



Prediction of Polymerization Degree Using  
Machine Learning: New Methodology for  
Assessing the Lifespan of Power Transformers

---

Rafael Fehlberg, Daniel Carrijo, Gabriel Gomes, Sofia Lopes,  
Rogério Flauzino, Renan Santa Rosa and Iony Patriota

EasyChair preprints are intended for rapid  
dissemination of research results and are  
integrated with the rest of EasyChair.

October 26, 2023

# Predição do Grau de Polimerização Utilizando Machine Learning: Nova Metodologia de Avaliação da Vida Útil de Transformadores de Potência

Rafael Prux Fehlberg\* Daniel Carrijo Polonio Araujo\*  
Gabriel de Souza Pereira Gomes\* Sofia Moreira de Andrade Lopes\*  
Rogério Andrade Flauzino\* Renan Ferreira Santa Rosa\*\*  
Iony Patriota de Siqueira\*\*\*

\* *Esc. Eng. de S. Carlos - EESC, Univ. de S. Paulo - USP, SP, Brasil*  
(e-mail: rafapf@usp.br)

\*\* *Treetech Tecnologia, Atibaia, SP, Brasil*

\*\*\* *Tecnix Engenharia e Arquitetura Ltda. Recife, PE, Brasil*

---

**Abstract:** The main indicator used to assess the condition of solid insulation in power equipment currently is the degree of polymerization (DP). This work presents a systematic study where machine learning techniques are used to estimate DP from 2-fal and other indicators. The results are promising, indicating that 2-fal, CO<sub>2</sub>/CO, the Chendong formula, and equipment power can be used together to better predict the current value of its service life.

**Resumo:** O principal indicador de avaliação do estado da isolação sólida em equipamentos de potência atualmente é o grau de polimerização (GP). Este trabalho apresenta um estudo sistemático onde técnicas de machine learning são utilizadas para estimar o GP a partir do 2-fal e de outros indicadores. Os resultados são promissores, indicando que 2-fal, CO<sub>2</sub>/CO, a fórmula de Chendong e a potência do equipamento podem ser empregadas em conjunto para melhor previsão do valor atual da sua vida útil.

*Keywords:*

Aging, Power Transformer, Assets, Degree of Polymerization

*Palavras-chaves:*

Envelhecimento, Transformador de Potência, Ativos, Grau de Polimerização

---

## 1. INTRODUÇÃO

Para a tomada de decisões acerca da manutenção e operação de transformadores de potência, um dos principais fatores é a vida útil. Estudos têm sido desenvolvidos de modo a estimar a vida útil da isolação de transformadores, como Elele et al. (2022), Shutenko and Ponomarenko (2022) e Oria et al. (2019). Atualmente, estudos correlacionam o grau de polimerização (GP) ao envelhecimento do papel isolante do transformador, de forma que o valor de GP pode ser utilizado para estimar a vida útil dos transformadores, Münster et al. (2018).

Por conta da dificuldade da realização de ensaios de GP, estudos conseguiram com sucesso relacionar o grau de polimerização da isolação com o nível de compostos furânicos presentes no óleo isolante. O composto 2-furfuraldeído (2-fal) é o mais aplicado. Assim como o uso desses compostos, alguns trabalhos correlacionam a proporção dos gases CO<sub>2</sub> e CO com o tempo de vida útil do transformador, porém sempre considerando técnicas separadas, como em Aciu et al. (2021).

É proposta uma nova abordagem para a estimação do GP, através da unificação dos métodos, utilizando a aprendizagem de máquinas no modelamento do comportamento do transformador, objetivando uma melhor previsão do tempo de vida.

## 2. VISÃO GERAL DO PROBLEMA

A principal causa de falha em transformadores de força é o envelhecimento de sua isolação. Muitos subprodutos são gerados através do aquecimento e resfriamento do óleo em contato com o papel isolante. A decomposição térmica de papel impregnado com óleo gera os gases CO e CO<sub>2</sub>. A taxa da geração aumenta exponencialmente com o aumento da temperatura do papel. Compostos furânicos também são gerados pela degradação do papel isolante impregnado com óleo e são consideravelmente solúveis em óleo. Todos esses compostos podem ser utilizados para avaliar o estado do transformador, auxiliando a evitar falhas devido ao envelhecimento. Porém seus valores não são indicativos diretos do estado da isolação sólida.

