



UNIVERSIDADE FEDERAL DO CEARÁ
CENTRO DE TECNOLOGIA
DEPARTAMENTO DE ENGENHARIA ELÉTRICA

GLADSON RENATO QUEIROZ VIDAL

**USO DE MEIOS COMPUTACIONAIS NA DETECÇÃO DE
FALHAS INCIPIENTES EM TRANSFORMADORES DE POTÊNCIA COMO MÉTODO
ALTERNATIVO À ANÁLISE DE GASES DISSOLVIDOS**

FORTALEZA

2022

GLADSON RENATO QUEIROZ VIDAL

USO DE MEIOS COMPUTACIONAIS NA DETECÇÃO DE
FALHAS INCIPIENTES EM TRANSFORMADORES DE POTÊNCIA COMO MÉTODO
ALTERNATIVO À ANÁLISE DE GASES DISSOLVIDOS

Trabalho de conclusão de curso apresentado a coordenação do curso de engenharia elétrica da Universidade Federal do Ceará, como requisito parcial à obtenção do título de bacharel em engenharia elétrica.

Orientador: Prof. Dr. Fernando Marcelo Antunes.

FORTALEZA

2022

Dados Internacionais de Catalogação na Publicação
Universidade Federal do Ceará
Sistema de Bibliotecas
Gerada automaticamente pelo módulo Catalog, mediante os dados fornecidos pelo(a) autor(a)

V691u Vidal, Gladson Renato Queiroz.

Uso de meios computacionais na detecção de falhas incipientes em transformadores de potência como método alternativo à análise de gases dissolvido / Gladson Renato Queiroz Vidal. – 2023.
50 f. : il. color.

Trabalho de Conclusão de Curso (graduação) – Universidade Federal do Ceará, Centro de Tecnologia, Curso de Engenharia Elétrica, Fortaleza, 2023.

Orientação: Prof. Dr. Fernando Marcelo Antunes.

1. Machine Learning. 2. Falhas incipientes. 3. Transformadores. I. Título.

CDD 621.3

GLADSON RENATO QUEIROZ VIDAL

USO DE MEIOS COMPUTACIONAIS NA DETECÇÃO DE
FALHAS INCIPIENTES EM TRANSFORMADORES DE POTÊNCIA COMO MÉTODO
ALTERNATIVO À ANÁLISE DE GASES DISSOLVIDOS

Trabalho de conclusão de curso apresentado a coordenação do curso de engenharia elétrica da Universidade Federal do Ceará, como requisito parcial à obtenção do título de bacharel em engenharia elétrica.

Aprovada em: 09/12/2022.

BANCA EXAMINADORA

Prof. Dr. Fernando Marcelo Antunes (Orientador)
Universidade Federal do Ceará (UFC)

Msc. Anderson Jhones Passos Nascimento
Universidade Federal do Ceará (UFC)

Eng. Gabriel Marçal da Cunha Pereira Carvalho
Universidade Federal do Ceará (UFC)

Aos meus pais, Maria Veronildes Queiroz Vidal
e Gleidson Vicente Fernandes Vidal.

AGRADECIMENTOS

Pensei muitas vezes sobre como iniciar esse texto e acho que me faltam palavras para saber como encerrá-lo... Provavelmente faltarão muitas pessoas nesses agradecimentos, pois é impossível resumir toda uma trajetória de 5 anos em apenas uma página. A todos que de alguma forma contribuíram para o meu desenvolvimento acadêmico, profissional e pessoal e que por algum motivo não citei aqui, deixo o meu mais sincero muito obrigado. Porém existem aqueles que contribuíram ativamente na caminhada que tem sido esses últimos meses e a estes não poderia deixar de nomear toda a gratidão que sinto.

Maria Elisa, Pâmela, Isabelle, Morghana e Gabe, agradeço por estarem a tanto tempo na minha vida, que hoje preciso parar e fazer contas sempre que tento explicar para alguém como algumas pessoas aparecem na sua vida e simplesmente ficam, sem que haja uma cobrança, ou mesmo um motivo que não e simplesmente o amor. Mesmo que estejamos fisicamente separados (grande parte do tempo), sei que sempre posso contar com vocês... e saibam que sempre podem contar comigo.

Larissa, Dani e Ana Virgília, nem o modelo estatístico mais realista conseguiu prever a pandemia, acho que ele também não conseguiria prever nossa amizade, foram muitos encontros e desencontros e hoje só posso dizer que amo vocês de graça (perdão por furar quase todos os papéis, juro que estou melhorando).

Ilana e Juliana, nossa amizade era previsível, lugar certo, hora certa, pessoas certas... Mas a parte que a previsibilidade não diz é como essa relação evoluiria a partir dali. Ela também não fala como será daqui pra frente mas espero que continuemos existindo nos próximos capítulos uns dos outros.

Davi Mendes, você foi uma das minhas principais inspirações enquanto liderança e hoje posso lhe chamar de amigo! Compartilhamos sonhos, choramos juntos, rimos juntos, crescemos juntos... Ansioso para o que vem depois.

Davi Veras, obrigado por ouvir todos os meus surtos. Não te odeio! (mesmo que você duvide disso), na verdade te considero um amigo que não esperava que fosse ser.

Por fim, não poderia deixar de agradecer também ao Professor Fernando Antunes, uma das pessoas que acreditou desde o primeiro momento que esse trabalho seria possível. E a UFC por proporcionar os anos mais loucos e talvez os melhores anos da minha vida.

“We can only see a short distance ahead, but we can see plenty there that needs to be done.”
(TURING, 1950, p. 460).

RESUMO

Os transformadores de potência são equipamentos essenciais para a operação do sistema elétrico, tanto no sentido econômico quanto no técnico. Diante disso, qualquer esforço despendido na identificação de falhas torna-se relevante. Este trabalho objetiva identificar os métodos computacionais com melhor desempenho na classificação de falhas incipientes em transformadores com base na concentração dos gases dissolvidos em óleo. Para tanto, foi utilizada uma base com 201 registros de análises cromatográficas (ou análise de gases dissolvidos) disponibilizadas no *IEEE DATAPORT*. Os dados foram tratados, segregados em dois conjuntos (um contendo o *dataset* original normalizado e outro contendo as razões entre os gases mediante metodologia definida por Rogers) e submetidos a classificação por método tradicional e computacionais (*K-nearest neighbor*, *Naïve bayes* e *Multilayer perceptron*). Para o *dataset* original, o *Multilayer perceptron* obteve precisão de 0,80, o *K-nearest neighbor* 0,60 e o *Naïve bayes* 0,36. Por sua vez, para o *dataset* das razões, o *Multilayer perceptron* obteve uma precisão menor (0,70), mas tanto o *K-nearest neighbor*, quanto o *Naïve bayes* melhoraram seu desempenho alcançando os valores de 0,70 e 0,63, respectivamente. É possível concluir que usar o método das razões no pré-processamento dos dados da análise dos gases dissolvidos, melhorou o desempenho de dois classificadores (*K-nearest neighbor* e *Naïve bayes*) e piorou o desempenho de um deles (*Multilayer perceptron*), além disso, dos classificadores implementados, o *Multilayer perceptron* e o *K-nearest neighbor*, obtiveram taxas de acerto superiores, respectivamente 0,80 e 0,70, ao método tradicional descrito por Rogers (0,69), sendo este melhor apenas que o *Naïve Bayes* (0,67).

Palavras-chave: *Machine Learning*; Falhas incipientes; Transformadores.

ABSTRACT

Power transformers are essential equipment for the operation of the electrical system, both economically and technically. In view of this, any effort devoted to identifying failures becomes relevant. This work aims to identify the computational methods with better performance in the classification of incipient failures in transformers based on the concentration of gases dissolved in oil. For this purpose, a database with 201 records of chromatographic analyzes (or dissolved gas analysis) available on the IEEE DATAPORT was used. The data were treated, segregated into two sets (one containing the original normalized dataset and the other containing the ratios between gases using a methodology defined by Rogers) and submitted to classification using traditional and computational methods (K-nearest neighbor, Naïve bayes and Multilayer perceptron). For the original dataset, the Multilayer perceptron obtained a precision of 0.80, the K-nearest neighbor 0.60 and the Naïve bayes 0.36. In turn, for the ratios dataset, the Multilayer perceptron obtained a lower precision (0.70), but both the K-nearest neighbor and the Naïve bayes improved their performance reaching values of 0.70 and 0.63, respectively. It is possible to conclude that using the method of ratios in the pre-processing of data from the analysis of dissolved gases, improved the performance of two classifiers (K-nearest neighbor and Naïve bayes) and worsened the performance of one of them (Multilayer perceptron), in addition, of the implemented classifiers, the Multilayer perceptron and the K-nearest neighbor, obtained higher success rates, respectively 0.80 and 0.70, than the traditional method described by Rogers (0.69), which is only better than Naïve Bayes (0.67).

Keywords: *Machine Learning*; Incipient Failure; Power transformers.

