



Universidade Federal de Campina Grande
Centro de Engenharia Elétrica e Informática
Curso de Graduação em Engenharia Elétrica

JOSÉ GOMES DA SILVA NETO

TRABALHO DE CONCLUSÃO DE CURSO

**CIÊNCIA DE DADOS APLICADA AO DESENVOLVIMENTO DE MODELO
DE PREDIÇÃO DE FALHAS EM TURBOGERADORES**

Campina Grande, Paraíba.

Agosto de 2022

Ciência de Dados Aplicada ao Desenvolvimento de Modelo de Predição de Falhas em Turbogeneradores

Trabalho de Conclusão de Curso submetido ao curso de Bacharelado em Engenharia Elétrica na Universidade Federal de Campina Grande, Campus Campina Grande, como parte da exigência para obtenção do grau de bacharel em Engenharia Elétrica.

Área de concentração: Sistemas Elétricos de Potência

Orientador: Dr. George Rossany Soares Lira.

Campina Grande, Paraíba.

Agosto de 2022

Ciência de Dados Aplicada ao Desenvolvimento de Modelo de Predição de Falhas em Turbogeneradores

Como parte dos requisitos necessários para obtenção do grau de Bacharel em Ciências no Domínio da Engenharia Elétrica, segue a Proposta de Trabalho de Conclusão de Curso de Bacharelado, em Engenharia Elétrica, submetida à Coordenação de Graduação, do mesmo curso, da Universidade Federal de Campina Grande.

Aprovado em: ____ / ____ / _____

Prof. Dr. George Rossany Soares Lira.
Orientador

Prof. Dr. Pablo Bezerra Vilar
Avaliador

RESUMO

Este trabalho apresenta uma nova abordagem para a análise de faltas em turbogeradores. O sistema proposto tem por finalidade antecipar a visualização de uma eventual intercorrência no processo de operação, a fim de proporcionar ampliação à longevidade do equipamento. A partir de informações de falhas no sistema, associadas às leituras dos sensores locais, foi implementado um algoritmo de informação mútua a fim de correlacionar as duas informações. Para o banco de dados base foi utilizada a ferramenta matriz ampliada, fornecendo diferentes perspectivas, para um mesmo objeto de estudo. Por meio da matriz ampliada, torna-se possível, para o sistema, confrontar resultados oriundos de uma mesma base de dados, indicando quais leituras apresentam maior confiabilidade para a análise. Essa ferramenta foi de imprescindível importância para o cálculo de informação mútua, realizado no código exposto. Ao longo do trabalho, o código sofreu diversas otimizações, que foram resumidas em duas macroanálises, a primeira evidenciando o método de concatenação utilizado e a segunda elucidando o método de cálculo da informação mútua. Por fim, um tratamento de dados foi implementado, possibilitando uma redução na base de dados em 86% e uma redução no tempo de execução em 77%.

Palavras-chave: Informação mútua, matriz ampliada, análise de faltas, confiabilidade, tratamento de dados.

ABSTRACT

This work presents a new approach to fault analysis in turbogenerators. The purpose of the proposed system is to anticipate the visualization of a possible intercurrent in the operation process, in order to increase the longevity of the equipment. Based on information about system failures, associated with readings from local sensors, a mutual information algorithm was implemented in order to correlate the two pieces of information. For the base database, the extended matrix tool was used, providing different perspectives for the same object of study. Through the expanded matrix, it becomes possible for the system to compare results from the same database, indicating which readings present greater reliability for the analysis. This tool was essential for the calculation of mutual information, performed in the exposed code. Throughout the work, the code underwent several optimizations, which were summarized in two macroanalyses, the first evidencing the concatenation method used and the second elucidating the method of calculating the mutual information. Finally, a data treatment was implemented, allowing a reduction in the database by 86% and a reduction in execution time by 77%.

Keywords: Mutual information, extended matrix, fault analysis, reliability, data treatment.

LISTA DE ABREVIATURAS E SIGLAS

ML – Machine Learning

DL – Deep Learning

FPSO – Floating Production, Storage and Offloading

SIN – Sistema Interligado Nacional

UEP – Unidade Estacionária de Produção

BPD – Barris Por Dia

TAG – Sigla inglesa para etiqueta ou rótulo

DF – DataFrame

LISTA DE FIGURAS

Figura 1: Diagrama esquemático de uma plataforma FPSO

Figura 2: Fotografia de plataforma FPSO

Figura 3: Diagrama unifilar da plataforma analisada

Figura 4: Fluxograma do trabalho

Figura 5: Base de dados original

Figura 6: Exemplo de formação da primeira matriz ceifada

Figura 7: Exemplo de formação da segunda matriz ceifada

Figura 8: Exemplo de formação da terceira matriz ceifada

Figura 9: Exemplo de matriz ampliada

Figura 10: Matriz ampliada

Figura 11: Vetores de informação mútua, antes e após a normalização

Figura 12: Vetores df , antes e após a filtragem

Figura 13: Vetor de falhas original

Figura 14: Vetor de falhas binário

LISTA DE TABELAS

Tabela 1: Tempos de execução de concatenação

Tabela 2: Tempos de execução do cálculo de informação mútua

SUMÁRIO

Sumário	10
1 Introdução	10
1.1 Justificativa	11
1.2 Hipótese.....	12
1.3 Objetivo Geral.....	12
1.4 Objetivos Específicos.....	13
2 Fundamentação Teórica	13
2.1 Plataforma Tipo FPSO	15
2.2 Sistema Elétrico	17
3 Metodologia	18
3.1 Base de Dados	18
3.2 Matriz Ampliada	19
3.3 Cálculo da Informação Mútua.....	22
3.4 Primeira Otimização.....	24
3.5 Segunda Otimização.....	25
4 Resultados e discussões.....	27
5 Conclusão	29
6 Referências.....	31
Apêndice 1 – Recorrência dos resultados – método de regressão	33
Apêndice 2 – Recorrência dos resultados – método de classificação.....	36

1 INTRODUÇÃO

Este trabalho aborda uma forma de obter informações por inspeção de um banco de dados, podendo fornecer limpeza, seleção e orientação guiada de acordo com o objetivo proposto. Este conjunto de medidas pode ser resumida por análise de dados.

Para as análises realizadas ao longo do exposto foi utilizado, como fonte de dados, um conjunto de leituras igualmente espaçadas no domínio do tempo. Este conjunto de leituras evidencia o sensoriamento de um turbogerador localizado em uma plataforma de petróleo e gás, do tipo FPSO (unidade flutuante de produção, armazenamento e transferência de petróleo e gás).

Turbogerador é um equipamento transdutor que converte energia cinética de moléculas de gás ou vapor em energia elétrica. Sob uso de um tubo direcionador para o encaminhamento dessas moléculas, uma turbina é posicionada para captar o seu movimento, da melhor forma, e convertê-lo sob uso de uma máquina elétrica girante, acoplada ao eixo dessa turbina. Vale salientar que esta tecnologia de geração de energia elétrica é a opção mais utilizada em navios e plataformas de petróleo, como mostra o exposto a seguir:

Segundo Fernando Martins, vice-presidente da GE para Óleo e Gás na América Latina, as usinas hidrelétricas são as principais fornecedoras de energia elétrica para as cidades brasileiras, no caso das plataformas, sistemas específicos atendem às demandas energéticas. Ainda de acordo com ele, a alternativa mais comum às “hidrelétricas” das plataformas são as turbinas a gás, principalmente os modelos aeroderivados que, como o nome denuncia, são adaptações dos equipamentos tradicionais que propulsionam os aviões a jato (PORTAL PETRÓLEO E ENERGIA).

O trabalho foi inteiramente desenvolvido em formato computacional, sob uso das plataformas *spyder* e *google colab*, como ferramentas de desenvolvimento dos algoritmos que foram elaborados sob uso da linguagem de programação *python*. Ambas as plataformas são de código livre, sendo a primeira de uso em máquina local e a segunda em nuvem.

Vale salientar que o estudo oferece análise de dados para um banco de dados específico composto por séries temporais. No entanto, podendo ser expandido para os

mais diversos tipos de conjuntos de dados que possuam como referência um alvo ao qual os dados possam ser relacionados.

Os métodos numéricos para obtenção de informação são utilizados como módulos pré-estabelecidos e consagrados nas análises numéricas dentro da linguagem de programação utilizada, provenientes da biblioteca *scikit-learn*.

1.1 JUSTIFICATIVA

Analisando o cenário atual, em que a demanda por produtos e serviços precisa se alinhar às mudanças de forma rápida, a presença de falhas se torna um fator decisivo à produção. Seguindo essa linha de raciocínio, o desenvolvimento de modelos de predição de falhas vem a se tornar uma ferramenta indispensável para a otimização dos processos industriais. No caso das plataformas de petróleo essa situação se torna ainda mais agravante, pois a fonte única de fornecimento elétrico é proveniente do sistema de turbogeração.