Para indicar o estágio de envelhecimento de transformadores isolados por papel impregnado de óleo de maneira precisa, podemos utilizar a medição do GP (grau de polimerização), que indica quantos monômeros existem em uma cadeia de celulose. Para medir o GP de transformadores é necessário um pedaço do papel para análise em laboratório. Para remover um pedaço do papel, é necessário desligar e abrir o transformador, o que gera um alto custo de manutenção. Devido a isso, diversos métodos ao longo dos anos foram desenvolvidos para estimar o GP a partir de ensaios mais simples, que mensuram

subprodutos do envelhecimento, como gases  $\text{CO}_2$  e  $\text{CO}$  e também 2-furfuraldeído (2-FAL). Nesta linha, foram propostas diversas fórmulas para calcular a relação entre a concentração de 2-fal e o valor de GP. Das propostas, podemos citar as de Chendong, Scholnick, De Pablo e Pahlavanpour (Teymouri and Vahidi (2017)), porém, essas fórmulas apresentam valores discrepantes, mostrando que apenas os valores de 2-fal não são suficientes para o cálculo preciso do valor de GP. Esse trabalho propõe uma solução capaz de calcular o valor de GP empregando os valores de 2-fal e da relação  $\text{CO}_2/\text{CO}$  e aprendizagem de máquina.

### 2.1 Descrição dos dados e abordagem utilizada

Neste trabalho, o método proposto é baseado em dados fornecidos por outros trabalhos realizados na mesma área. Podemos dividir os dados em dois grandes grupos, sendo o primeiro deles extraído de Assuncao (2007), apresentado na Tabela 1. Esses dados possuem diversas medições de 2-fal e as medições de GP correspondentes.

Tabela 1. Dados extraídos de Assuncao (2007)

2-fal [ppb]	GPM
60	1044
80	971
...	...
...	...
2551	214
2784	197

O segundo conjunto de dados utilizado foi extraído de Teymouri and Vahidi (2017). A Tabela 2 apresenta dados contendo as medições dos transformadores, como as concentrações dos compostos furânicos que foram medidos usando cromatografia líquida de alta pressão; a relação  $\text{CO}_2/\text{CO}$  calculada a partir da cromatografia gasosa; GP, medido mediante a análise em laboratório; e a potência do transformador (MVA). O segundo conjunto de dados é mais completo que o primeiro e apresenta mais informações que serão consideradas na modelagem do problema.

Tabela 2. Dados de referência dos Transformadores

TR.	Pot. (MVA)	$\text{CO}_2/\text{CO}$	2-fal (ppb)	GPM
1	95	6,2	10	1089
2	105	6,4	13	1042
3	96	6,5	18	983
4	...	...	...	...
5	...	...	...	...
28	85	8,7	1060	243
29	154	8,85	1090	235
30	123	8,9	1180	214

Este trabalho foi desenvolvido conforme o fluxograma apresentado na Figura 1.

### 2.2 Análise exploratória e tratamento inicial dos dados

A análise exploratória dos dados foi realizada separadamente, devido a dois fatores: os dados têm origem diferente e possuem atributos diferentes. Depois, conclusões foram obtidas a partir da comparação dos resultados.

Os dados de Assuncao (2007) possuem apenas dois atributos, sendo eles: 2-fal e GP. Inicialmente na Tabela 3 são apresentados parâmetros estatísticos básicos desses dados.

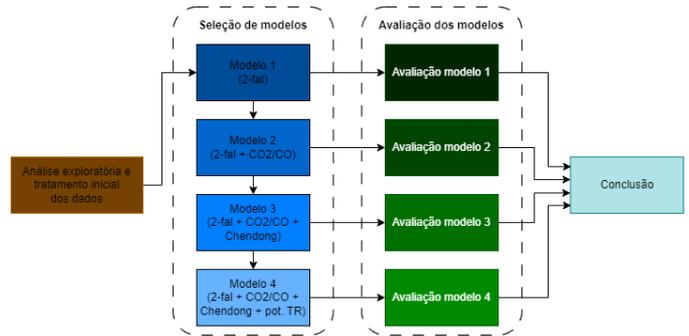


Figura 1. Metodologia utilizada

Tabela 3. Parâmetros estatísticos dos dados

	GPM	2-fal (ppb)
count	36,000000	36,000000
média	470,861111	1176,877778
std	233,966072	863,447395
min	197,000000	60,000000
50 %	415,500000	1007,500000
max	1044,000000	2784,000000

Metade dos dados estão abaixo de 415 de GP. Somado a isso, a média e a mediana são diferentes, indicando que a distribuição estatística dos dados não se comporta como uma normal. Esse fato é verificado através do histograma apresentado na Figura 2.