LISTA DE FIGURAS

Figura 01	Transformador Ideal com núcleo de ferro	16
Figura 02	Representação gráfica do método de Duval	22
Figura 03	Interpretação gráfica do método de Dörnenburg.....	23
Figura 04	<i>K-nearest neighbor</i> para k igual a quatro com p (características) igual a dois.	29
Figura 05	Ilustração do <i>Naïve Bayes</i>	30
Figura 06	Ilustração da rede <i>Perceptron</i>	31
Figura 07	Ilustração da fronteira de separação (neurônio com duas entradas).....	32
Figura 08	Ilustração de uma rede MLP.....	33

LISTA DE GRÁFICOS

Gráfico 01	Precisão do classificador pelo valor de η para o <i>dataset</i> original.....	38
Gráfico 02	Precisão do classificador pelo valor de η para o <i>dataset</i> das razões.....	39
Gráfico 03	Classificadores computacionais contra método de Rogers.....	47

LISTA DE TABELAS

Tabela 01	Razões entre as concentrações dos gases	21
Tabela 02	Gás chave e falhas associadas.....	21
Tabela 03	Faixas das razões definidas no método de Dörnenburg.....	22
Tabela 04	Faixas das razões definidas no método de Rogers.....	24
Tabela 05	Quantidade de registros por tipo de falha.....	36
Tabela 06	Métricas do <i>K-nearest neighbor</i> para <i>dataset</i> original.....	42
Tabela 07	Métricas do <i>Multilayer perceptron</i> para <i>dataset</i> original.....	42
Tabela 08	Métricas do <i>Naïve Bayes</i> para <i>dataset</i> original.....	43
Tabela 09	Acurácia para cada classificador aplicado ao <i>dataset</i> original.....	43
Tabela 10	Métricas do <i>K-nearest neighbor</i> para <i>dataset</i> das razões.....	44
Tabela 11	Métricas do <i>Multilayer perceptron</i> para <i>dataset</i> das razões.....	44
Tabela 12	Métricas do <i>Naïve bayes</i> para <i>dataset</i> das razões.....	45
Tabela 13	Acurácia para cada classificador aplicado ao <i>dataset</i> das razões.....	45
Tabela 14	Acurácia dos classificadores computacionais contra método de Rogers.....	46

LISTA DE ABREVIATURAS E SIGLAS

ABNT Associação Brasileira de Normas Técnicas

MLP *Multilayer perceptron*

KNN *K-nearest neighbor*

NB *Naïve Bayes*

DGA Análise de gases dissolvidos

RNA Redes Neurais Artificiais

ML *Machine Learning*

SUMÁRIO

1	INTRODUÇÃO	14
2	REVISÃO BIBLIOGRÁFICA	15
2.1	Transformadores de potência e falhas incipientes	15
2.2	Análise dos gases dissolvidos (DGA).....	19
2.2.1	<i>Método do gás chave</i>	21
2.2.2	<i>Triângulo de Duval (Duval's Triangle)</i>	21
2.2.3	<i>Razões de Dörnenburg</i>	22
2.2.4	<i>Método das Razões de Rogers (Rogers Ratio Method)</i>	23
2.3	Limitações das técnicas convencionais.....	24
2.4	Inteligências Computacionais	24
2.5	Construção do Conhecimento	25
2.6	<i>Machine Learning</i> e Técnicas de Classificação	26
2.6.1	<i>k-Nearest neighbor</i>	27
2.6.2	<i>Naïve-Bayes</i>	29
2.6.3	<i>Multilayer perceptron (MLP)</i>	30
3	PROPOSTA DE PESQUISA	35
3.1	Conjunto de dados avaliado	35
3.2	Linguagem de Programação e Bibliotecas utilizadas	36
3.3	Classificadores e parâmetros utilizados	36
3.3.1	<i>Parâmetros do classificador K-Nearest Neighbor (K vizinhos mais próximos)</i>	36
3.3.2	<i>Parâmetros do classificador Multilayer Perceptron (Perceptron multicamadas)</i>	37
3.3.3	<i>Parâmetros do classificador Gaussian Naïve Bayes</i>	39
3.4	Aplicação dos Classificadores	39
3.5	Classificação por métodos convencionais.....	40
4	RESULTADOS E DISCUSSÕES	41
4.1	Execução 1: <i>Dataset</i> original	41
4.2	Execução 2: <i>Dataset</i> das razões	44
4.3	Comparação entre os métodos computacionais e método tradicional.....	46
5	CONCLUSÕES	48
	REFERÊNCIAS	49

1 INTRODUÇÃO

Os transformadores de potência, possuem grande importância no sistema elétrico, por se tratarem de equipamentos elevadores e abaixadores dos níveis de tensões e correntes, fato que é extremamente relevante nos sistemas de transmissão e distribuição de energia. Diante disso, sua manutenção e antecipação a falhas tornam-se temas essenciais no mercado e na academia, o que acaba por motivar a busca por avaliação dos sistemas de isolamento desses equipamentos com o intuito de obter diagnósticos quanto a presença de falhas ainda na fase inicial (BARBOSA, 2013).

A análise de gases dissolvidos do inglês *dissolved gas analysis* (DGA), também conhecida como análise cromatográfica, vem sendo utilizada ao longo dos anos como forma de identificar falhas incipientes em transformadores de potência. Tal avaliação é comparada constantemente a um exame sanguíneo devido a similaridades informadas entre os resultados. Enquanto um exame de sangue auxilia médicos na verificação de situações de ameaças a saúde humana, a análise do óleo isolante pode ajudar a tomada de decisão quanto à sua manutenção e operação (DING et. al., 2011).

Existem diferentes técnicas com o intuito de interpretar os resultados das análises cromatográficas, sendo os mais conhecidos: Gás chave (PUGH; WAGNER, 1961, IEEE, 2008, triângulo de Duval (DUVAL, 1989), razões de Dörnenburg (DÖRNENBURG, GERBER, 1967) e Rogers (ROGERS, 1978). Porém, tais métodos têm como característica comum o problema da não-decisão, uma vez que as normas que os apresentam, possuem apenas algumas combinações típicas das razões entre os gases dissolvidos no óleo isolante. Dessa forma, naturalmente observa-se uma tendência à substituição dos diagnósticos feitos por meio de testes convencionais por aqueles realizados de maneira automática incorporando técnicas computacionais no processo de diagnóstico (MORAIS; ROLIM, 2006).

Muitos trabalhos, por exemplo, tem proposto a utilização de Redes Neurais Artificiais para estabelecer o diagnóstico de falhas incipientes e obtido resultados promissores advindos da capacidade da RNA aprender as conexões escondidas entre as concentrações de gases dissolvidos e os tipos de falhas (BARBOSA, 2013).

Diante disso, este trabalho objetiva avaliar diferentes métodos computacionais, mais especificamente, *k-nearest neighbors*, *Naïve Bayes* e *Multilayer Perceptron*, na identificação de falhas incipientes em transformadores de potência, através das informações de gases dissolvidos no óleo isolante. Identificando, qual o melhor modelo e sob quais condições produzem-se as melhores taxas de acerto.

2 REVISÃO BIBLIOGRÁFICA

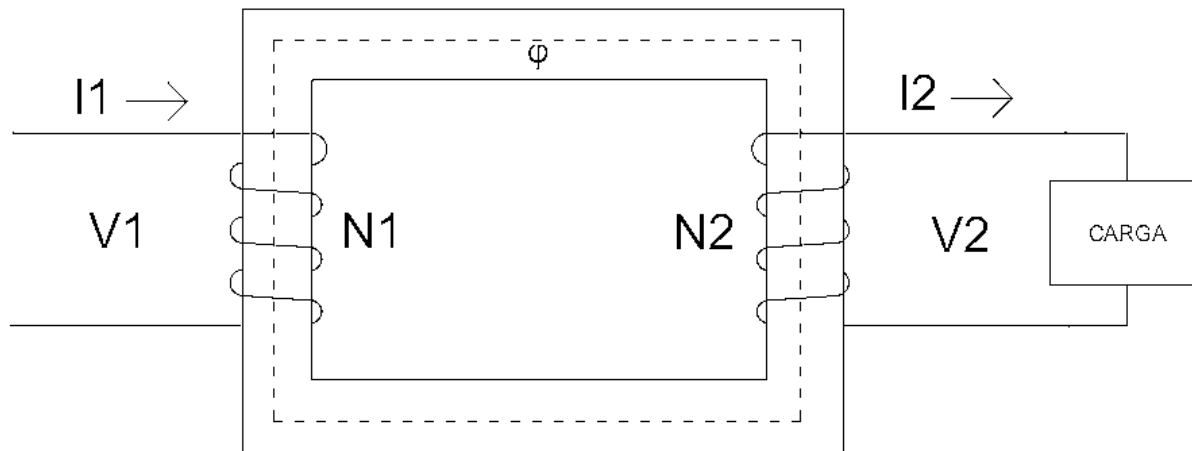
Neste capítulo apresenta-se a base teórica, que servirá de embasamento para as questões abordadas nesta pesquisa. Nele encontram-se apresentados os conceitos de transformador, falhas incipientes, metodologias convencionais de análise de gases dissolvidos, *machine learning* e técnicas de classificação.