Partindo do pressuposto de que a manutenção para a prevenção é menos custosa, em comparação com a manutenção para a correção, aliada a um modelo em que o sistema não sofre paradas não programadas, o setor de manutenção, que tinha uma imagem de gerador de custos para a empresa, passou a ter papel importante na competitividade. A competitividade exige mudanças rápidas, adaptação às exigências do mercado, redução de custos e alto desempenho do processo produtivo (CAMPOS, 2013). Desta forma, a manutenção preditiva vem a ser um fator que associa a redução de custos ao acréscimo da confiabilidade do sistema elétrico, por redução da quantidade de paradas não programadas, na plataforma.

O armazenamento de dados, oriundos do sensoriamento de um equipamento, é um método que há tempos é utilizado para a elaboração de um histórico de operação. Tais dados revelam a forma como o dispositivo foi operado, os comportamentos operacionais cíclicos e a ocorrência de faltas. Com o avanço tecnológico exponencial dos dispositivos de memória, e os adventos dos componentes embarcados, a tarefa de armazenar dados encontra-se cada vez mais simples e segura.

1.2 HIPÓTESE

A forma de armazenamento de dados se modificou e atualmente, a coleta de dados discreta, com uma taxa de amostragem pré-determinada, é um método utilizado amplamente. De acordo com Santos et al. (2022) um fluxo de dados é uma sequência ordenada de amostras, que são obtidas continuamente ao longo tempo como uma série temporal, e a ordem de suas amostras é uma característica crucial do problema.

Desta maneira, os métodos de análise de dados também sofreram adequações e inovações, podendo oferecer novas informações a partir desse novo formato de banco de dados. Segundo Gondra (2008) a não existência de técnicas universais de identificação de falhas torna a utilização de métodos especializados em identificação de padrões, como as técnicas de *Machine Learning* (ML), extremamente promissora.

Em concordância com Pinto (2021), *Machine Learning*, ou em português, Aprendizado de máquina é uma subárea de estudos dentro de Inteligência Artificial (IA). IA é uma tecnologia que tem a capacidade de criar sistemas que podem simular a inteligência humana, a parte de ML é a que se dedica a encontrar padrões, aprender com os dados e prever resultados futuros.

Conforme Lei et al. (2013) máquinas rotativas são comumente usadas em sistemas mecânicos e desempenham um papel importante em aplicações industriais. Em sintonia com Sun et al. (2022) segurança e confiabilidade são fatores importantes nas operações industriais. Na maquinaria industrial moderna, as máquinas rotativas representaram mais de 80% da estrutura instalada. Também em concordância com Brito et al. (2022) os modelos tradicionais de ML não devem ser abandonados apesar dos recentes avanços da *Deep Learning* (DL): isso porque ainda vale a pena investigar o aprendizado estatístico em Diagnóstico Inteligente de Falhas com a revolução do *big data*, uma vez que as teorias de aprendizado estatístico têm bases teóricas rigorosas, que promovem a construção de modelos diagnósticos com parâmetros, características e resultados de fácil compreensão.

1.3 OBJETIVO GERAL

Paralelamente ao alinhamento às mudanças de forma rápida, imposta pela demanda por produtos e serviços, o desenvolvimento computacional também se encontra em velocidade elevada. Desta forma, o número de operações matemáticas executadas por

segundo sofre acréscimos a passos largos. Esse fato possibilita uma análise de dados cada vez mais rápida e robusta. Portanto, modelos de análise de dados, sob uso de ferramentas computacionais, podem vir a ser grandes aliados na obtenção de informações importantes. Tomando como referência Liu et al. (2022) métodos orientados a dados, como aprendizado de máquina, podem ser usados para aprender padrões de falha de máquinas a partir de dados de monitoramento de condição.

Partindo deste princípio, este trabalho propõe realizar uma análise de dados, obtidos a partir do sensoriamento de turbogeradores utilizados em uma plataforma de produção de petróleo e gás, com o objetivo de desenvolver um método para relacionar as leituras às falhas. Partindo da implementação de um modelo baseado em ML para a análise e tratamento de um banco de dados, proveniente do sensoriamento de uma turbina a gás, onde as amostras foram coletadas minuto a minuto, possibilitando um diagnóstico precoce.

A proposta surgiu baseada em Castro (2020), que utilizou o software *PowerFactory* para analisar a confiabilidade do sistema elétrico de uma plataforma do tipo FPSO (*Floating Production Storage and Offloading*).

1.4 OBJETIVOS ESPECÍFICOS

- Analisar um banco de dados proveniente de leituras a partir de sensores em um turbogerador;
- Desenvolver um método para correlacionar as informações de falhas às leituras;
- Otimizar o algoritmo;

2 FUNDAMENTAÇÃO TEÓRICA

Em linha com Ma et al. (2021), técnicas eficazes de diagnóstico de falhas garantem a confiabilidade e segurança das máquinas industriais, evitando perdas de vidas e propriedades causadas por falhas. Além disso, em concordância com Li et al. (2019), as turbinas a gás são o núcleo da geração de energia e são construídas em várias formas,

como motores aeronáuticos, turbinas a gás marinhas, turbinas a gás industriais, microturbinas a gás e turbinas a gás para serviços pesados.

Para este tipo de plataforma, as falhas se tornam ainda mais críticas, pois devido a características intrínsecas à plataforma, as pausas acarretam em perdas significativas na produção, afetando diretamente a relação entre o tempo de operação da plataforma e sua vida útil.

Em congruência com Tran et al. (2021) o desenvolvimento do método de detecção de vibração é essencialmente importante para a automação e para evitar erros humanos durante o processo de usinagem. Ainda mais, em consonância com Mousavi et al. (2022), a estrutura complicada e os comportamentos dependentes da carga das turbinas a gás provocam algumas dificuldades na identificação das causas-raiz da deterioração do desempenho, levando a falhas inesperadas em tais sistemas.

Neste contexto, as indústrias buscam práticas eficientes na manutenção, com o objetivo de melhorar a disponibilidade da produção. Uma técnica proposta na literatura e que está sendo, aos poucos, implementada na indústria é utilizar inteligência artificial para estimar as performances dos equipamentos, prever a necessidade de manutenção, possíveis modos de falhas e aspectos semelhantes com antecedência (BUKSHSH et al, 2019). Conforme detalhado por Liu et al. (2018), existem três tarefas básicas de diagnóstico de falhas: (1) determinar se o equipamento está normal ou não; (2) encontrar uma causa incipiente de falha e sua razão; (3) prever a tendência de desenvolvimento de falhas.

De acordo com Benouioua et al. (2022), a estratégia de diagnóstico é desenvolvida levando em conta os seguintes requisitos:

- Uma instrumentação limitada: a instrumentação não intrusiva confere um aspecto “genérico” ao sistema de diagnóstico, tornando-o mais facilmente adaptável.
- Um método simples, mais facilmente implementável numa perspectiva de diagnóstico em tempo real com vista à aplicação. Assinaturas de falhas com tempos de computação reduzidos, combinadas com uma abordagem de diagnóstico orientada por dados, são usadas propositalmente. As ferramentas matemáticas implementadas devem resultar em um algoritmo de diagnóstico de alto desempenho, tanto em termos de taxa de classificação (identificação) dos modos de operação considerados, quanto de “custo/tempo de cálculo”.

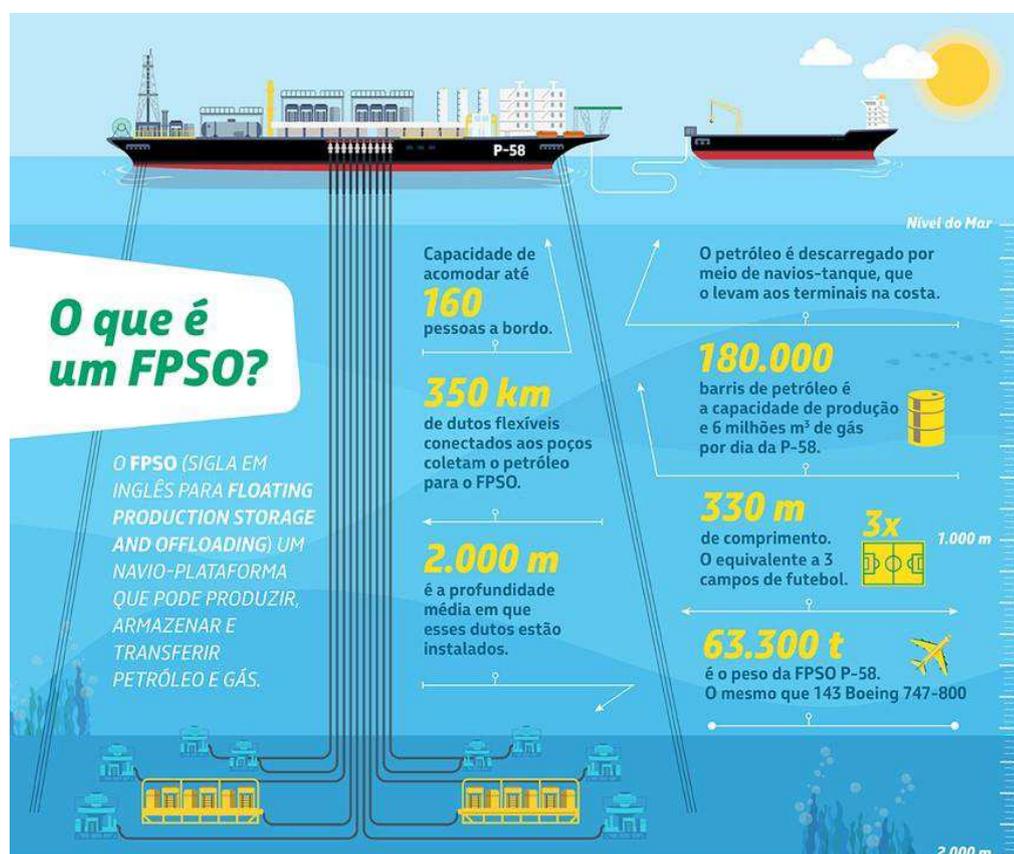
- A prevenção de falhas, ou degradações irreversíveis, com bastante antecedência.