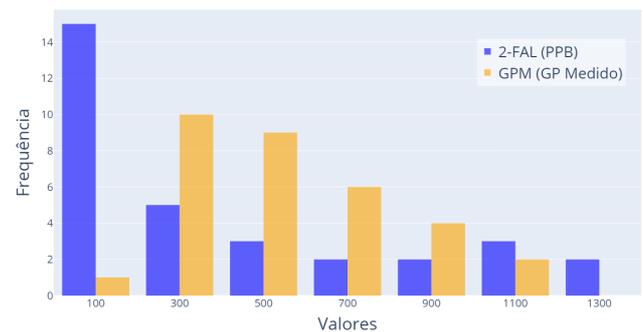


Figura 2. Histograma dos dados

Conforme é possível observar, a maior parte dos dados se concentra para valores de GP e 2-fal baixos, criando um deslocamento à esquerda dos dados. A relação entre esses dois atributos pode ser verificada na Figura 3.

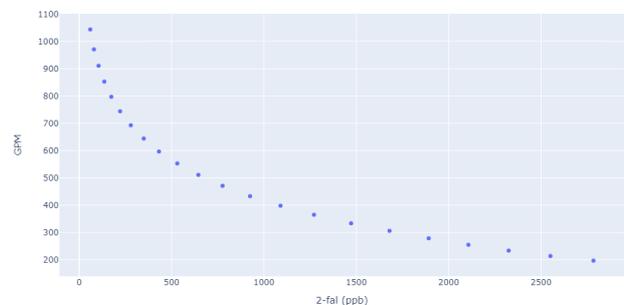


Figura 3. Relação entre GP e 2-fal

A relação entre os dados aparenta se ajustar a uma curva do tipo exponencial decrescente. De fato, os dados parecem

terem sido tratados e não serem os dados brutos das medições, pois não existem pontos que se deslocam para fora da curva.

A descrição dos dados [Teymouri and Vahidi \(2017\)](#) pode ser visualizada na Tabela 4.

Tabela 4. Descrição dos dados

	Pot. (MVA)	CO <sub>2</sub> /CO	2-fal (ppb)	GPM
count	32,000000	32,000000	32,000000	32,000000
mean	86,531250	7,655312	413,093750	554,781250
std	39,240355	0,797840	420,861126	272,054161
min	10,000000	6,200000	10,000000	191,000000
50 %	95,500000	7,760000	262,000000	509,500000
max	154,000000	9,200000	1285,000000	1089,000000

Dados de 32 transformadores foram utilizados, sendo que suas potências variavam de 10 MVA até 154 MVA, ou seja, transformadores de médio e grande porte. Além disso, a distribuição de GP para esse caso estava menos deslocada, uma vez que a média e a mediana estavam mais próximas, conforme apresentado no histograma da Figura 4.

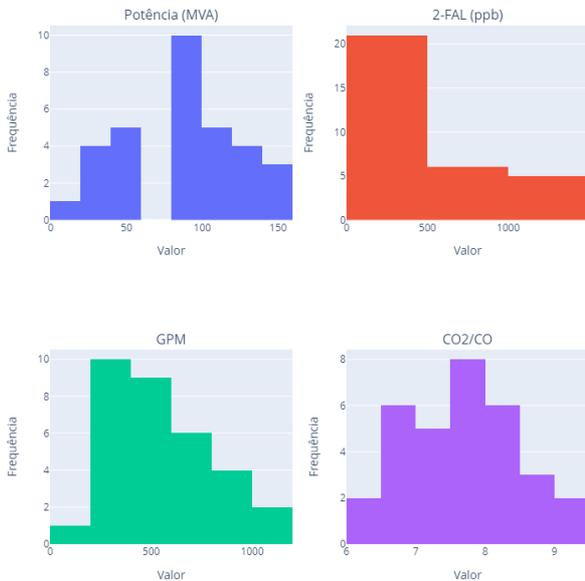


Figura 4. Histograma dos atributos

A relação entre GP e 2-fal pode ser verificada na Figura 5.