2.1 Transformadores de potência e falhas incipientes

O Transformador é um equipamento utilizado para converter diferentes níveis de tensão entre si, sem alterar a magnitude da potência entregue a carga, dessa forma, garante fornecimento de energia elétrica e preserva a segurança operacional do processo. Ele possui uso extensivo em hidrelétricas, sendo utilizado para elevar o nível de tensão que estas fornecem e reduzir o nível de corrente, o que resulta em diminuição de perdas nos condutores, tornando possível transmissão de energia por longas distâncias; em cidades, para aumentar o nível de corrente e diminuir o nível de tensão, medida necessária para garantir a segurança dos habitantes e viabilizar o uso de eletrônicos de baixa tensão e em indústrias, onde seu uso está associado a adequação dos níveis de tensão e corrente para diferentes tipos de equipamentos que possuem necessidades específicas para sua atuação (FONSECA, 2014).

O seu funcionamento está associado ao princípio de indução mútua entre duas (ou mais) bobinas ou circuitos que encontram-se indutivamente acoplados, portanto seus circuitos não são ligados fisicamente, ou seja, não há conexão condutiva entre eles (KOSOW,1982), como observamos na figura abaixo.

Figura 01 – Transformador Ideal com núcleo de ferro



Fonte: elaborada pelo autor

Em que:

I_1 = Corrente primária;

I_2 = Corrente secundária;

ϕ = Fluxo magnético;

V_1 = Tensão no Primário;

V_2 = Tensão no Secundário;

N_1 = Número de espiras no primário;

N_2 = Número de espiras no secundário;

Quando aplicamos uma tensão elétrica V_1 (instantaneamente positiva), a direção da corrente primária I_1 produz a direção do fluxo magnético ϕ . É induzida uma força eletromotriz no primário, que de acordo com a convenção dos pontos e com a lei de Lenz, produz uma polaridade positiva na parte superior da bobina primária que se opõe instantaneamente a tensão aplicada V_1 . De maneira semelhante, no secundário, para a direção de ϕ especificada (sentido horário), a polaridade positiva da força eletromotriz deve obrigatoriamente produzir um fluxo desmagnetizante oposto a ϕ . Uma carga ligada aos terminais do secundário produz uma corrente I_2 , que circula em resposta à polaridade da força eletromotriz no secundário e produz um fluxo desmagnetizante (KOSOW,1982).

Da relação estabelecida, determina-se uma igualdade entre as forças magnetomotrizes do primário e secundário do transformador:

$$FMM_1 = FMM_2$$

$$I_1' \cdot N_1 = I_2 \cdot N_2$$

$$\frac{I_2}{I_1'} = \frac{N_1}{N_2} = \alpha$$

Onde α é a relação das espiras primárias para as secundárias também conhecida como relação de transformação. A importância de α decorre do fato que ela é fixa (não constante), para qualquer transformador dado (já construído). Com isso, a componente de carga da corrente primária pode ser calculada para qualquer valor da corrente de carga secundária. Além disso, é possível estender, por meio da quantificação de Neumann da lei de Faraday, a noção de relação de transformação também para as tensões (KOSOW,1982). Dessa forma, temos que:

$$\alpha = \frac{I_2}{I_1'} = \frac{N_1}{N_2} = \frac{V_1}{V_2}$$

Para um transformador ideal, sem perdas, não existindo fluxos dispersos primários nem secundários (reatâncias de dispersão nulas), podemos dizer que:

$$V_1 \cdot I_1 = V_2 \cdot I_2$$

Esta equação comprova a definição fundamental de um transformador como um dispositivo que transfere energia de um circuito para outro.

Para Georgilakis (2009) os transformadores podem ser classificados de acordo com dois critérios primordiais:

1. Classificação de transformadores quanto ao uso:

- a) Transformadores de distribuição: utilizados em redes de distribuição para transmitir energia de uma rede de média tensão para uma rede de baixa tensão. Sua potência nominal está geralmente entre 50 e 1600 kVA.
- b) Transformadores de potência: utilizados em usinas de geração para elevar a tensão e em subestações de transmissão para aumento ou rebaixamento de tensão. Em sua maioria apresentam potência maior do que 2 MVA.

- c) Autotransformadores: utilizados para limites pequenos de transformação de tensão, para conexões com sistemas de energia elétrica de diferentes tensões, para a partida de motores de corrente alternada, etc.
- d) Transformadores de teste: utilizados para a realização de testes de desempenho em alta e extra-alta tensão.
- e) Transformadores de potência especiais: utilizados em aplicações especiais
- f) Transformadores de instrumentação: utilizados para realizar medições precisas de tensão ou de corrente.
- g) Transformadores de telecomunicações: utilizados em aplicações de telecomunicações com o intuito de garantir uma melhor reprodução do sinal em uma larga faixa de frequências e tensões.

2. Classificação de transformadores quanto ao meio de isolamento:

- a) Transformadores isolados a óleo: o isolamento é feito com óleo mineral ou óleo sintético.
- b) Transformadores isolados a seco: o resfriamento é implementado por meio da circulação natural de ar e os enrolamentos são isolados com materiais específicos.
- c) Transformadores isolados com resina: o transformador com isolamento de resina de epoxy é um transformador isolado a seco com resina epoxy fundida a vácuo

Um componente básico da grande maioria dos transformadores de potência é o óleo, que promove a refrigeração e o isolamento dos circuitos elétricos e magnéticos. A popularidade do óleo mineral, oriundo do petróleo se deve, nesta aplicação, à disponibilidade, ao baixo custo e às suas excelentes propriedades dielétricas e refrigerantes (GEORGILAKIS, 2011). Ademais, as características químicas do óleo são os principais indicativos de falhas no

seu funcionamento, podendo estes serem determinados via análise dos gases dissolvidos também conhecida como DGA (ZENG, GUO, ZHU, et al., 2019).

As principais falhas associadas aos transformadores de potência, possuem como origem o envelhecimento do equipamento ou como resultado da incidência de curto-circuitos e sobrecargas, devido o regime de trabalho a que estes são expostos (BENGTSON, 1996; BETTA et al., 2000; KOVACEVIC; DOMINELLI, 2003; TENBOHLEN; FIGEL, 2000).

Podemos associar as falhas incipientes em transformadores aos esforços térmicos, em que os sobreaquecimentos são os principais agentes causadores e aos esforços elétricos, que estão relacionados às descargas elétricas no interior do equipamento. As falhas com características térmicas advêm de correntes elétricas que circulam o equipamento e das trocas térmicas com a temperatura ambiente. Os esforços de origem elétrica correspondem a tensão e a frequência de operação. É importante ressaltar que esses eventos podem ocorrer de maneira isolada ou simultânea. Somado a isto, há possibilidade de aparecimento de bolhas no sistema de isolamento líquido, devido à presença de umidade originada de contaminações do processo industrial e impurezas no material, causando assim uma diminuição da capacidade elétrica do isolante (AZEVEDO, 2009).

2.2 Análise dos gases dissolvidos (DGA)

Os transformadores de potência envolvidos em situações de falhas, apresentam dissolvidos em seu óleo isolante alguns gases: hidrogênio (H_2), monóxido de carbono (CO), dióxido de carbono (CO_2), metano (CH_4), etileno (C_2H_4), acetileno (C_2H_2) e etano (C_2H_6). Também podem ser encontrados alguns hidrocarbonetos como C_3 e C_4 , além de gases não-combustíveis como o oxigênio (O_2) e o nitrogênio (N_2), porém estes não influenciam no diagnóstico de falhas, por não serem significativos no processo de decomposição do óleo (ABNT, 1982; MILASCH, 1984).

Cada gás é gerado como resultado de um processo físico-químico, em sua maioria indesejado, em decorrência do funcionamento do equipamento.

Quando submetidos a faixa de temperatura de 150 °C a 500 °C, entendidas como baixas temperaturas, o óleo mineral sofre uma decomposição em gases de baixo peso molecular, como hidrogênio (H_2) e metano (CH_4) e pequenos traços de gases pesados, como etileno (C_2H_4) e etano (C_2H_6). Conforme a temperatura aumenta, na região da falha, ocorre um aumento da produção de gases pesados. De início, a concentração do etano se eleva mais rapidamente sendo seguido pelo etileno. Porém, quando a falha causa um aumento

exacerbado da temperatura, nota-se a produção do acetileno (C_2H_2). A presença desse gás, indica aos especialistas uma condição extrema no equipamento, considerando este como o último nível de alarme em termos de falhas incipientes (BARBOSA, 2008; IEEE, 2008).

Quando ocorrem descargas elétricas e arcos intermitentes de baixa energia, observa-se uma elevada produção de hidrogênio (H_2), com uma pequena concentração de metano (CH_4) e acetileno (C_2H_2). Conforme aumenta-se a intensidade das descargas de energia, e como consequência aumenta-se também a temperatura, observa-se crescer a produção de etileno (C_2H_4) e acetileno (C_2H_2). Quando a concentração de gases dissolvidos é alta e ocorre uma alteração na temperatura, observa-se uma mudança no valor da solubilidade de saturação do óleo, fazendo com que os gases, antes dissolvidos, sejam liberados na forma de bolhas. A existência dessas bolhas, altera a condição de rigidez elétrica, o que leva a uma ruptura do isolamento (BARBOSA, 2013).