Dessa forma, faz-se necessária a presença de um sistema de análise de falhas robusto, e com entrega de resultados em tempo hábil, promovendo assim a solução para eventuais faltas que poderiam vir a ocorrer no sistema elétrico da plataforma. De acordo com Ma et al. (2021), técnicas eficazes de diagnóstico de falhas garantem a confiabilidade e segurança das máquinas industriais, evitando perdas de vidas e propriedades causadas por falhas.

2.1 PLATAFORMA TIPO FPSO

Este trabalho propõe realizar uma análise de dados, obtidos a partir do sensoriamento de turbogeradores, utilizados em uma plataforma de produção de petróleo e gás. A figura 1, a seguir, resume algumas informações importantes a respeito deste tipo de plataforma.

Figura 1: Diagrama esquemático de uma plataforma FPSO



Fonte: Petrobras

A plataforma, do tipo FPSO (unidade flutuante de produção, armazenamento e transferência de petróleo e gás), tem capacidade de processar até 180 mil barris de óleo e 12 milhões de m³ de gás por dia, o que representa 6% da produção operada pela Petrobras, contribuindo para o crescimento previsto da produção da companhia (PETRÓLEO BRASILEIRO S.A. 2022).

Segundo Castro (2020), a especificidade do estudo que trata de um sistema típico de um FPSO se deu pela necessidade de melhor explorar esse assunto dentro de um universo particular que conta com grandes volumes de energia e que trabalha de forma isolada de outros sistemas elétricos.

Além dos fatores geográficos, evidenciando a indisponibilidade de conexão ao SIN (Sistema Interligado Nacional), a presença do fator econômico também é um forte influenciador ao desenvolvimento de um sistema preditivo eficiente. A operação de uma UEP tipo FPSO, que produz mais de 150.000 barris por dia (bpd), envolve custos operacionais diários superiores a US\$ 550.000,00 (Brasil Energy Insight, 2019), com lucro cessante diário em torno de US\$ 6.000.000,00 para um custo de barril de petróleo da ordem de US\$ 40,00 (Offshore Engineer, 2020).

Ao passo que a figura 1, anteriormente apresentada, mostrou um diagrama esquemático da plataforma aqui abordada, a figura 2, abaixo, exhibe uma fotografia real deste tipo de plataforma.

Figura 2: Fotografia de plataforma FPSO

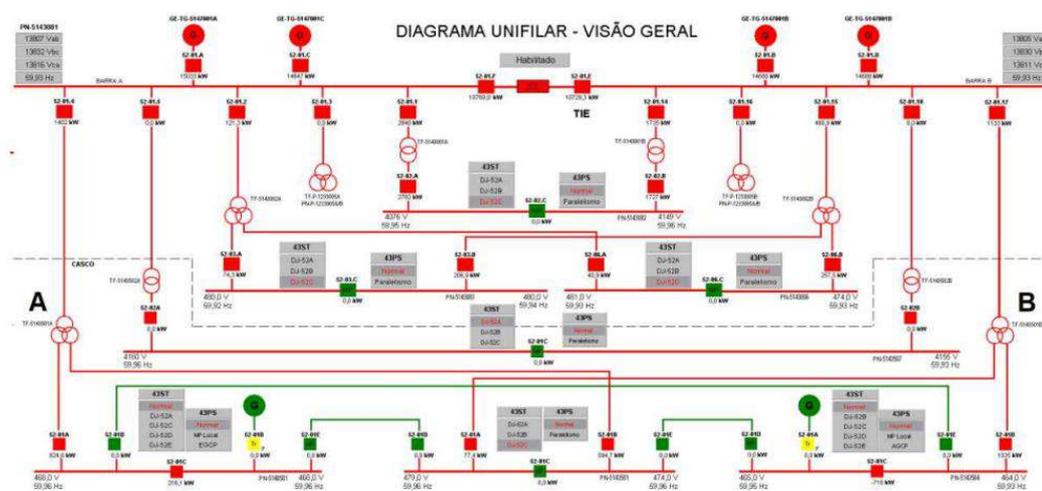


Fonte: Petrobras

2.2 SISTEMA ELÉTRICO

De acordo com Castro (2020), o sistema de geração principal dessa unidade conta com quatro turbogeradores (GE-TG-5147001A/B/C/D), acionados por turbinas a gás, com potência nominal de 25 MW / 31,25 kVA, tensão nominal de 13,8 kV e frequência nominal de 60 Hz que alimentam tanto o sistema elétrico do *topside* (parte superior de uma plataforma de petróleo que compreende a superfície; contém todos os equipamentos de perfuração, produção e processamento) quanto o casco da plataforma. O sistema elétrico do casco conta com um gerador auxiliar (GE-UG5262501) e um gerador de emergência (GE-UG-5261501) com potência nominal de 1,8 MW cada e tensão nominal de 480 V. Esses geradores entram em operação em caso de falha de todo o sistema de geração principal, nos testes semanais ou em caso de alguma necessidade operacional. Na condição normal de operação do sistema elétrico apenas os geradores principais necessários para atendimento à demanda da plataforma são utilizados. Em relação ao sistema de distribuição de energia elétrica, os níveis de tensão encontrados na plataforma são: 13,8 kV, 4,16 kV, 690 V, 480 V, 220 V e 110 V. No sistema elétrico do *topside*, os níveis de tensão são 13,8 kV, 4,16 kV, 690 V, 480 V, 220 V e 110 V, já no sistema elétrico do casco os níveis de tensão encontrados são 13,8 kV, 4,16 kV, 480 V, 220 V e 110 V. Abaixo, a figura 3 traz um diagrama esquemático do sistema elétrico da plataforma.

Figura 3: Diagrama unifilar da plataforma analisada

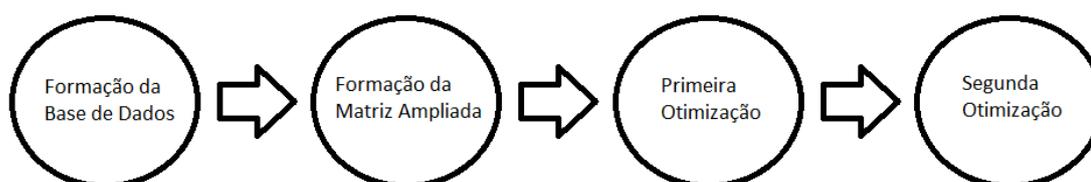


Fonte: Castro, 2020.

3 METODOLOGIA

Neste capítulo serão introduzidos os processos utilizados para a realização da análise de dados. O fluxograma indicado na Figura 4 resume as principais etapas que serão aprofundadas.

Figura 4: Fluxograma do trabalho



Fonte: autoria própria

3.1 BASE DE DADOS

A base de dados aqui analisada consta de leituras sensoriais realizadas minuto a minuto por um período de seis dias. Cada um dos sensores recebe um título que referencia sua descrição de funcionalidade e posição, denominado TAG. Cada sensor tem suas leituras armazenadas formando um vetor coluna, onde cada linha é uma leitura e a coluna tem por título a sua TAG. A concatenação dos sensores forma uma base dados de todas as leituras realizadas. Associado a esta base, uma coluna que informa a ocorrência de falhas é adicionada lateralmente, possuindo título *Flag_Falha*. Este conjunto de dados, é finalmente apresentado como uma matriz que dispõe de 8761 linhas e 56 colunas, que pode ser visualizado, parcialmente, na figura 4 a seguir.