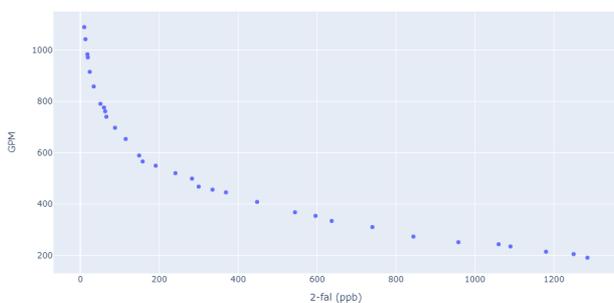


Figura 5. Relação GP e 2-fal

A relação também se assemelha a uma exponencial negativa, porém esses dados parecem ter sido menos tratados, uma vez que o ajuste a uma curva não é perfeito. Além disso, é possível também verificar a relação entre GP e CO<sub>2</sub>/CO, conforme a Figura 6.

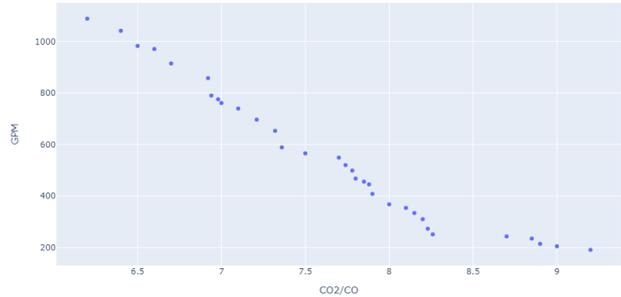


Figura 6. Relação GP e CO<sub>2</sub>/CO

Isso é um forte indicativo de que o valor da relação CO<sub>2</sub>/CO também pode ser utilizada para prever o GP do transformador. Para analisarmos melhor essas relações, convém que uma matriz de correlação seja montada, conforme a Tabela 5. Nela, é possível observar que a correlação entre CO<sub>2</sub>/CO e 2-fal possui um alto valor (0,93) e ambos apresentam alta correlação com o valor de GP. A alta correlação entre CO<sub>2</sub>/CO e 2-fal é um indicativo de que ambos os atributos possuem informação redundante quando relacionados ao GP. Essa alta correção está ligada ao fato de que ambos CO<sub>2</sub>/CO e 2-fal são produtos diretos do envelhecimento do papel, sendo produzidos de forma semelhante por sua degradação [Saha and Purkait \(2017\)](#). Todavia, como essa correlação não é perfeita, existe informação na relação CO<sub>2</sub>/CO que não está presente na medição de 2-fal, tornando possível a utilização do CO<sub>2</sub>/CO para realização de um ajuste fino em um modelo de aprendizagem de máquina.

Tabela 5. Matriz de Correlação

	Pot. (MVA)	CO <sub>2</sub> /CO	2-fal (ppb)	GPM
Pot. (MVA)	1,000000	0,344027	0,446999	-0,264492
CO <sub>2</sub> /CO	0,344027	1,000000	0,934144	-0,975286
2-fal (ppb)	0,446999	0,934144	1,000000	-0,879303
GPM	-0,264492	-0,975286	-0,879303	1,000000

A relação entre os principais atributos (CO<sub>2</sub>/CO, 2-fal e GP) pode ser visualizada na Figura 7.

Ambos os dados utilizados apresentam relação entre 2-fal e GP semelhante, próxima a uma exponencial decrescente. Os dados de [Assuncao \(2007\)](#) possuem um comportamento mais filtrado enquanto os dados [Teymouri and Vahidi \(2017\)](#) possuem um comportamento menos filtrado. Além disso, foi verificado que a relação CO<sub>2</sub>/CO, por mais que possua alta correlação com o 2-fal, pode ser empregada como atributo na previsão do grau de polimerização do transformador.

### 2.3 Modelos propostos

#### 2.3.1. Modelo de previsão do GP a partir do 2-fal

O primeiro modelo a ser criado aplicará apenas o valor de 2-fal para prever o valor do GP. Para isso, serão utilizados

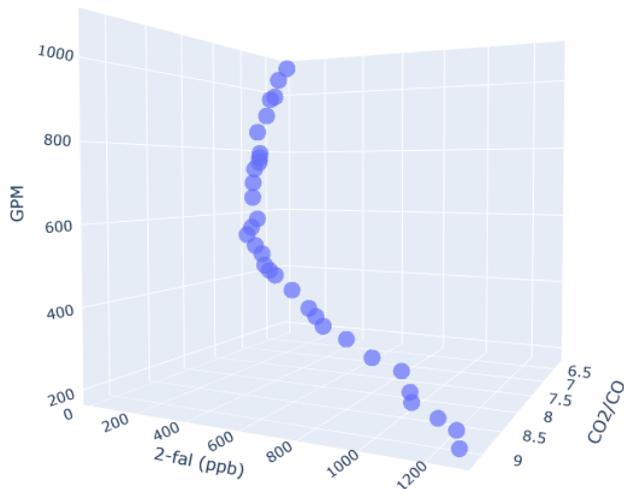


Figura 7. Relação entre GP, 2-fal e CO<sub>2</sub>/CO

os dados de Assuncao (2007) e o problema será tratado como um de regressão, no qual a saída GP será calculada / predita a partir da entrada 2-fal. Após isso, o modelo será testado com os dados de Teymouri and Vahidi (2017).