Se as descargas elétricas atingirem valores de alta energia, como arcos elétricos, a produção de acetileno (C_2H_2) torna-se relevante, pois a temperatura se eleva entre 800 °C e 2800 °C. As possíveis causas da formação destes pontos quentes são: centelhamentos, ocasionados pela fragilidade nas conexões ou interrupções de corrente no comutador; descargas de contorno ou arcos de potência propagados pela ruptura dielétrica do óleo entre os enrolamentos, espiras ou entre espiras e massa (BARBOSA,2008).

Por refletir os acontecimentos no interior do transformador e por incidir um processo que é capaz de levar a falhas severas, como curto-circuitos internos, a DGA no óleo é extremamente importante, tanto para a verificação das condições de funcionamento bem como para a manutenção do equipamento (BARBOSA,2013).

Há décadas as técnicas de análise são aplicadas à nível mundial como etapa do processo de manutenção, sendo padronizadas por normas nacionais e internacionais, como ABNT NBR 7070 (ABNT, 1981), IEC 599 (IEC, 1999) e IEEE C57.104 (IEEE, 2008). A base de sua aplicação consiste em submeter uma amostra do óleo isolante à cromatografia em fase gasosa. A cromatografia objetiva quantificar as concentrações dos gases dissolvidos no óleo. Os dados obtidos podem ser expressos por meio das concentrações individuais de cada gás, através de razões entre as concentrações dos gases ou como porcentagens em relação à concentração total de gases combustíveis. A grande maioria das teorias de DGA baseiam-se no uso das razões, presentes na tabela 01, das concentrações dos gases dissolvidos no óleo (BARBOSA, 2013).

Tabela 01 – Razões entre as concentrações dos gases

Notação	R1	R2	R3	R4	R5
Razão	CH_4/H_2	C_2H_2/C_2H_4	C_2H_2/CH_4	C_2H_6/C_2H_2	C_2H_4/C_2H_6

Fonte: IEEE (2008)

Dentre os métodos mais conhecidos para a interpretação dos resultados cromatográficos estão: o método do gás chave (PUGH; WAGNER, 1961, IEEE, 2008), o triângulo de Duval (DUVAL, 1989) e os métodos das razões de Dörnenburg (DÖRNENBURG, GERBER, 1967) e Rogers (ROGERS, 1978).

2.2.1 Método do gás chave

O princípio do método consiste em estabelecer uma relação entre o gás dissolvido no óleo com a temperatura de sua geração. Uma vez correlacionados, determinam-se as falhas correspondentes ao gás mais predominante, este é chamado de “gás chave”. Dessa forma, para cada gás temos uma falha, como podemos observar na tabela 02:

Tabela 02 – Gás chave e falhas associadas

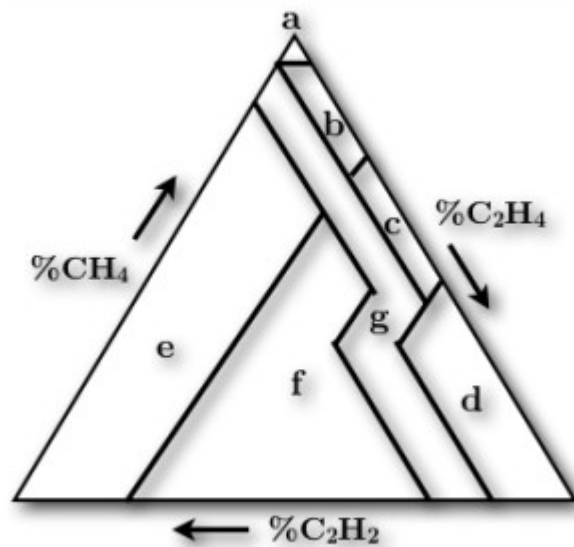
Gás Chave	Falha Associada
Etileno	Falha térmica no óleo
Hidrogênio	Falha elétrica de baixa Energia ou corona
Acetileno	Falha elétrica de alta energia ou arco elétrico

Fonte: Elaboração Própria

2.2.2 Triângulo de Duval (Duval's Triangle)

A aplicação desse método necessita do cálculo dos percentuais do metano (CH_4), etileno (C_2H_4) e acetileno (C_2H_2) em relação a soma das concentrações destes gases gerados em p.p.m. ($CH_4+C_2H_4+C_2H_2$). Os percentuais são então usados como coordenadas em um sistema triangular de mapeamento das falhas, como observamos na figura 02.

Figura 02 – Representação gráfica do método de Duval



Fonte: BARBOSA, 2013

Em que, a representa: Descargas parciais; b: Falhas térmicas de baixa temperatura (temperatura < 300°C); c: Falhas térmicas de média temperatura (300°C < temperatura < 700°C); d: Falhas térmicas de alta temperatura (temperatura > 700°C); e: Descargas elétricas de baixa energia; f: Descargas elétricas de alta energia; g: Indeterminada (falha térmica ou elétrica).

2.2.3 Razões de Dörnenburg

O método de Dörnenburg especifica três tipos gerais de falhas, sendo estas: sobreaquecimento, descargas de baixa energia (corona) e descargas de alta energia (arcos elétricos). A categorização utiliza-se das razões R1, R2, R3 e R4, definidas anteriormente na tabela 01. Na tabela 03, observamos as faixas das razões definidas para o método.

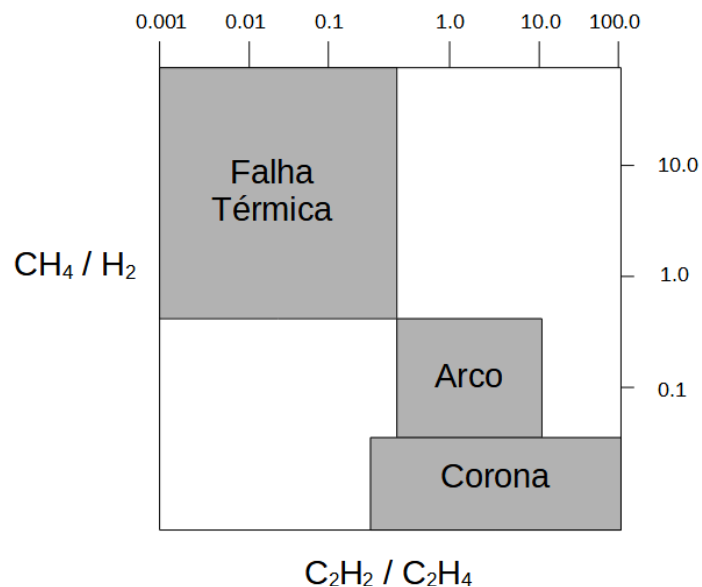
Tabela 03 – Faixas das razões definidas no método de Dörnenburg.

Falha Incipiente	R1	R2	R3	R4
Decomposição Térmica	>1,0	<0,75	<0,3	>0,4
Corona (descargas de baixa energia)	<0,1	-	<0,3	>0,4
Arcos Elétricos(descargas de alta energia)	>0,1 e <1,0	>0,75	>0,3	<0,4

Fonte: Adaptado de IEEE (2008)

A Figura 03 é uma interpretação gráfica do método. Nela é extremamente perceptível regiões, onde não é definida nenhuma associação com qualquer diagnóstico de falha. Dessa forma, para os casos onde os valores das razões levem a estas regiões, o diagnóstico não poderá ser realizado. Este problema é conhecido por problema de não-decisão. O problema de não-decisão é comum nos métodos convencionais, pois as normas apresentam apenas certas combinações típicas das razões. Quando um caso se apresenta fora dessas combinações, outros métodos são necessários para que se possa determinar a falha.

Figura 03 – Interpretação gráfica do método de Dörnenburg



Fone: Adaptado de LIMA(2005)

2.2.4 Método das Razões de Rogers (Rogers Ratio Method)

De forma semelhante ao método de Dörnenburg, o método de Rogers faz uso das razões entre os gases para estabelecer um diagnóstico de falhas. Originalmente o método utiliza-se das razões R1, R2 e R5 e C_2H_6/CH_4 , porém após sua revisão no IEEE C57.104, no IEC 599 e na ABNT NBR-7274 houve a exclusão da razão C_2H_6/CH_4 , que não era utilizada na identificação de falhas. É a técnica mais utilizada em empresas concessionárias de energia (GOMEZ, 2013). Porém, quando há ocorrência simultânea de falhas, o diagnóstico pode gerar ambiguidade, pois as faixas de variação das relações se sobrepõem, gerando assim, mais de uma interpretação possível (zonas de não-decisão) (GOMEZ,2013; SENNA,2010). A tabela 4, a seguir mostra as diferentes faixas de razões entre gases e a falha associada.

Tabela 04 – Faixas das razões definidas no método de Rogers.