Figura 5: Base de dados original

	Flag_Falha	P74_BE- 5147501A- A	P74_BE- 5147501A- B	P74_IT- 5147501A	P74_PDIT- 5147520A	P74_PDIT- 5147522A- A	P74_PDIT- 5147522A- B	P74_PIT- 5147501A	P74_PIT- 5147530A- A	P74_PIT- 5147530A- B	...
0	0.0	0.000000	0.000000	0.259979	0.000939	-0.000050	0.001054	-0.000053	-0.000186	0.000039	...
1	0.0	0.000000	0.000000	0.258057	0.000916	0.000052	0.001288	-0.000056	-0.000198	0.000032	...
2	0.0	0.000000	0.000000	0.260939	0.000950	-0.000046	0.000997	-0.000056	-0.000186	0.000038	...
3	0.0	0.000575	0.000575	0.033965	0.001077	-0.000435	-0.000434	0.023230	0.002672	0.002702	...
4	0.0	0.000419	0.000419	0.025137	0.002248	0.001125	0.000693	0.016815	0.002369	0.002390	...
...
8756	0.0	0.000000	0.000000	0.030971	0.000100	0.000095	-0.000363	-0.000004	-0.000021	0.000017	...
8757	0.0	0.000000	0.000000	0.031049	0.000103	0.000112	-0.000370	-0.000003	-0.000021	0.000017	...
8758	0.0	0.000000	0.000000	0.030805	0.000098	0.000102	-0.000363	-0.000004	-0.000021	0.000017	...
8759	0.0	0.000000	0.000000	0.031002	0.000100	0.000099	-0.000363	-0.000006	-0.000021	0.000017	...
8760	0.0	0.000000	0.000000	0.030544	0.000097	0.000099	-0.000354	-0.000006	-0.000021	0.000017	...

8761 rows x 56 columns

Fonte: autoria própria.

3.2 MATRIZ AMPLIADA

Devido a limitação da base de dados, pois conta com um período de tempo delimitado, uma técnica para a ampliação da visualização das informações foi implementada.

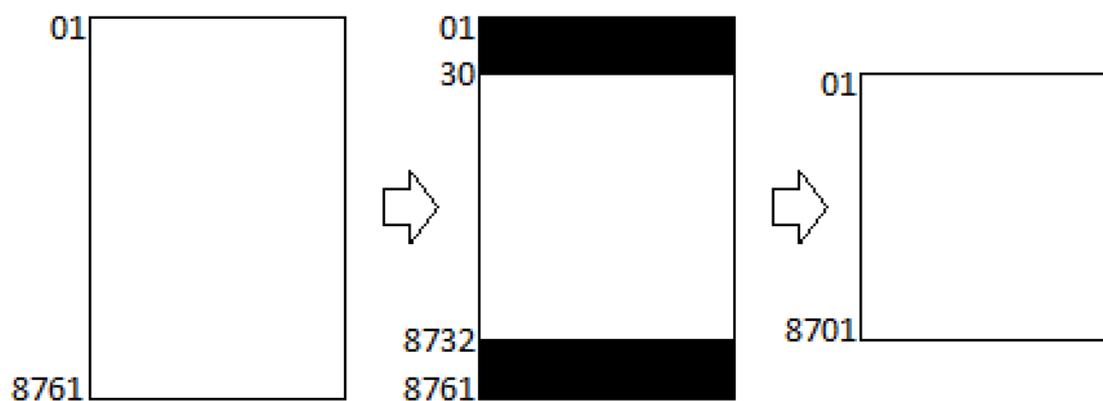
Dado o arquivo original de dados a informação de falha é a referência a ser tomada para que se inicie a análise de dados. Desta forma, o vetor *Flag_Falha* foi retirado e a operação ocorre com as duas bases de dados, vetor de falhas (8761, 1) e leituras dos sensores (8761, 55), de forma separada.

Fazendo uso dessas bases de dados uma seleção central foi implementada, retirando dados das extremidades inicial e final destas matrizes, reduzindo-as. Este processo ocorreu de forma sequencial, retirando linhas do início e fim da matriz original, e armazenando o conteúdo selecionado em uma variável inicialmente vazia. Para cada seleção o número total de linhas retirada sempre foi o mesmo, variando apenas as quantidades inicial e final. Este processo recursivo foi operacionalizado sob uso de um laço controlado do tipo *for*.

Sob uso de um método de redução controlada, há a possibilidade de uma redução da base de dados original, conseqüentemente, proporcionando uma visão diferente, pois agora tem-se uma nova série de dados observada, apesar de formada a partir de uma base

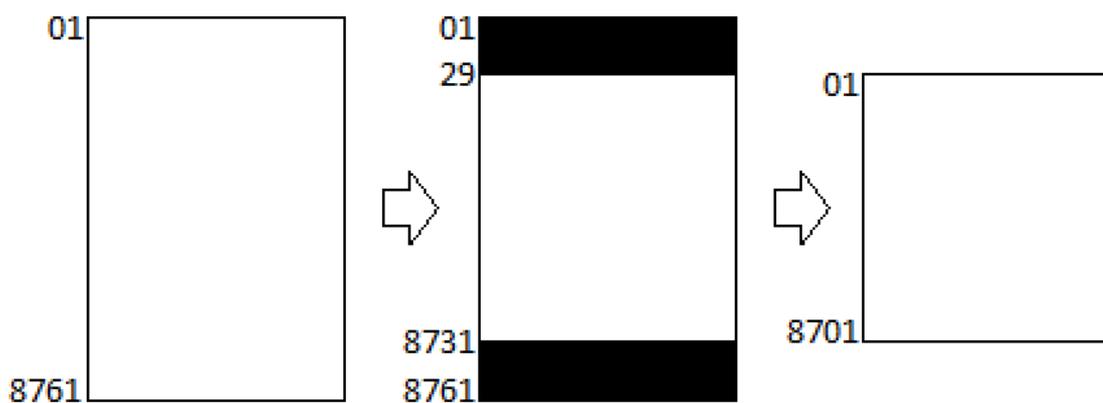
de dados previamente conhecida. O processo, anteriormente abordado, é evidenciado nas figuras 5, 6 e 7 a seguir.

Figura 6: Exemplo de formação da primeira matriz ceifada



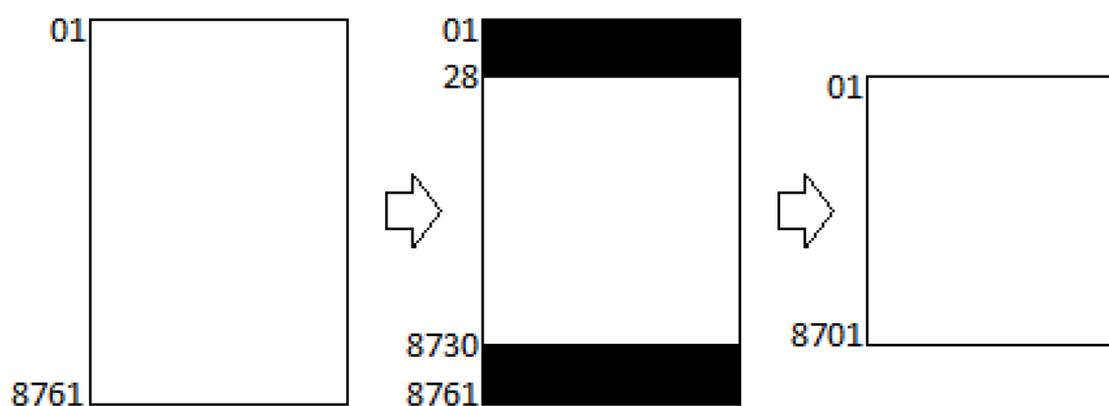
Fonte: autoria própria

Figura 7: Exemplo de formação da segunda matriz ceifada



Fonte: autoria própria

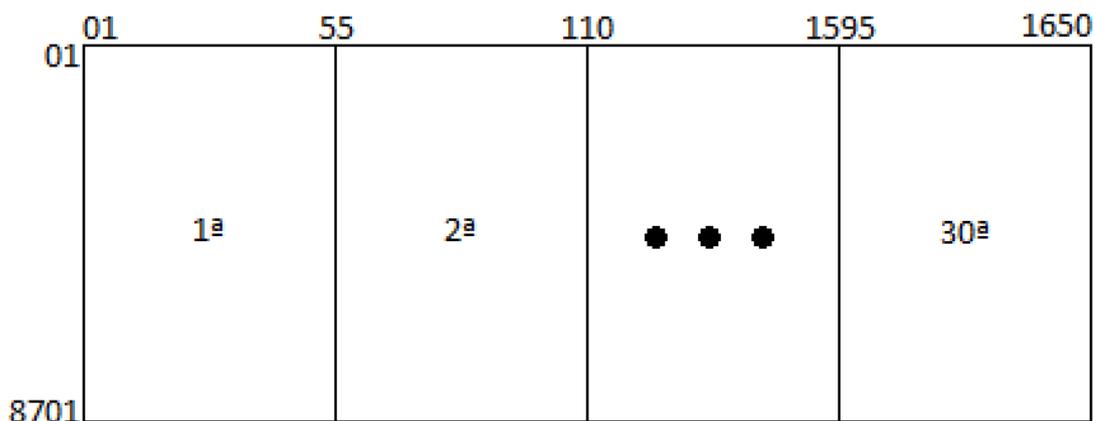
Figura 8: Exemplo de formação da terceira matriz ceifada



Fonte: autoria própria

O código aqui implementado apresentou uma redução de 60 linhas, retirando inicialmente as 30 leituras iniciais e finais, posteriormente as 29 iniciais e 31 finais, e assim por diante.

