of Petroleum Engineers. Texas, 1988. p. 351- 360.
- ANDREOLLI, Ivanildo. Introdução à elevação e escoamento monofásico e multifásico de petróleo. **Rio de Janeiro: Interciência**, 2016.
- ELLISON, B.T., GALLAGHER, C.T., FROSTMAN, L.M., e S.E. LORIMER. "**The Physical Chemistry of Wax, Hydrates, and Asphaltene**". Offshore Technology Conference, Houston, Texas, May 2000.
- FACELI, K., LORENA, A. C., GAMA e J., CARVALHO, C. P. L. F. **Inteligência artificial: uma abordagem de aprendizado de máquina**. Rio de Janeiro: LTC, 2011.
- FILHO, R. F. **Agenda de Apoio as Operações de Completação**, 2013.
- GARCIA, J. E. L. **A Completação de Poços no Mar**. Salvador, BA: 1997.
- HAUSLER, R. H., KRISHNAMURTHY, R. M., and BRENT W.A. S. "**Observation of Productivity Loss in Large Oil Wells due to Scale Formation without Apparent Production of Formation Brine**." CORROSION 2015, Dallas, Texas, March 2015.
- HIRSCH, R. L.; BEZDEK, R; WEDLING, R. **Peaking of World Oil Production: Impacts, Mitigation, Risk Management**. ASPO, maio de 2005. Disponível em: <http://large.stanford.edu/courses/2014/ph240/liegl1/docs/hirschreport.pdf>. Acesso em: 03/05/2023.
- JÚNIOR, W. F. VARGAS, R. E. V, KOMATI, K. S, GAZOLLI, K. A. S. **Deteção de anomalias em poços produtores de petróleo usando aprendizado de máquina**. Congresso Brasileiro de Automática-CBA. 2020.
- KOWSARI, K. MAIMANDI, K. J. HEIDARYSAFA, M. MENDU, S. BARNES, L. E. BROWN, D. E. **Text classification algorithms: a survey**. Information, Multidisciplinary Digital Publishing Institute, v. 10, n. 4, p. 150, 2019
- MEGLIO, F. PETIT, N. ALSTAD, V. KAASA, G. **Stabilization of slugging in oil production facilities with or without upstream pressure sensors**. Journal of Process Control, v. 22, n. 4, p. 809–822, 2012.
- MOORE. W. R.; KONOPCZYNSKI. M. R; **Implementation of Intelligent Well Completions Within a Sand Control Environment**. IADC/SPE 77202, Jakarta, 2002.
- NASCIMENTO, R. S. F. do. **Deteção de anomalias em poços de produção de petróleo offshore com a utilização de autoencoders e técnicas de reconhecimento de padrões**. 2021. 88 p. Dissertação (Mestrado em Engenharia de Sistemas e Automação) – Universidade Federal de Lavras, Lavras, 2022.
- SANTOS, Mayara de Jesus Rocha. **Deteção de problemas de garantia de escoamento a partir da utilização de ferramentas de aprendizado de máquina**.

2020. 110 f. Dissertação (Mestrado) - Curso de Programa Francisco Eduardo Mourão Saboya de Pós-Graduação em Engenharia Mecânica, Engenharia de Mecânica, Universidade Federal Fluminense, Niterói, 2020.

SCHMIDT, Z.; DOTY, D.; DUTTA-ROY, K. **Severe Slugging in Offshore Pipeline Riser-Pipe Systems**. Society of Petroleum Engineers Journal, v. 25, n. 1, p. 27-38, fev. 1985.

THOMAS, et al. **Fundamentos de Engenharia de Petróleo**. Rio de Janeiro: Interciência, 2001.

VARGAS, R. E. V. **Base de Dados e Benchmarks para Prognóstico de Anomalias em Sistemas de Elevação de Petróleo**. Tese (Doutorado) — Universidade Federal do Espírito Santo, Vitória - ES, 8 2019

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.