Falha Incipiente	R1	R2	R5
Condição Normal	>0,1 a <1,0	<0,1	<1,0
Corona (descargas de baixa energia)	<0,1	<0,1	<1,0
Arcos Elétricos(descargas de alta energia)	0,1 a 1,0	0,1 a 0,3	>3,0
Falha Térmica $T < 300\text{ }^{\circ}\text{C}$	>0,1 a <1,0	<0,1	0,1 a 3,0
Falha Térmica $300^{\circ}\text{C} < T < 700^{\circ}\text{C}$	>1,0	<0,1	0,1 a 3,0
Falha Térmica $T > 700^{\circ}\text{C}$	>1,0	<0,1	>3,0

Fonte: Adaptado de AMORA (2013)

2.3 Limitações das técnicas convencionais

Mesmo sendo normatizadas e aplicadas há anos pelas concessionárias de energia as técnicas de análise de gases dissolvidos não são fundamentadas em uma ciência exata, e sim resultado de evidências empíricas e por isso a interpretação de seus resultados está sujeita à variação. Somado a esse fator, temos ainda a influência dos diferentes métodos de fabricação dos equipamentos, dos seus sistemas de circulação e conservação do óleo, além do próprio histórico de cargas aplicadas ao transformador (BARBOSA, 2013).

Para um diagnóstico eficiente, há necessidade de um especialista na área, que tenha uma experiência extensa e grande conhecimento das técnicas de análise. Ademais, o técnico não deve restringir-se a uma única metodologia, pois este é um problema complexo e muito específico para cada caso. Assim sendo, os diagnósticos que são baseados na análise de gases dissolvidos são muito dependentes da experiência de especialistas humanos, e por isso, muitas vezes são insatisfatórios e incompletos (BARBOSA, 2013; CASTILLO, 2003).

2.4 Inteligências Computacionais

Frente as limitações das técnicas convencionais de diagnósticos diversas metodologias baseadas no uso de inteligência computacional foram propostas para a avaliação dos dados provenientes da cromatografia. O objetivo é diminuir as incertezas envolvidas no processo e aumentar a precisão dos diagnósticos de falhas em transformadores isolados a óleo. Tais técnicas possuem como vantagem principal a manipulação de grandes volumes de dados, o que pode facilitar a detecção e identificação das falhas. As técnicas utilizadas mais frequentemente são redes neurais artificiais, sistemas especialistas e lógica difusa (MORAIS, 2004).

2.5 Construção do Conhecimento

Tradicionalmente a transformação de dados em conhecimento advém do processamento destes por especialistas, entretanto, vivemos nos dias atuais uma mudança de paradigma, que torna esse procedimento cada vez mais inviável, seja devido à automação dos processos, ou pela elevada quantidade de dados relacionados que podem se encontrar em fontes distintas. Diante disso, Fayyad, Shapiro e Smyth (1996) definiram o conceito de Descoberta de Conhecimento das bases de dados, também conhecido como *Knowledge Discovery in Databases (KDD)*.

O processo de KDD consiste em uma série de etapas a serem executadas a fim de gerar um conhecimento específico ao final de sua realização, sendo estas: limpeza, integração, seleção e transformação dos dados, mineração, avaliação ou pós-processamento e visualização dos resultados (AMO; ROCHA, 2003).

Quando discutiram Mineração de Dados ou *Data Mining* Fayyad, Shapiro e Smyth (1996), definiram como um “passo no processo de descoberta de conhecimento que consiste na realização da análise dos dados e na aplicação de algoritmos de descoberta que, sob certas limitações computacionais, produzem um conjunto de padrões de certos dados”. Todavia, mesmo tendo um alto desempenho no processamento de grandes quantidades de informações, tal processo não é puramente automático, uma vez que existe uma série de ferramentas para execução dos algoritmos, bem como a análise dos resultados, em que ambas tarefas precisam ser efetuadas por seres humanos (CAMILO; SILVA, 2009).

Ainda nesse contexto, Bueno e Viana (2012), alertam para a importância em se considerar dois aspectos: tarefa e técnica de mineração. A tarefa é nada mais que a especificação do objetivo de busca do dado, ou seja, a informação e o tipo de categoria que se deseja obter. Já a técnica, refere-se ao método utilizado para realizar a tarefa definida, ou seja, o que será feito para atender a tarefa especificada.

Quando fala-se de mineração de dados, existem diferentes tipos de tarefas passíveis a serem exploradas, sendo as mais comuns (CAMILO; SILVA, 2009):

- Classificação, que é a análise das características dos dados para relacionar o objeto classificado a uma categoria específica previamente definida. Para tanto, cada registro possui atributos de predição, assim como a classe a que pertence. Esse método é comumente utilizado na identificação de pessoas, classificação de objetos, diagnóstico de doenças e falhas em equipamentos;

- Regressão, consiste na tarefa de identificação do registro através de um valor numérico não categórico. A regressão se subdivide em regressão linear (usada quando a variável preditora possui uma relação linear com a predição) e não linear (usada quando a variável preditora possui uma relação não-linear com a predição);
- Clusterização, é definida como a tarefa de agrupamento dos registros tendo como base as similaridades existentes entre eles, segmentando-os assim em subconjuntos ou grupos;
- Associação, que é a tarefa de identificação de relacionamento de atributos, através de estruturas condicionais “SE..ENTÃO”. Nesses casos o objetivo da tarefa desempenhada é entender como um atributo α implica em β ;
- Descrição, que representa a tarefa de descrever os padrões demonstrados pelos dados, sendo utilizado na análise exploratória.

A escolha da tarefa dá-se em detrimento do problema e tipo de conhecimento que espera-se ser gerado, e para cada tipo de tarefa, existem diversos tipos de técnicas que podem ser utilizadas isoladamente ou separadamente para a realização das atividades de mineração de dados. A decisão pela aplicação desta ou daquela técnica, parte do objetivo do estudo, desempenho computacional e grau de precisão de cada técnica aplicada a cada situação (CAMILO; SILVA, 2009).

2.6 Machine Learning e Técnicas de Classificação

O *Machine Learning* (ML) ou Aprendizado de Máquina é uma área da Inteligência computacional que estuda a elaboração de diferentes métodos que tornam possível a extração de conhecimento a partir de um conjunto de dados. Dessa forma, é avaliado a capacidade de um algoritmo aprender a melhorar o seu desempenho na execução de uma determinada tarefa por meio das experiências anteriores (MITCHELL, 1997). Tais técnicas foram desenvolvidas tendo como bases diferentes áreas do conhecimento, como sistemas biológicos, que geraram as redes neurais e algoritmos genéticos, assim como aprendizado simbólico, representado pelas árvores de decisão ou *decision trees*, ou também, de teorias estatísticas, a exemplo de *Naïve Bayes* (HAN; KRAMER, 2006).

As técnicas ML podem ser classificadas em dois paradigmas, aprendizado supervisionado e não supervisionado, e a sua escolha determina como o aprendizado ocorrerá. A distinção dessa classificação se faz presente nos seguintes aspectos:

- Aprendizado Supervisionado, consiste no treinamento dos algoritmos a partir de uma categorização inicial dos dados, assim sendo, estes são compostos pelo objeto de entrada e o valor de saída esperado. Uma boa analogia é a da figura de um professor que corrige a resposta obtida. Dessa forma, o treinamento do algoritmo se dá pela análise dos dados de treinamento para a produção de uma saída já apresentada, podendo, posteriormente, ser aplicado na classificação de outros dados do mesmo domínio. Técnicas como *k-Nearest Neighbour* (k-NN), *Naïve Bayes*, entre outros, são exemplos de algoritmos que usam esse tipo de aprendizado (NIEUWENHOFF, 2017).
- Aprendizado não-supervisionado consiste, por sua vez, na identificação de uma estrutura não conhecida sem uma saída definida, ou seja, sem a figura do professor. O algoritmo aprende sozinho a agrupar as entradas a partir das suas relações, categorias, padrões, entre outros, que identifica com base em alguma medida de similaridade entre os atributos. *Clustering*, *k-means*, Estatísticas, são exemplos de algoritmos que se utilizam desse aprendizado (NIEUWENHOFF, 2017).

Este trabalho consiste na aplicação de algoritmos computacionais para a classificação de falhas incipientes em transformadores com base na concentração dos gases dissolvidos no óleo. Para tanto, foram considerados nesse estudo três técnicas de *machine learning*, de aprendizado supervisionado, sendo estas: *k-Nearest neighbor* (K-NN), *Naïve-Bayes* e *Multilayer perceptron* (MLP).

2.6.1 *k-Nearest Neighbor*

O *K-nearest-neighbor* é um dos mais simples e fundamentais métodos de machine learning e geralmente a primeira escolha para estudos de classificação quando não se tem nenhum ou quase nenhum conhecimento prévio sobre a distribuição dos dados. O método foi desenvolvido a partir da necessidade de realizar análises de discriminantes quando estimativas confiáveis de densidade de probabilidade são desconhecidas ou difíceis de se determinar. Fix e Hodges, 1951, foram os primeiros a introduzir o método que posteriormente ficaria conhecido como regra KNN em um relatório não publicado da escola de medicina de aviação da força Aérea dos EUA, mais tarde, em 1967, algumas das propriedades formais da regra foram elaboradas (PETERSON, 2009).