Figura 9: Exemplo de matriz ampliada



Fonte: autoria própria

A disposição final é composta por uma matriz com 30 matrizes de informação reduzida, concatenadas lateralmente, formando um arquivo com 1650 colunas e 8701 linhas. Esta base de dados conjunta é denominada matriz ampliada.

A formação da matriz ampliada traz à tona diferentes perspectivas para um mesmo conjunto de dados. Sob uso de um algoritmo que forme um perfil para cada conjunto de leituras associadas ao mesmo sensor, pode-se analisar os resultados, a partir de um confronto entre as leituras.

Por mais que a base de dados seja a mesma, esta ferramenta possibilitará uma forma de validação do método de análise utilizado, pois, agora existem diferentes intervalos de tempo, levando a disposições de leituras distintas, para cada um dos sensores. Assim, possibilitando a comparação entre os valores de informação mútua, para um mesmo sensor.

Abaixo segue a imagem 9, evidenciando o acréscimo do número de colunas em comparação com a imagem 4, da base de dados original.

Figura 10: Matriz ampliada

	0	1	2	3	4	5	6	7	8	9	...
0	0.000412	0.000412	0.025208	0.002350	0.002890	0.001323	0.016493	0.002315	0.002337	0.016709	...
1	0.000412	0.000412	0.024614	0.002390	0.002974	0.001464	0.016482	0.002306	0.002329	0.016685	...
2	0.000415	0.000415	0.025021	0.002413	0.002618	0.000992	0.016610	0.002332	0.002353	0.016829	...
3	0.000416	0.000416	0.025008	0.002598	0.002644	0.000634	0.016658	0.002338	0.002361	0.016876	...
4	0.000416	0.000416	0.025899	0.002289	0.001115	0.000604	0.016658	0.002325	0.002346	0.016880	...
...
8696	0.000000	0.000000	0.030304	0.000094	0.000095	-0.000367	-0.000006	-0.000023	0.000015	0.000468	...
8697	0.000000	0.000000	0.030293	0.000094	0.000092	-0.000363	-0.000006	-0.000023	0.000015	0.000473	...
8698	0.000000	0.000000	0.030450	0.000096	0.000094	-0.000359	-0.000005	-0.000023	0.000015	0.000467	...
8699	0.000000	0.000000	0.030100	0.000091	0.000091	-0.000353	-0.000005	-0.000023	0.000015	0.000474	...
8700	0.000000	0.000000	0.029957	0.000088	0.001519	-0.000360	0.019254	0.000006	0.000046	0.000480	...

8701 rows x 1650 columns

Fonte: autoria própria

3.3 CÁLCULO DA INFORMAÇÃO MÚTUA

De acordo com Li (1990), a informação mútua é uma medida da dependência entre duas variáveis. Se as duas variáveis são independentes, a informação mútua entre elas é zero. Se as duas são fortemente dependentes, uma é função da outra, e a informação mútua entre elas é grande. Existem outras interpretações da informação mútua, por exemplo, a informação armazenada em uma variável sobre outra variável e o grau de previsibilidade da segunda variável conhecendo a primeira. Claramente, todas essas interpretações estão relacionadas à mesma noção de dependência e correlação.

Desta forma, após a análise previamente realizada, foi implementada uma análise de informação mútua, à matriz ampliada, sob uso de um algoritmo de regressão linear, que identificou a relação entre os sensores e o vetor de falhas. O algoritmo abordado trata da função *mutual_info_regression*, obtida com o auxílio da biblioteca *scikit-learn* e mencionada pela sigla *sklearn*. Essa ferramenta é comumente utilizada para a prática do *machine learning* na linguagem de programação *Python*.

Dada a função, sua operação segue-se segundo o método de regressão linear, correlacionando os vetores de entrada com um vetor alvo, inicialmente determinado, e nomeado *target_vector*. Este método, de regressão linear, formula o cálculo de informação mútua entre cada um dos vetores da matriz ampliada e o vetor *Flag_Falha*, resultando em um valor numérico. Os resultados são armazenados em um vetor que trazia

em cada argumento um valor associado ao respectivo sensor. Após o armazenamento os valores são normalizados, tomando como referência o valor mais elevado dentre os resultados. Seguem abaixo, na figura 10, os dois vetores de informação mútua, anterior e posterior à normalização.

Figura 11: Vetores de informação mútua, antes e após a normalização

The figure shows two screenshots of a Jupyter Notebook. The top screenshot displays a pandas DataFrame with 11 columns (0-10) and 1 row. The values are: 0.0, 0.002876, 0.0, 0.000712, 0.000681, 0.00016, 0.0002, 0.000365, 0.000638, 0.000271, 0.001084. The bottom screenshot displays a pandas DataFrame with the same 11 columns and 1 row, but with normalized values: 0.0, 0.339387, 0.0, 0.084053, 0.080315, 0.018838, 0.023632, 0.043103, 0.098853, 0.032036, 0.127894.

Fonte: autoria própria

Finalizando o código foi implementado ao vetor *mutual_information* o título de cada valor, relacionando-o ao seu respectivo sensor e número de matriz. Posteriormente foi implementada uma filtragem, selecionando apenas os sensores que possuem valor de informação mútua superior a 5% do valor normalizado. Este valor é um parâmetro de filtragem, podendo ser alterado, conforme as análises sejam mais refinadas. Criou-se a variável *df* que armazena os dados de informação mútua após a filtragem. Segue abaixo a figura 11, que apresenta o vetor *df* inicialmente copiado do vetor *mutual_information* e posterior ao processo de filtragem.

Figura 12: Vetores *df*, antes e após a filtragem

The figure shows two screenshots of a Jupyter Notebook. The top screenshot shows a pandas DataFrame with 10 columns labeled with sensor IDs and matrix numbers (e.g., P74_BE-5147501A-A 0, P74_BE-5147501A-B 1, etc.). The values are: 0.0, 0.339387, 0.0, 0.084053, 0.080315, 0.018838, 0.023632, 0.043103, 0.098853, 0.032036. The bottom screenshot shows a pandas DataFrame with 10 columns labeled with sensor IDs and matrix numbers (e.g., P74_BE-5147501A-B 1, P74_PDIT-5147520A 3, etc.). The values are: 0.339387, 0.084053, 0.080315, 0.098853, 0.127894, 0.058592, 0.050514, 0.060347, 0.174113, 0.123139.

Fonte: autoria própria

Foi identificado que alguns sensores apresentaram valores significativamente inferiores, demonstrando uma baixa correlação ao vetor de falhas. A partir desta perspectiva, foi proposta uma redução da base de dados, reduzindo os custos computacionais. Após a construção desta base de dados reduzida, obteve-se uma matriz ampliada com 218 colunas, em comparação à originalmente elaborada com 1650 colunas, desta forma podendo reduzir em até 86% o custo computacional.

3.4 PRIMEIRA OTIMIZAÇÃO

Foi identificado que o laço iterativo que elabora a matriz ampliada faz uso de um grande número de operações repetidas, devido a elevada quantidade de dados presentes na base original. Também foi visto que o código possui dois pontos de maior relevância para o tempo de execução:

- Implementação da matriz ampliada;
- Execução do cálculo de informação mútua;

Apenas o primeiro gargalo é afetado pela mudança implementada, que diz respeito à elaboração da matriz ampliada. Portanto, o código foi dividido nestas duas partes, agora tendo marcações de tempo individuais para cada etapa, e, conseqüentemente, duas marcações de tempo para cada execução completa.

Após uma quantificação do total de operações foi visto que, para implementação da matriz ampliada, a função *append*, proveniente da biblioteca *numpy*, responsável por concatenar os dados, para elaboração da base de dados, era utilizada 280.352 vezes. Por este motivo, foi proposta a substituição desta função por uma de custo operacional reduzido. O teste foi efetuado utilizando a função *concat*, oriunda da biblioteca *pandas*.

O código teve sua primeira parte executada por 10 vezes, sob uso de cada uma das funções, como evidencia a tabela 1 a seguir.