Os dados serão divididos entre treinamento e teste, e o grau do polinômio será definido através de um problema de otimização, cujo objetivo é obter a maior média de erro médio ao longo dos folds. A proporção para treinamento e teste considerada foi de 80 % e 20 %, pois a quantidade de dados é pequena. Os erros médios para cada grau polinomial no teste podem ser visualizados na Tabela 6.

Tabela 6. Mean Average Error ao longo dos folds para cada grau de polinômio

Grau	1	2	3	4	5
MAE	-25,10	-24,4	-24,7	-24,35	-23,22
Grau	6	7	8	9	10
MAE	-21,91	-20,66	-19,01	-18,76	-18,51
Grau	11	12	13		
MAE	-14,33	-17,33	-33,69		

O menor erro médio é obtido para o polinômio de grau 11. Aplicando esse polinômio nos dados de teste, obtêm-se um erro médio de 211. O resultado do treinamento e do teste podem ser visualizados nas Figura 8 e Figura 9.

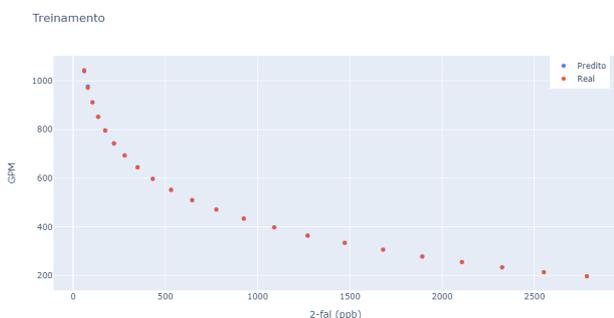


Figura 8. Relação entre GP predito e GP medido para o treinamento

Nos dados de teste percebe-se um offset entre as curvas. Por mais que o formato seja bastante semelhante, esse

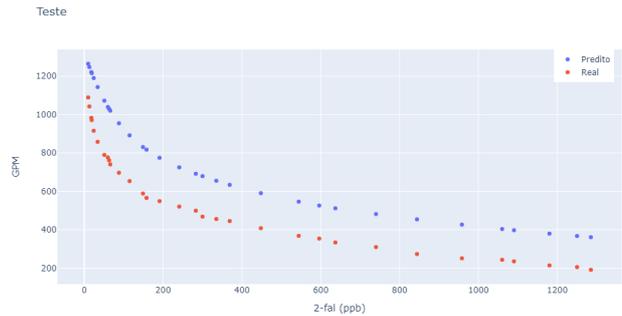


Figura 9. Relação entre GP predito e GP medido para o teste

offset leva a um erro médio de 211. Esse erro pode ser crítico principalmente para transformadores no fim de vida, uma vez que o valor predito indicaria um estado de deterioração avançado, mas não crítico. Faz-se necessária a consideração de mais atributos além do 2-fal para que a realização dessas previsões possa ser mais precisa.

### 2.3.2. Modelo de previsão do GP a partir do 2-fal e da relação CO<sub>2</sub>/CO

O segundo modelo proposto aplica também a relação CO<sub>2</sub>/CO, como apresentado em Teymouri and Vahidi (2017), o uso desta relação pode ser empregado para prever o GP de um transformador e possui alta taxa de correlação com o GP. Além disso, utilizar as medições de CO<sub>2</sub>/CO apresenta vantagens como:

- 2-fal é gerado em diferentes formas e, portanto, não é exclusivamente o resultado da degradação de celulose Allan (1993)
- 2-fal não é estável em temperaturas acima 110 °C;
- A presença de oxigênio, água e ácidos afeta a concentração de 2-fal Prevost (2005).

Diferentemente do primeiro, no qual temos apenas um atributo de entrada (2-fal), para esse segundo modelo, convém aplicar machine learning. Como a quantidade de dados é pequena, será utilizado um modelo de aprendizagem de máquina simples. Além disso, devido à alta colinearidade entre a relação CO<sub>2</sub>/CO e 2-fal, será necessária a aplicação de um modelo que penalize atributos com alta correlação, como o Ridge Regression. As métricas obtidas no treinamento e no teste podem ser visualizadas na Tabela 7.