O Classificador K-nearest-neighbor é baseado na distância entre uma amostra de teste e as amostras de teste especificadas (PETERSON, 2009).

Seja x_a um ponto, da amostra teste, no espaço em que x_i é uma entrada com p

características, $x_a = (x_{i1}, x_{i2}, \dots, x_{ip})$ e um outro ponto $x_b = (x_{l1}, x_{l2}, \dots, x_{lp})$, da amostra de treinamento, $d(x_a, x_b)$ é a distância entre esses pontos (PETERSON, 2009).

A distância é generalizada pela matemática como uma métrica para um dado espaço. A mais comum é a **distância euclidiana**, onde:

$$d(x_a, x_b) = \sqrt{(x_{i1} - x_{l1})^2 + (x_{i2} - x_{l2})^2 + \dots + (x_{ip} - x_{lp})^2}$$

Porém são possíveis outros tipos de distância como por exemplo, a **distância Manhattan**, em que:

$$d(x_a, x_b) = |x_{i1} - x_{l1}| + |x_{i2} - x_{l2}| + \dots + |x_{ip} - x_{lp}|$$

Ou a **distância Minkowski**:

$$d(x_a, x_b) = (|x_{i1} - x_{l1}|^p + |x_{i2} - x_{l2}|^p + \dots + |x_{ip} - x_{lp}|^p)^{\frac{1}{p}}$$

Nota-se que a distância *minkowski* é uma generalização para a distância *manhattan* e a distância euclidiana, uma vez que podemos obtê-las diretamente da fórmula quando p é respectivamente igual a 1 e 2 (PETERSON, 2009).

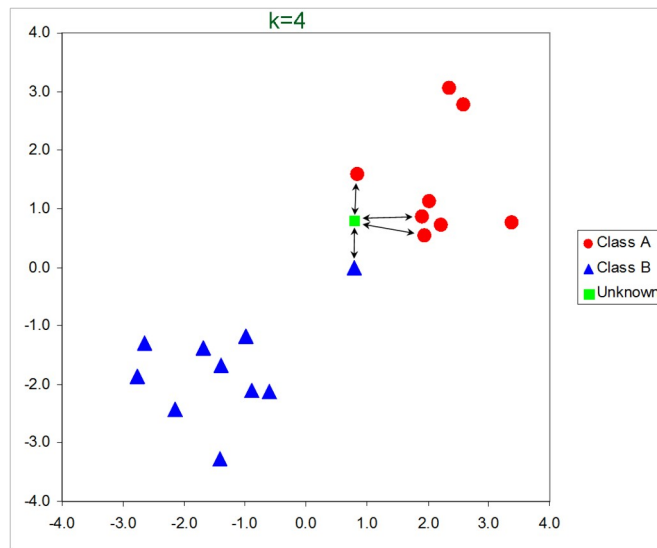
Tendo calculado todas as distâncias, entre conjunto teste e treinamento, a classe prevista da amostra teste é igual à classe definida do seu vizinho de treinamento mais próximo, onde m_i é o vizinho mais próximo de x_a (PETERSON, 2009). Formalmente, temos:

$$d(m_i, x_a) = \min_j [d(m_j, x)]$$

Para k vizinhos mais próximos, a predição da classe da amostra de teste x_a é igual à classe mais frequente entre as k amostras de treinamento mais próximas (PETERSON, 2009).

Na figura 04, temos um scatterplot com 19 amostras que ilustram o funcionamento do KNN. Para os 4 mais próximos pontos vizinhos da amostra, a classe mais frequente é a em vermelho, portanto o ponto é classificado como pertencente a classe vermelho.

Figura 04 – *K-nearest neighbor* para k igual a quatro com p (características) igual a dois



Fonte: Peterson, 2009

2.6.2 Naïve-Bayes

As metodologias de aprendizagem bayesianas utilizam um modelo de probabilidade, baseado no conhecimento inicial do problema, sendo este combinado com exemplos de treinamentos para determinar a probabilidade final de classificação (MITCHELL, 1997).

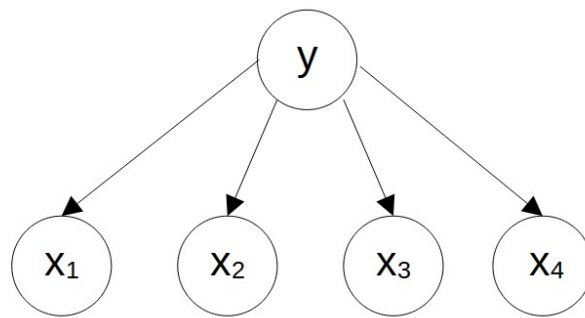
O Algoritmo Naïve Bayes (NB), tem como base o Teorema de Bayes, o qual, determina a probabilidade de um evento ocorrer dada a probabilidade de outro evento que já ocorreu (MITCHELL, 1997).

A equação a seguir, ilustra a probabilidade condicional de A dado B, ou seja, a probabilidade da hipótese A ocorrer, dado que B é verdadeiro, onde A e B são eventos e $P(B) \neq 0$.

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

O classificador é chamado de naïve, do inglês ingênuo, pois assume que os atributos do conjunto de dados são independentes, ou seja, cada atributo x_i tem influência a classe y mas não influencia os outros atributos x_{i+1} a x_n (MITCHELL, 1997). Na Figura 05, observamos uma representação gráfica do conceito especificado.

Figura 05 – Ilustração do Naïve Bayes



Fonte: Gallagher, Madden e D'arcy (2015)

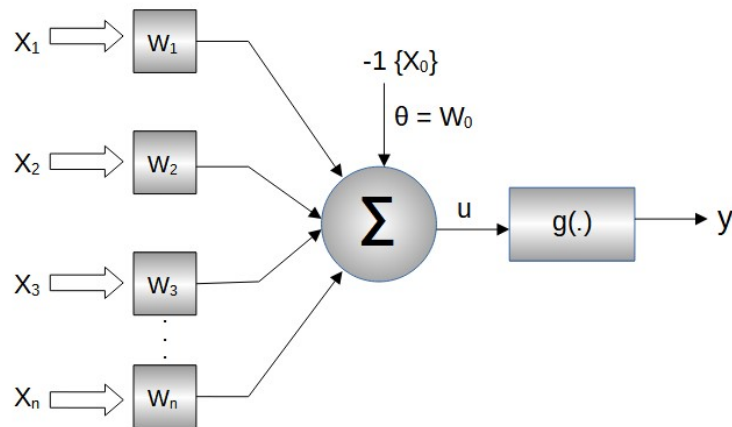
O NB executa a classificação realizando uma leitura dos dados de treinamento uma única vez para estimar as probabilidades. Ademais, o classificador pode ser executado a partir de dois modelos: binário ou multinomial (MITCHELL, 1997).

2.6.3 Multilayer perceptron (MLP)

As redes neurais artificiais, ou simplesmente RNA, são algoritmos de *machine learning* que cumprem tarefas que vão desde o reconhecimento de padrões até a aproximação de funções. Elas são baseadas no conceito de neurônio biológico, envolvendo estruturas matemáticas capazes de aprender por meio de treinamentos (NIEUWENHOFF, 2017).

O *Perceptron* é a forma mais simples de configuração de uma RNA. Sua simplicidade advém de sua estrutura, uma vez que ele é composto por apenas uma camada neural, tendo também apenas um neurônio artificial nesta camada única. Na Figura 06, observa-se uma ilustração de uma rede Perceptron com n sinais de entrada, definidos pelo problema a ser mapeado, e apenas uma saída, uma vez que este tipo de rede é composto por apenas um único neurônio. O Conjunto de pesos sinápticos $\{w_i\}$, é responsável por ponderar as influências de cada entrada, frente ao objetivo funcional atribuído ao neurônio. O valor resultante desse processo é somado ao limiar de ativação $\{\theta\}$, e repassado como argumento da função de ativação. O resultado é a saída $\{y\}$, que é função de u , sendo este o potencial de ativação (SILVA; SPATTI; FLAUZINO, 2016).

Figura 06 – Ilustração da rede Perceptron



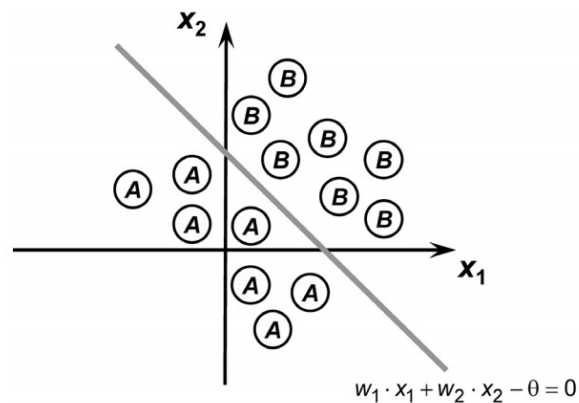
Fonte: Silva, Spatti e Flauzino (2016)

Matematicamente, podemos resumir o processamento realizado pelo Perceptron através das seguintes expressões:

$$\begin{cases} u = \sum_{i=1}^n w_i \cdot x_i - \theta \\ y = g(u) \end{cases}$$

Em termos práticos, pode-se interpretar o comportamento como o de um classificador de padrões cuja função é dividir classes, que sejam linearmente separáveis. Na figura 07, observamos uma representação gráfica do funcionamento do Perceptron de duas entradas, com sua fronteira de separabilidade. Para três entradas (três dimensões), a fronteira e separação seria representada por um plano; para dimensões superiores, tais fronteiras seriam hiperplanos (SILVA; SPATTI; FLAUZINO, 2016).