Tabela 1: Tempos de execução de concatenação

	append	concat
1 ^a	07m02s	10m56s
2 ^a	04m04s	11m25s
3 ^a	03m01s	12m58s
4 ^a	04m22s	10m50s
5 ^a	04m22s	10m52s
6 ^a	04m52s	11m17s
7 ^a	04m08s	11m46s
8 ^a	05m01s	11m07s
9 ^a	04m15s	11m15s
10 ^a	04m29s	12m01s

Fonte: autoria própria

3.5 SEGUNDA OTIMIZAÇÃO

Foi visto que a operação de regressão linear consumia um tempo de execução de 25 min 28 s em média, representando aproximadamente 84% do tempo total do código.

De acordo com Dao (2022), em uma regressão linear, as variáveis preditoras são chamadas de variáveis dependentes ou explicativas; e as variáveis de resposta são conhecidas como variáveis dependentes. Através da formação de um múltiplo de uma equação de regressão linear, é possível descrever e entender como mudanças em uma ou mais variáveis preditoras estão associadas a mudanças na resposta. Em contrapartida, um sistema de classificação apenas agrupa os objetos que possuem características semelhantes. Desta forma foi possível deduzir que um algoritmo de classificação seria menos custoso computacionalmente, otimizando a execução do processo, sem acarretar em perdas significativas de informações. Por esta razão, foi proposta a substituição deste método de análise de informação mútua pelo método de classificação.

O algoritmo de classificação faz uso da função *mutual_info_classification*, novamente oriunda da biblioteca *scikit-learn*. Esta função opera segundo o método de

classificação, correlacionando os vetores de entrada com um vetor alvo, inicialmente determinado, e nomeado *target_vector*. A grande diferença entre as funções *mutual_info_regression* e *mutual_info_classification* dá-se pelo fato que o *target_vector* do algoritmo de regressão é um vetor de valores contínuos, ao passo que para o correspondente código de classificação, o vetor alvo obrigatoriamente precisa ser de formação binária. Para isto, foi necessário o tratamento do vetor de falhas, convertendo os dados numéricos em dados binários. No entanto, o vetor de falhas apresentava distribuição contínua, variando entre 0 e valores quaisquer acima do valor nulo. Como forma de adaptar esse conjunto ao formato de classificação, foram determinados dois possíveis resultados, apresentando valor unitário para falha vigente e valor nulo para o estado oposto. As figuras 12 e 13 evidenciam este processo de formatação.

Figura 13: Vetor de falhas original

index	
2759	0.001305274040633667
1540	0.001133672122888551
2225	0.0007550728150869634
1694	0.0007286752327823565
1684	0.0006868676572548059
5420	0.0004522201530430856
7881	0.0004483055962393888
7751	0.0004480202190384807
8470	0.000426770220505536
0	0.0

Fonte: autoria própria

Figura 14: Vetor de falhas binário

index	
1540	1.0
1684	1.0
1694	1.0
2225	1.0
2759	1.0
5420	1.0
7751	1.0
7881	1.0
8470	1.0
0	0.0

Fonte: autoria própria

Novamente houveram 10 execuções do código, no entanto, desta vez com a segunda parte contemplada para análise, elencada pela tabela 2 a seguir.

Tabela 2: Tempos de execução do cálculo de informação mútua

	Método de regressão	Método de classificação
1 ^a	24m23s	02m02s
2 ^a	26m31s	02m06s
3 ^a	18m16s	02m08s
4 ^a	26m08s	02m06s
5 ^a	26m27s	02m05s
6 ^a	26m17s	02m05s
7 ^a	26m37s	02m05s
8 ^a	26m51s	02m11s
9 ^a	26m36s	02m05s
10 ^a	26m40s	02m08s

Fonte: autoria própria

4 RESULTADOS E DISCUSSÕES

Analisando a primeira otimização foram obtidos os valores médios de 4 min 46 s e 11 min 27 s respectivamente, para as funções *append* e *concat*. Portanto, foi validado que esta situação se tratava de um caso em estado ótimo, sob uso da função *append*.

As funções foram analisadas mais a fundo e foi visto que a função *append* é uma função menos versátil, em comparação com a função *concat* que apresenta mais atributos. Porém, apesar da função *concat* possuir mais atributos e fornecer uma maior variedade de implementações, para a aplicação utilizada, o método *append* supre todas as necessidades e oferece uma melhor resposta em processamento, sendo, portanto, a opção mais indicada para esta situação.

Analisando a segunda otimização, o método de classificação obteve um tempo para execução de 2 min 6 s em média, representando aproximadamente 8% do tempo necessário para implementação do método de regressão, que apresentou tempo médio de 25 min e 28 s. Agora, o código que inicialmente apresentava tempo de execução de 30

min 14 s, apresenta após as otimizações tempo de execução de 6 min 52 s, resultando em uma redução de tempo de aproximadamente 77%.

Dado que o mesmo código, sob uso da mesma base de dados, foi executado por 10 vezes, o resultado ótimo seria um cenário em que os sensores apresentassem o mesmo valor para o coeficiente de informação mútua. No entanto, sob uso do método de regressão para análise de informação mútua na matriz ampliada, os sensores que apresentaram correlação elevada, variaram para cada parte da matriz, proporcionando resultados distintos para cada execução do código. Dessa forma, evidenciando-se que para cada intervalo de dados, analisando um mesmo sensor, há diferença na quantidade de informação, relacionada à falha vigente. Desta forma, o método de regressão apresentou baixa confiabilidade aos resultados, oferecendo dubiedade para o valor de informação mútua, e por isso potencialmente não sendo a melhor opção para a análise de faltas no sistema aqui estudado.

Em vista do resultado anteriormente apresentado o método de análise de dados foi substituído. Agora sob uso do método de classificação, evidenciou-se uniformidade nos sensores com valor de informação mútua elevada, oferecendo confiabilidade aos resultados. Dentre as duas análises apresentadas, o método de classificação foi escolhido como a opção ótima, para ser a ferramenta de análise de dados. Devido à recorrência dos resultados e sensores, o método agora utilizado eleva a acurácia do modelo, associada a um tempo de execução intensamente reduzido.

Devido a extensão do vetor de informação mútua, mesmo após o tratamento de dados aplicado ao arquivo original, foi implementada uma seleção que apresenta apenas os sensores de maior recorrência de aparição. Os apêndices 1 e 2 elucidam estas análises, evidenciando as 10 execuções do código e apresentando os 10 sensores que mais se tornaram recorrentes na análise de informação mútua, para cada uma das execuções.

Desta forma, analisando a coleta de dados em geral, a redução de forma otimizada de um banco de dados, concentrando as leituras mais significativas, relacionadas ao objetivo de análise, pode ser a chave para reduzir a possibilidade de um diagnóstico incorreto, evitando complexidades que poderiam vir a surgir durante o processo.

No entanto, o sistema apresentado não foi cogitado a ser especialista em nenhum momento da análise. Visto que trata-se de um setor crítico da plataforma, faz-se necessária a presença de um operador experiente e presente *in-loco* para validar as análises e averiguar a veracidade do resultado. A redundância é um *modus-operandi* que reduz riscos e eleva a confiabilidade de um setor. No caso da análise de faltas não é

diferente. Por isso, o código foi desenvolvido para auxiliar o operador, oferecendo uma ferramenta que vem a otimizar o tempo de execução de uma análise de dados, reduzindo o número de leituras de acordo com uma base de dados exemplo. No entanto, somente o operador poderá vir a alterar o método adicionando algum sensor que, por ventura, seja crucial à análise correta.

5 CONCLUSÃO

A análise de informação mútua é uma ferramenta fundamental para analisar uma grandeza mediante a ótica de outras variáveis. Neste trabalho ficou evidenciada a utilidade e robustez dessa ferramenta.

Tornou-se evidente a importância da análise do cenário previamente à aplicação do método, e a elucidação de informações correlacionadas em um sistema complexo. A matriz ampliada mostrou-se uma ferramenta fundamental para a validação dos resultados, mediante confronto entre valores obtidos parcialmente de diferentes partes do banco de dados, fornecendo a possibilidade de visualização das lacunas existentes no algoritmo por evidenciar a desconformidade dos resultados.

Vale salientar que a execução repetitiva do código foi a ferramenta utilizada para averiguar a confiabilidade do código. A partir deste método houve a constatação de que o modelo de regressão oferecia resultados desconformes. Como pode ser averiguado no apêndice A, os sensores que apresentaram repetição, com valor de informação mútua acima de 5 %, variaram ao longo das execuções do código. A apresentação de resultados distintos, para uma mesma base de dados, poderia acarretar em riscos à ferramenta desenvolvida. Desta forma, o resultado apresentado poderia orientar o operador a analisar uma informação falsa, a respeito do turbogerador, influenciando-o a uma possível conclusão incorreta. Esta informação foi crucial para a otimização do modelo utilizado, o que acarretou em melhorias significativas nos resultados e no tempo de execução.