Tabela 7. Métricas do modelo com 2-fal e a relação CO<sub>2</sub>/CO

	R2	MAE
Treinamento	0,93	58,41
Teste	0,87	77,82

Conforme é possível perceber, ocorreu um pequeno overfitting aos dados de treinamento. Todavia o resultado apresentado pelo teste foi bastante satisfatório, atingindo um erro médio de 77,82, bem menor quando comparado ao primeiro modelo. Os resultados gráficos podem ser visualizados nas Figura 10 e Figura 11. Verifica-se que o erro não é linear ao longo das amostras.

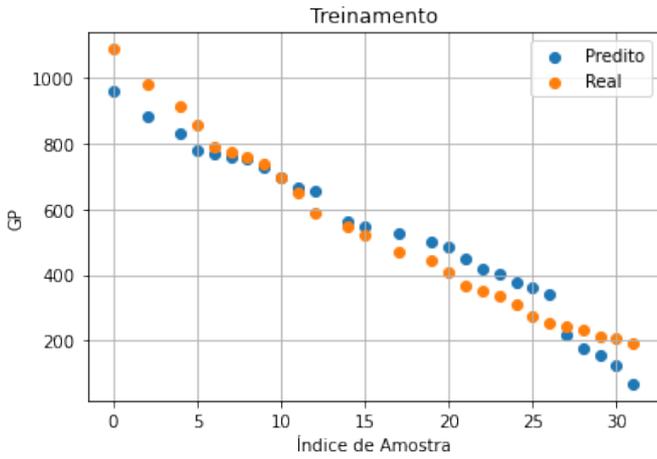


Figura 10. Relação entre GP predito e GP medido para o treinamento

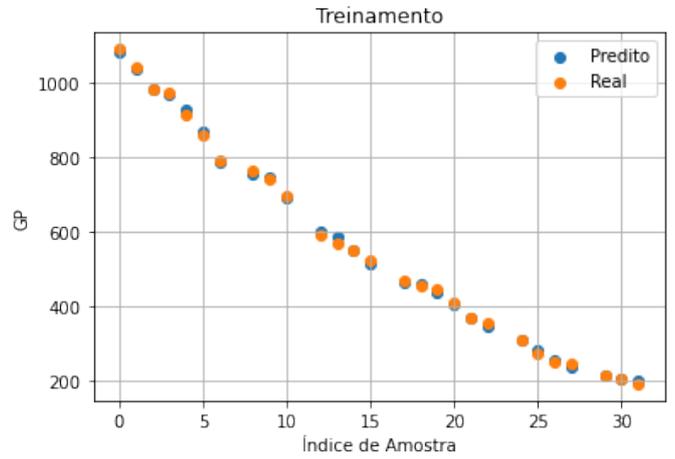


Figura 12. Relação entre GP predito e GP medido para o treinamento

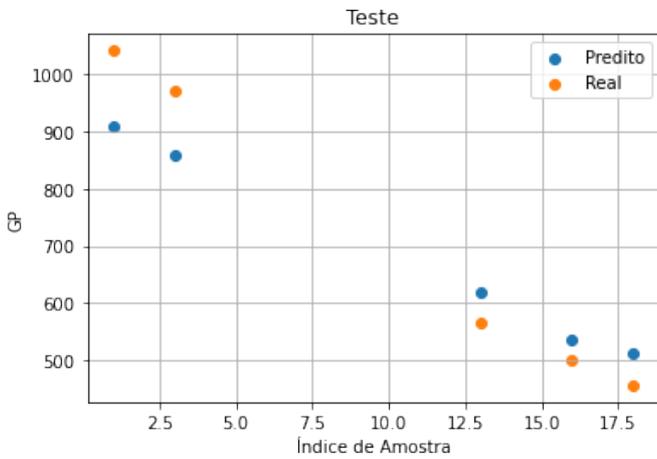


Figura 11. Relação entre GP predito e GP medido para o teste

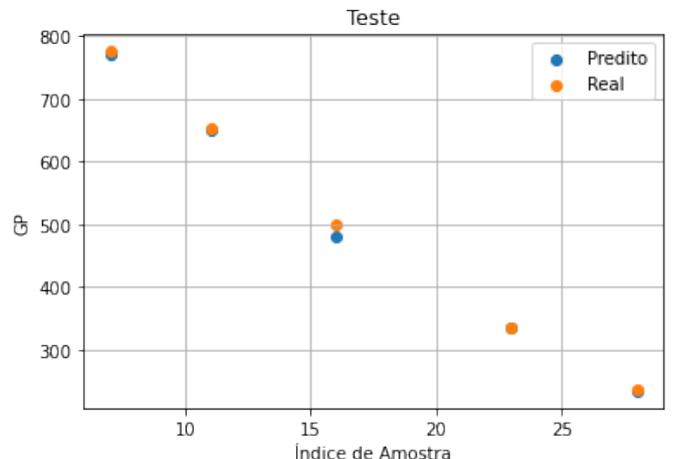


Figura 13. Relação entre GP predito e GP medido para o teste

### 2.3.3. Modelo de previsão do GP a partir do 2-fal, da relação $CO_2/CO$ e da fórmula de Chendong

Como a quantidade de dados utilizada para treinamento desse modelo é pequena, apenas a base de [Teymouri and Vahidi \(2017\)](#), decidiu-se agregar conhecimento de outros artigos. Para isso, aplicou-se a fórmula de Chendong (Equação 1) como um atributo para o cálculo do GP real. A fórmula de Chendong possui uma razoável precisão para previsão do GP a partir das medições de 2-fal. Todavia, ainda assim, possui deficiências, principalmente por considerar apenas 2-fal.

$$GP = \frac{1,51 - \log(2fal)}{0,0035} \quad (1)$$