Figura 07 – Ilustração da fronteira de separação (neurônio com duas entradas)

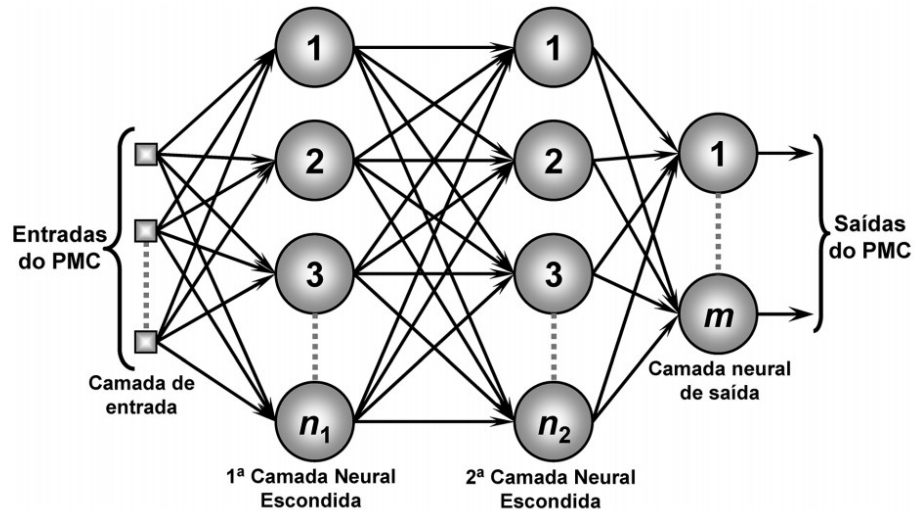


Fonte: Silva, Spatti e Flauzino (2016)

O Perceptron simples (camada única), lida apenas com problemas de saídas binárias, o que limita o número de classes em duas. Ademais, somente consegue convergir se as classes envolvidas com o problema a ser mapeado forem linearmente separáveis. Situações diferentes a esta, ou seja, mais de duas classes e problemas não linearmente separáveis, somente podem ser resolvidas por intermédio de uma rede *Multilayer perceptron* (MLP) (SILVA; SPATTI; FLAUZINO, 2016).

As redes *Multilayer perceptron* são caracterizadas pela presença de ao menos uma camada intermediária, conhecida como *hidden layer*, camada escondida, ou ainda, camada oculta, de neurônios. Ficando esta, situada entre a camada de entrada e a respectiva camada neural de saída. Dessa forma, as redes MLP possuem no mínimo duas camadas (SILVA; SPATTI; FLAUZINO, 2016).

Figura 08 – Ilustração de uma rede MLP



Fonte: Silva, Spatti e Flauzino (2016)

O treinamento das redes MLP é feito utilizando o algoritmo *backpropagation*, também conhecido como regra Delta generalizada, mediante a aplicação de duas fases específicas: Fase *forward* ou propagação adianta e fase *backward* ou propagação reversa. Na fase *forward*, os atributos $\{x_1, x_2, \dots, x_n\}$ de uma amostra do conjunto de dados de treinamento são inseridos nas entradas da MLP e propagados camada a camada até a saída. Nessa fase não há atualização dos pesos sinápticos, que só serão atualizados na fase *backward*. A fase *backward* é iniciada após calcular o erro entre a saída obtida e a saída esperada, uma vez que o erro é utilizado na atualização dos pesos, que é feito recursivamente (SILVA; SPATTI; FLAUZINO, 2016).

A especificação da topologia da rede mais apropriada para realizar o mapeamento para um problema específico é usualmente efetuada de forma empírica, pois o dimensionamento depende, entre outros fatores, do algoritmo de aprendizagem utilizado, da forma como as matrizes de pesos foram iniciadas, da complexidade do problema a ser mapeado, da disposição espacial das amostras e, por sua vez, da qualidade do conjunto de treinamento disponível. Porém, algumas metodologias podem ser aplicadas para determinar as melhores topologias candidatas, como por exemplo, a validação cruzada (*cross-validation*), ou ainda nos casos em que define-se apenas uma única camada escondida:

$$n_1 = 2 \cdot n + 1 \quad \{\text{Método de Kolmogorov}\}$$

$$2 \cdot \sqrt{n} + n_2 \leq n_1 \leq 2 \cdot n + 1 \quad \{\text{Método de Fletcher-Gloss}\}$$

Onde:

n_1 = Quantidade de neurônios na camada escondida

n = Número de entradas na rede

n_2 = Quantidade de neurônios na camada de saída

Alternativamente, existe também uma heurística comumente utilizada e implementada em plataformas como a *Weka (Waikato environment for knowledge analysis)*, sendo definida como:

$$n_1 = \frac{n + n_c}{2}$$

Onde:

n_c = Número de classes do problema

3 PROPOSTA DE PESQUISA

Neste capítulo apresenta-se a metodologia utilizada para a realização dessa pesquisa. Os tópicos foram subdivididos em: conjunto de dados avaliado, linguagem de programação e bibliotecas utilizadas, classificadores e parâmetros utilizados, aplicação dos classificadores e classificação por métodos convencionais.

3.1 Conjunto de dados avaliado

Para a realização desta pesquisa foi utilizado um conjunto de dados (*dataset*) contendo informações dos gases dissolvidos no óleo de transformadores de potência e suas respectivas falhas incipientes associadas. O conjunto de dados foi organizado por Enwen Li e disponibilizado *online* no IEEE DATAPORT, sob regime de licença “*Free Cultural Works*”, permitindo o estudo, a aplicação, a cópia e/ou modificação por qualquer pessoa e para qualquer finalidade, desde que sejam atribuídos créditos a seu criador.

O *dataset* original é composto por 7 (sete) colunas e 201 registros, sendo estas:

NM: identificador do número do registro;

H2: concentração em PPM de hidrogênio (H₂) dissolvido no óleo do transformador;

CH4: concentração em PPM de metano (CH₄) dissolvido no óleo do transformador;

C2H6: concentração em PPM de etano (C₂H₆) dissolvido no óleo do transformador;

C2H4: concentração em PPM de etileno (C₂H₄) dissolvido no óleo do transformador;

C2H2: concentração em PPM de acetileno (C₂H₂) dissolvido no óleo do transformador;

TYPE: identificador da falha, podendo ser este: *partial discharge*, *Spark discharge*, *Arc discharge*, *High-temperature overheating*, *Middle-temperature overheating*, *Low-temperature overheating*, *Low/Middle-temperature overheating*.

O *dataset* original sofreu um pré-processamento de dados, em que estes foram normalizados pela média.

A partir do *dataset* original, foi criado um segundo *dataset*, (para referências futuras, o chamaremos de *dataset* das razões) normalizado pela média e contendo as informações das razões R1 a R5 utilizada na maioria das aplicações da teoria de DGA (*Dissolved Gas Analysis*), conforme tabela 2.4 (IEEE, 2008). As razões que resultassem em divisões por zero, foram substituídas pela mediana da classe a que fosse correspondente.

Na tabela a seguir encontra-se a distribuição dos dados por tipo de falha.

Tabela 05 – Quantidade de registros por tipo de falha

Tipo de Falha	Número de Registros
Partial discharge	16
Spark discharge	49
<i>Arc discharge</i>	54
High-temperature overheating	38
Middle-temperature overheating	9
Low-temperature overheating	19
Low/Middle-temperature overheating	16
Total	201

Fonte: elaborada pelo autor.

3.2 Linguagem de programação e bibliotecas utilizadas

Para modelagem dos classificadores, bem como para sua execução, utilizou-se da linguagem *open source* Python (versão 3.10.0.), além das bibliotecas pandas (versão 1.3.4) para manipulações de objetos *dataframe*, numpy (versão 1.21.3) para manipulações e cálculos numéricos, scikit-learn (versão 1.0.2) para performar os classificadores, seaborn (versão 0.11.2) e matplotlib (versão 3.5.1), ambas para plotagem de gráficos.

3.3 Classificadores e parâmetros utilizados

Foram utilizados três métodos de classificação diferentes, para resolução do problema proposto. Um algoritmo de distância ao centróide do tipo K-nearest neighbors (k vizinhos mais próximos); Uma rede neural do tipo Multilayer Perceptron – MLP (Perceptron multicamadas) e um modelo de classificação probabilística do tipo Naïve Bayes.

Cada classificador possui diferentes possibilidades de manipulação de seus parâmetros. A seguir, encontram-se as escolhas adotadas em cada método de classificação implementado, bem como os motivos de sua aplicação.