Em contrapartida, sob uso do método de classificação, que pode ter os resultados averiguados no apêndice B, os sensores que apresentaram repetição, com valor de informação mútua acima de 5 %, não apresentaram variação alguma. Essa informação, da uniformidade dos resultados a partir de pontos de vista distintos, visto que foi utilizada

a matriz ampliada, da mesma forma que sob uso do método de regressão, atribui confiabilidade aos resultados, também a partir da execução repetitiva do código.

Pelo fato do vetor de falhas tratar-se de um conjunto de dados, igualmente espaçados no domínio do tempo, em conformidade às leituras dos sensores, foi visto que a sua informação deveria ser binária, para adequar-se ao modelo de classificação. Portanto a adaptação do vetor de falhas para um vetor binário foi fundamental para a otimização do sistema, possibilitando ainda a substituição do método de regressão linear para um método de classificação.

6 REFERÊNCIAS

BENOUIOUA, Djedjiga; CANDUSSO, Denis; HAREL, Fabien; PICARD, Pierre. **Diagnosis of fuel cells using instantaneous frequencies and envelopes extracted from stack voltage signal**. *International Journal of Hydrogen Energy*. 2022.

BRITO, Lucas C.; SUSTO, Gian Antonio; BRITO, Jorge N.; DUARTE, Marcus A. V. **An explainable artificial intelligence approach for unsupervised fault detection and diagnosis in rotating machinery**. *Mechanical Systems and Signal Processing*. 2002.

BUKSH, Z. A. et al, **Predictive maintenance using tree-based classification techniques: A case of railway switches**, *Transportation Research Part C: Emerging Technologies*, Volume 101, 2019.

CASTRO, Patrício Fernandes de. **Análise de Confiabilidade do Sistema Elétrico de uma Plataforma de Produção de Petróleo e Gás**. João Pessoa, 2020. Dissertação (Mestrado) – UFPB/CEAR.

DAO, Phong B.. **Condition monitoring and fault diagnosis of wind turbines based on structural break detection in SCADA data**. *Renewable Energy*. 2021.

GONDRA, Iker. **Applying machine learning to software fault-proneness prediction**. *Journal of Systems and Software*. 2008.

LEI, Yaguo; LIN, Jing; HE, Zhengjia; ZUO, Ming J. **A review on empirical mode decomposition in fault diagnosis of rotating machinery**. *Mechanical Systems and Signal Processing*. 2013.

LI, Jingchao; YING, Yulong; JI, Chunlei. **Study on gas turbine gas-path fault diagnosis method based on quadratic entropy feature extraction**. *IEEE Access*. 2019.

LI, Wentian. **Mutual Information Functions Vs Correlation Functions**. *Journal of Statistical Physics*, 1990.

LIU, Hui; LIU, Zhenyu; JIA, Weiqiang; ZHANG, Donghao; TAN, Jianrong. **A Novel Imbalanced Data Classification Method Based on Weakly Supervised Learning for Fault Diagnosis**. *IEEE Transactions on Industrial Informatics*. 2022.

LIU, Ruonan; YANG, Boyuan; ZIO, Enrico; CHEN, Xuefeng. **Artificial intelligence for fault diagnosis of rotating machinery: A review**. *Mechanical Systems and Signal Processing*. 2018.

MA, Liang; DING, Yu; WANG, Zili; WANG, Chao; MA, Jian; LU, Chen.. **An interpretable data augmentation scheme for machine fault diagnosis based on a**

sparsity-constrained generative adversarial network. Experts Systems with Applications. 2021.

MOUSAVI, Mehdi; CHAIBAKHSH, Ali; JAMALI, Ali; KORDESTANI, Mojtaba; SAIF, Mehrdad. **A New Fault Diagnosis Approach for Heavy-Duty Gas Turbines.** IEEE/ASME Transactions on Mechatronics. 2022.

OFFSHORE ENGINEER. New Oil and Gas Project Spending to Fall 68% in 2020 if Oil Price Stays Low. Disponível em: <https://www.oedigital.com/news/476916-new-oil-and-gas-project-spending-to-fall-68-in-2020-if-oil-price-stays-low>. Acesso em: 25 Jun 2022.

PETRÓLEO BRASILEIRO S.A. – PETROBRÁS. Rio de Janeiro, 1953-2020. Disponível em: https://petrobras.com.br/en_us/fatos-e-dados/iniciamos-a-operacao-da-primeira-plataforma-definitiva-no-campo-de-mero-no-pre-sal-da-bacia-de-santos.html. Acesso em: 21 Jun. 2022.

PORTAL PETRÓLEO E ENERGIA. Disponível em: <https://www.petroleoenergia.com.br/turbinas-energia-nas-plataformas-depende-de-modelos-aeroderivados-e-movidos-a-gas/>. Acesso em: 26 Jun. 2022.

PINTO, Marcela Coury. **Sistema de manutenção preditiva de falhas em válvulas em um processo industrial utilizando inteligência artificial.** Trabalho de Conclusão de Curso – Universidade Federal de Uberlândia – UFU Faculdade de Engenharia Elétrica. 2021.

SANTOS, Mailson Ribeiro; COSTA, Bruno Sielly Jales; BEZERRA, Clauber Gomes; GUEDES, Luiz Affonso. **An evolving framework for fault diagnosis of dynamic systems.** Software Impacts. 2022.

SUN, Hongchun; CAO, Xu; WANG, Changdong; GAO, Sheng. **An interpretable anti-noise network for rolling bearing fault diagnosis based on FSWT.** Measurement. 2022.

TRAN, Minh-Quang; ELSISI, Mahmoud; LIU, Meng-Kun. **Effective feature selection with fuzzy entropy and similarity classifier for chatter vibration diagnosis.** Measurement. 2021.

APÊNDICE 1 – RECORRÊNCIA DOS RESULTADOS – MÉTODO DE REGRESSÃO

Figura A1.1: 1ª execução - 10 resultados mais relevantes

1	Frequência ▼	Posição
P74_BE-5147501A-B	19	2
P74_BE-5147501A-A	17	1
P74_PDIT-5147522A-B	13	6
P74_TE-5147509A-D	13	28
P74_TE-5147511A-A	13	33
P74_TE-5147516A-B	12	36
P74_TE-5147524A-A	12	45
P74_VT-5147502A	12	53
P74_PIT-5147530A-B	11	9
P74_TE-5147508A-A	11	17

Show 10 per page

Fonte: autoria própria

Figura A1.2: 2ª execução - 10 resultados mais relevantes

1	Frequência ▼	Posição
P74_TE-5147509A-D	19	28
P74_BE-5147501A-A	17	1
P74_PDIT-5147522A-B	17	6
P74_TE-5147503A	15	14
P74_TE-5147511A-A	15	33
P74_TIT-5147501A-B	15	50
P74_BE-5147501A-B	14	2
P74_TE-5147508A-C	14	19
P74_VT-5147501A	14	52
P74_ZT-5147503A-B	14	55

Show 10 per page

Fonte: autoria própria

Figura A1.3: 3ª execução - 10 resultados mais relevantes

1	Frequência ▼	Posição
P74_BE-5147501A-B	14	2
P74_BE-5147501A-A	12	1
P74_TE-5147511A-A	10	33
P74_VT-5147502A	10	53
P74_TE-5147509A-D	9	28
P74_ZT-5147503A-A	9	54
P74_PIT-5147530A-A	8	8
P74_PT-5147502A-B	8	11
P74_PT-5147503A-A	7	12
P74_TE-5147508A-E	7	21

Show 10 per page

Fonte: autoria própria

Figura A1.4: 4ª execução - 10 resultados mais relevantes

1	Frequência ▼	Posição
P74_BE-5147501A-A	14	1
P74_TE-5147510A-A	12	31
P74_BE-5147501A-B	11	2
P74_PIT-5147501A	11	7
P74_TE-5147509A-D	10	28
P74_TE-5147508A-C	9	19
P74_TE-5147508A-H	9	24
P74_LT-5147501A	8	3
P74_TE-5147508A-D	8	20
P74_TE-5147511A-A	8	33

Show 10 per page

Fonte: autoria própria

Figura A1.5: 5ª execução - 10 resultados mais relevantes

1	Frequência ▼	Posição
P74_BE-5147501A-B	17	2
P74_TE-5147509A-D	14	28
P74_BE-5147501A-A	13	1
P74_TE-5147506A-B	11	16
P74_TE-5147511A-A	11	33
P74_TIT-5147518A	11	51
P74_ZT-5147503A-A	11	54
P74_LT-5147501A	10	3
P74_PIT-5147530A-A	10	8
P74_TE-5147508A-E	10	21