Desse modo, para esse terceiro modelo, os atributos de entrada considerados foram a medição de 2-fal em ppb, a medição da relação  $CO_2/CO$  e o GP calculado a partir da fórmula de Chendong. Os resultados são apresentados nas Figura 12 e Figura 13.

O resultado obtido foi significativamente melhor, estando bem mais próximo tanto para o treinamento quanto para o teste. A Figura 14 mostra os resultados apenas da equação de Chendong. Esses resultados podem também ser verifi-

cados na Tabela 8. O erro médio caiu significativamente do modelo 2 para o modelo 3, indicando que a adição do atributo Chendong agregou conhecimento útil ao modelo.

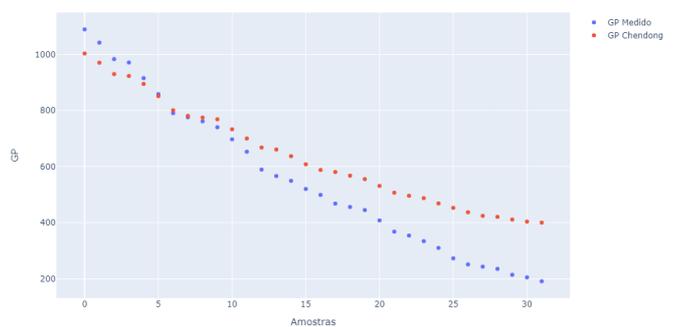


Figura 14. Relação entre GP real e GP Chendong

### 2.3.4. Modelo de previsão do GP a partir do 2-fal, da relação $CO_2/CO$ , da fórmula de Chendong e da potência do Transformador

Por fim, o último atributo adicionado ao modelo foi a potência do transformador. Por mais que esta se mostrasse

Tabela 8. Métricas do modelo com 2-fal, relação CO<sub>2</sub>/CO e GP calculado (Chendong)

	R2	MAE
Treinamento	0,99	5,84
Teste	0,99	8,37

sem correlação com o GP, os resultados obtidos mostraram que ela contém informação relevante para a previsão. A Tabela 9 apresenta esses resultados.

Tabela 9. Métricas do modelo com 2-fal, relação CO<sub>2</sub>/CO, GP calculado (Chendong) e a Potência do Transformador

	R2	MAE
Treinamento	0,99	5,88
Teste	0,99	6,01

Percebe-se que o erro médio caiu ainda mais no teste. Todavia, como o último valor já estava suficientemente baixo, a diferença é quase imperceptível a nível de gráfico. A distribuição estatística do erro pode ser vista na Figura 15, e é levemente deslocada para a direita, fato reforçado pelo valor de MAE positivo.

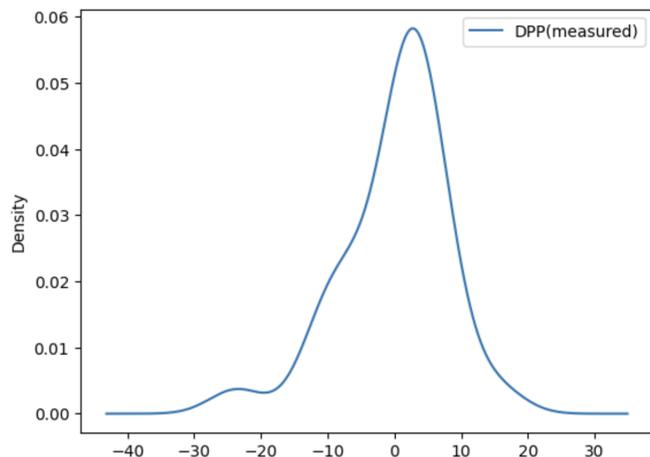


Figura 15. Distribuição de erro do modelo com os dados de Chendong

### 3. CONCLUSÃO

Consequente ao envelhecimento da isolação sólida do transformador são gerados diversos indicadores, como o 2-fal e CO<sub>2</sub>/CO. O poder preditivo de cada um dos indicadores isoladamente é insuficiente para a criação de um modelo preciso que se ajuste aos diversos transformadores existentes. Porém, quando tratados de forma unificada, esses dois indicadores compõem um modelo robusto com capacidade de prever os valores de GP de modo mais preciso.

Modelos de aprendizagem de máquina necessitam de uma grande quantidade de dados para funcionarem adequadamente, e por isso, foi também adotada a equação de Chendong como entrada para o modelo, contornando a falta de dados e possibilitando uma melhor precisão. Percebeu-se que a informação de Potência do Transformador, por mais que não possua alta correlação com o valor de GP,

apresenta conhecimento importante para o cálculo do envelhecimento utilizando o GP, auxiliando na diminuição do erro médio absoluto.

Apesar do modelo proposto não ter sido validado com um caso real distinto aos dados de teste, ele é expansível, e pode ser retreinado assim que novos dados estejam disponíveis. O modelo atual serve de ferramenta para auxílio aos especialistas para uma melhor estimativa da vida útil dos transformadores utilizando o cálculo de GP.

### 4. ORCID IDS DOS AUTORES

Rafael Prux Fehlberg ; Daniel Carrijo Polonio Araujo ; Gabriel de Souza Pereira Gomes ; Sofia Moreira de Andrade Lopes ; Rogério Andrade Flauzino ; Renan Ferreira Santa Rosa .

### REFERÊNCIAS

- Aciu, A.M., Nicola, C.I., Nicola, M., and Nițu, M.C. (2021). Complementary analysis for dga based on duval methods and furan compounds using artificial neural networks. *Energies*, 14(3), 588.
- Allan, D. (1993). Practical life-assessment technique for aged transformer insulation. *IEEE Proceedings A (Science, Measurement and Technology)*, 140(5), 404–408.
- Assuncao, T.C.B.N. (2007). *Contribuição à modelagem e análise do envelhecimento de transformadores de potência*. Ph.D. thesis, Universidade Federal de Minas Gerais.
- Elele, U., Nekahi, A., Arshad, A., and Fofana, I. (2022). Towards online ageing detection in transformer oil: A review. *Sensors*, 22(20), 7923.
- Münster, T., Kinkeldey, T., Werle, P., Hämel, K., and Preusel, J. (2018). Investigation on ageing parameters of a thermally accelerated aged paper-oil-insulation in a hermetically sealed system. In *2018 Condition Monitoring and Diagnosis (CMD)*, 1–6. doi:10.1109/CMD.2018.8535626.
- Oria, C., Ortiz, A., Ferreño, D., Carrascal, I., and Fernández, I. (2019). State-of-the-art review on the performance of cellulosic dielectric materials in power transformers: Mechanical response and ageing. *IEEE Transactions on Dielectrics and Electrical Insulation*, 26(3), 939–954.
- Prevost, T.A. (2005). Thermally upgraded insulation in transformers. In *Proceedings Electrical Insulation Conference and Electrical Manufacturing Expo, 2005.*, 120–125. IEEE.
- Saha, T.K. and Purkait, P. (2017). *Transformer ageing: monitoring and estimation techniques*. John Wiley & Sons.
- Shutenko, O. and Ponomarenko, S. (2022). Analysis of ageing characteristics of transformer oils under long-term operation conditions. *Iranian Journal of Science and Technology, Transactions of Electrical Engineering*, 46(2), 481–501.
- Teymouri, A. and Vahidi, B. (2017). Co<sub>2</sub>/co concentration ratio: A complementary method for determining the degree of polymerization of power transformer paper insulation. *IEEE Electrical Insulation Magazine*, 33(1), 24–30. doi:10.1109/MEI.2017.7804313.