Show 10 per page

Fonte: autoria própria

Figura A1.6: 6ª execução - 10 resultados mais relevantes

1	Frequência ▼	Posição
P74_BE-5147501A-A	18	1
P74_PIT-5147530A-A	15	8
P74_TE-5147509A-D	15	28
P74_BE-5147501A-B	14	2
P74_VT-5147502A	13	53
P74_TE-5147504A	12	15
P74_TE-5147508A-C	12	19
P74_TE-5147503A	11	14
P74_PDIT-5147520A	10	4
P74_PT-5147502A-A	10	10

Show 10 per page

Fonte: autoria própria

Figura A1.7: 7ª execução - 10 resultados mais relevantes

1	Frequência ▼	Posição
P74_BE-5147501A-A	17	1
P74_BE-5147501A-B	12	2
P74_TE-5147511A-B	12	34
P74_ZT-5147503A-B	12	55
P74_TE-5147521A-B	11	40
P74_PDIT-5147522A-A	10	5
P74_TE-5147511A-A	10	33
P74_VT-5147502A	10	53
P74_ZT-5147503A-A	10	54
P74_PDIT-5147522A-B	9	6

Show 10 per page

Fonte: autoria própria

Figura A1.8: 8ª execução - 10 resultados mais relevantes

1	Frequência ▼	Posição
P74_BE-5147501A-A	17	1
P74_TE-5147509A-D	16	28
P74_TE-5147510A-A	15	31
P74_PIT-5147501A	14	7
P74_TE-5147508A-C	13	19
P74_TE-5147508A-G	13	23
P74_TE-5147521A-A	13	39
P74_BE-5147501A-B	12	2
P74_TE-5147508A-B	12	18
P74_TE-5147509A-B	12	26

Show 10 per page

Fonte: autoria própria

Figura A1.9: 9ª execução - 10 resultados mais relevantes

1	Frequência ▼	Posição
P74_BE-5147501A-A	14	1
P74_BE-5147501A-B	13	2
P74_TE-5147509A-D	13	28
P74_TE-5147508A-C	11	19
P74_LT-5147501A	10	3
P74_PIT-5147530A-B	10	9
P74_TE-5147509A-C	10	27
P74_TE-5147521A-A	10	39
P74_PDIT-5147522A-B	9	6
P74_TE-5147508A-D	9	20

Show 10 per page

1 2 3 4 5 6

Fonte: autoria própria

Figura A1.10: 10ª execução - 10 resultados mais relevantes

1	Frequência ▼	Posição
P74_BE-5147501A-A	15	1
P74_BE-5147501A-B	13	2
P74_TE-5147509A-D	12	27
P74_TE-5147509A-C	11	26
P74_VT-5147502A	11	51
P74_PIT-5147530A-B	10	9
P74_TE-5147509A-B	9	25
P74_TE-5147508A-B	8	18
P74_TE-5147508A-D	8	20
P74_TE-5147508A-E	8	21

Show 10 per page

1 2 3 4 5 6

Fonte: autoria própria

APÊNDICE 2 – RECORRÊNCIA DOS RESULTADOS – MÉTODO DE CLASSIFICAÇÃO

Figura A2.1: 1ª execução - 10 resultados mais relevantes

1	Frequência ▼	Posição
P74_TE-5147517A-A	9	37
P74_TE-5147516A-B	7	36
P74_LT-5147501A	6	3
P74_PDIT-5147522A-B	6	6
P74_TE-5147506A-B	6	16
P74_TE-5147511A-A	6	33
P74_TE-5147516A-A	6	35
P74_TE-5147509A-C	5	27
P74_TE-5147522A-B	5	42
P74_TE-5147524A-A	5	45

Show 10 per page

1 2 3 4 5 6

Fonte: autoria própria

Figura A2.2: 2ª execução - 10 resultados mais relevantes

1	Frequência ▼	Posição
P74_TE-5147517A-A	9	37
P74_TE-5147516A-B	7	36
P74_LT-5147501A	6	3
P74_PDIT-5147522A-B	6	6
P74_TE-5147506A-B	6	16
P74_TE-5147511A-A	6	33
P74_TE-5147516A-A	6	35
P74_TE-5147509A-C	5	27
P74_TE-5147522A-B	5	42
P74_TE-5147524A-A	5	45

Show 10 per page

1 2 3 4 5 6

Fonte: autoria própria

Figura A2.3: 3ª execução - 10 resultados mais relevantes

1	Frequência ▼	Posição
P74_TE-5147517A-A	9	37
P74_TE-5147516A-B	7	36
P74_LT-5147501A	6	3
P74_PDIT-5147522A-B	6	6
P74_TE-5147506A-B	6	16
P74_TE-5147511A-A	6	33
P74_TE-5147516A-A	6	35
P74_TE-5147509A-C	5	27
P74_TE-5147522A-B	5	42
P74_TE-5147524A-A	5	45

Show 10 per page

1 2 3 4 5 6

Fonte: autoria própria

Figura A2.4: 4ª execução - 10 resultados mais relevantes

1	Frequência ▼	Posição
P74_TE-5147517A-A	9	37
P74_TE-5147516A-B	7	36
P74_LT-5147501A	6	3
P74_PDIT-5147522A-B	6	6
P74_TE-5147506A-B	6	16
P74_TE-5147511A-A	6	33
P74_TE-5147516A-A	6	35
P74_TE-5147509A-C	5	27
P74_TE-5147522A-B	5	42
P74_TE-5147524A-A	5	45

Show 10 per page

1 2 3 4 5 6

Fonte: autoria própria

Figura A2.5: 5ª execução - 10 resultados mais relevantes

1	Frequência ▼	Posição
P74_TE-5147517A-A	9	37
P74_TE-5147516A-B	7	36
P74_LT-5147501A	6	3
P74_PDIT-5147522A-B	6	6
P74_TE-5147506A-B	6	16
P74_TE-5147511A-A	6	33
P74_TE-5147516A-A	6	35
P74_TE-5147509A-C	5	27
P74_TE-5147522A-B	5	42
P74_TE-5147524A-A	5	45

Show 10 per page

Fonte: autoria própria

Figura A2.6: 6ª execução - 10 resultados mais relevantes

1	Frequência ▼	Posição
P74_TE-5147517A-A	9	37
P74_TE-5147516A-B	7	36
P74_LT-5147501A	6	3
P74_PDIT-5147522A-B	6	6
P74_TE-5147506A-B	6	16
P74_TE-5147511A-A	6	33
P74_TE-5147516A-A	6	35
P74_TE-5147509A-C	5	27
P74_TE-5147522A-B	5	42
P74_TE-5147524A-A	5	45

Show 10 per page

Fonte: autoria própria

Figura A2.7: 7ª execução - 10 resultados mais relevantes

1	Frequência ▼	Posição
P74_TE-5147517A-A	9	37
P74_TE-5147516A-B	7	36
P74_LT-5147501A	6	3
P74_PDIT-5147522A-B	6	6
P74_TE-5147506A-B	6	16
P74_TE-5147511A-A	6	33
P74_TE-5147516A-A	6	35
P74_TE-5147509A-C	5	27
P74_TE-5147522A-B	5	42
P74_TE-5147524A-A	5	45

Show 10 per page

Fonte: autoria própria

Figura A2.8: 8ª execução - 10 resultados mais relevantes

1	Frequência ▼	Posição
P74_TE-5147517A-A	9	37
P74_TE-5147516A-B	7	36
P74_LT-5147501A	6	3
P74_PDIT-5147522A-B	6	6
P74_TE-5147506A-B	6	16
P74_TE-5147511A-A	6	33
P74_TE-5147516A-A	6	35
P74_TE-5147509A-C	5	27
P74_TE-5147522A-B	5	42
P74_TE-5147524A-A	5	45

Show 10 per page

Fonte: autoria própria

Figura A2.9: 9ª execução - 10 resultados mais relevantes

1	Frequência ▼	Posição
P74_TE-5147517A-A	9	37
P74_TE-5147516A-B	7	36
P74_LT-5147501A	6	3
P74_PDIT-5147522A-B	6	6
P74_TE-5147506A-B	6	16
P74_TE-5147511A-A	6	33
P74_TE-5147516A-A	6	35
P74_TE-5147509A-C	5	27
P74_TE-5147522A-B	5	42
P74_TE-5147524A-A	5	45

Show 10 per page

1 2 3 4 5 6

Fonte: autoria própria

Figura A2.9: 9ª execução - 10 resultados mais relevantes

1	Frequência ▼	Posição
P74_TE-5147517A-A	9	37
P74_TE-5147516A-B	7	36
P74_LT-5147501A	6	3
P74_PDIT-5147522A-B	6	6
P74_TE-5147506A-B	6	16
P74_TE-5147511A-A	6	33
P74_TE-5147516A-A	6	35
P74_TE-5147509A-C	5	27
P74_TE-5147522A-B	5	42
P74_TE-5147524A-A	5	45

Show 10 per page

1 2 3 4 5 6

Fonte: autoria própria