3.3.1 Parâmetros do classificador K-Nearest Neighbor (K vizinhos mais próximos)

O algoritmo de classificação K-nearest neighbor permite manipulação dos seguintes aspectos:

- **weights (pesos):** define a influência que cada instância de treinamento (X_q) tem sobre o classificador. Podendo ser “uniform”, em que todos os pesos uniformes, ou seja, todos os pontos em cada vizinhança são ponderados igualmente; ou “distance”, em que os vizinhos mais próximos de um ponto de consulta terão uma influência maior do que os vizinhos mais distantes;
- **algorithm (algoritmo):** determina o tipo de algoritmo utilizado para calcular os vizinhos mais próximos. Este pode ser “brute”, que calcula todas as distâncias dos vizinhos mais próximos, “ball_tree” ou “kd_tree”, que são algoritmos baseados em *projective trees*, dessa forma, estes categorizam os pontos tendo como base sua projeção em algum espaço de dimensão inferior (Kumar Zhang & Nayar, 2008);
- **metric (métrica):** define qual o tipo de métrica utilizada para calcular a distância, podendo ser “euclidean”, “manhattan”, “chebyshev”, “minkowski”, “mahalanobis”, etc;
- **n_neighbors ou K:** número de vizinhos utilizados para o cálculo do centroide de cada classe.

Utilizou-se como parâmetros para o classificador pesos uniformes (parâmetro `weight = “uniform”`), algoritmo de cálculo “brute” para as distâncias e métrica “minkowski”, com $p = 2$, que corresponde a distância euclidiana. O número de vizinhos mais próximos (K), foi definido arbitrariamente como 5.

3.3.2 *Parâmetros do classificador Multilayer Perceptron (Perceptron multicamadas)*

O Classificador MLP permite manipulação dos seguintes parâmetros:

- **hidden_layer_sizes:** uma *tupla* contendo o número de camadas ocultas e o número de neurônios da(s) camada(s) oculta(s);
- **Activation:** função de ativação da camada oculta, podendo ser, função identidade, logística, tangente hiperbólica ou unidade linear retificada (também conhecida como ReLu);
- **Solver:** o solucionador para otimização dos pesos, podendo ser: ‘lbfgf’, um otimizador na família de métodos quasi-Newton, ‘sgd’, referente ao gradiente descendente estocástico e ‘adam’, que é um otimizador baseado no SGD;
- **Tol:** tolerância para a otimização. Quando o erro (diferença entre o resultado esperado e o resultado obtido) não melhora pelo menos a tolerância definida, por dez épocas

consecutivas, o treinamento é interrompido;

- Max_iter: limite máximo de iterações (ou épocas) caso a tolerância não seja alcançada.

Para o problema definido nesse trabalho, optou-se por usar uma camada oculta. A quantidade de neurônios nesta foi definida inicialmente pelo Método de Kolmogorov, como se observa a seguir:

$$\eta_1 = 2 \cdot \eta + 1$$

Onde:

$$\eta_c = 5$$

Dessa forma,

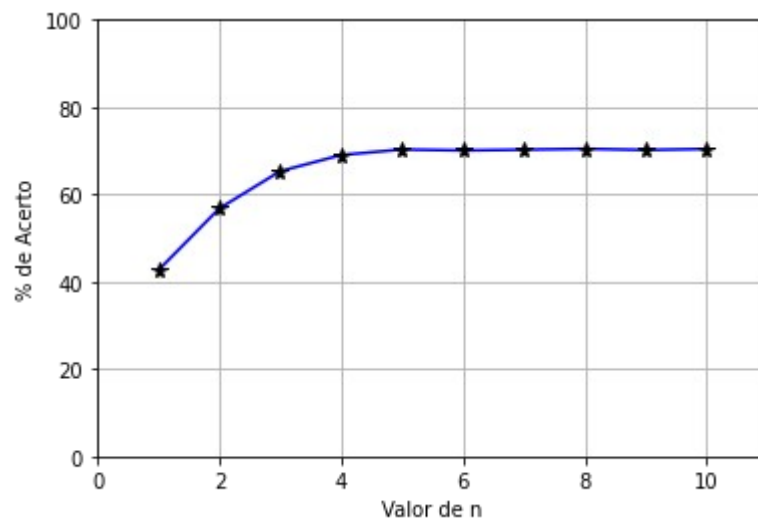
$$\eta_1 = 2 \cdot 5 + 1$$

$$\eta_1 = 11$$

Como Fletcher-Gloss utiliza o método de Kolmogorov como limite superior, estabeleceu-se onze (11) como o número máximo de neurônios possíveis e avaliou-se como números menores teriam efeito sobre a precisão do classificador para cada *dataset*, a fim de definir o número ideal. Para tanto, a rede neural foi performada por cem vezes, calculou-se uma média da precisão obtida e, por fim, escolheu-se o valor que gerou o melhor desempenho médio.

No Gráfico 01, observa-se que para o *dataset* original o melhor η foi $\eta = 6$. A partir do $\eta=4$, observamos também que o aumento do número de neurônios não tem influência significativa na precisão da rede.

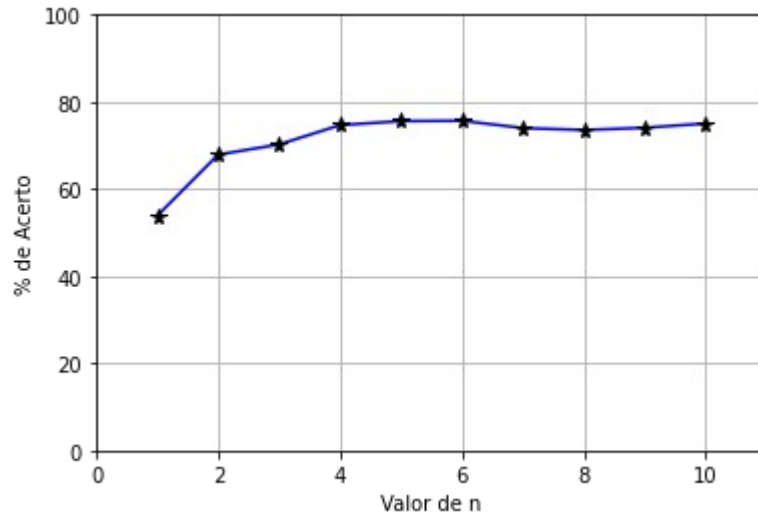
Gráfico 01 – Precisão do classificador pelo valor de η para o dataset original



Fonte: elaborada pelo autor.

Já no Gráfico 02, observa-se que para o dataset das razões o melhor η também foi $\eta = 6$, porém, observamos uma variação maior no desempenho ao mudarmos η .

Gráfico 02 – Precisão do classificador pelo valor de η para o dataset o dataset das razões



Fonte: elaborada pelo autor.

Por fim estabeleceu-se que, a função de ativação utilizada seria a tangente hiperbólica, o otimizador dos pesos o lgbf, a tolerância estabelecida com valor de 10^{-15} e o número máximo de épocas fixado em 20.000.

3.3.3 Parâmetros do classificador Gaussian Naïve Bayes

Os parâmetros do classificador Gaussian Naïve Bayes são apenas dois, sendo estes:

- prior: especifica as probabilidades prévias de cada classe;
- var_smoothing: parte da maior variância de todos os recursos que é adicionada às variâncias para estabilidade de cálculo.;

Estes foram setados nos valores padrões para o classificador, ou seja, sem probabilidade prévia das classes e com variância de 10^{-9} .

3.4 Aplicação dos classificadores

Para a aplicação dos classificadores, inicialmente separou-se os dados dos dois *datasets* em conjuntos de treinamento e de teste, nas proporções 70/30 respectivamente e de

forma aleatória. O conjunto de treinamento foi submetido a um processo de *oversampling smote* (*Synthetic Minority Oversampling Technique*) com o intuito de balancear os dados, melhorando assim, o desempenho dos modelos.

Os classificadores foram performados para os mesmos conjuntos de treinamento e teste, ao final, calcularam-se diferentes métricas (*precision*, *recall*, *f1-score* e *accuracy*), para avaliar o desempenho dos classificadores.

3.5 Classificação por métodos convencionais

Com o intuito de estabelecer uma comparação direta entre os métodos convencionais e os métodos computacionais, o *dataset* também foi classificado tendo como base a metodologia idealizada por Rogers e formalizada pela ABNT NBR-7274.

4 RESULTADOS E DISCUSSÕES

Este trabalho alcançou resultados para duas execuções efetuadas de três algoritmos de aprendizado de máquina (*K-nearest-neighbor*, *Naïve Bayes* e *Multilayer perceptron*) na classificação dos dados estudados, sendo estas: (i) execução dos algoritmos com o *dataset* original normalizado por meio do *z-score* e (ii) execução do *dataset* das razões. Ademais, também foram registrados os resultados da classificação por meio do método de Rogers, a fim de estabelecer uma comparação entre os métodos convencionais e os computacionais. As seções a seguir, apresentam os resultados para cada uma das execuções.

A acurácia indica a performance global do classificador. Entretanto, ela por si só, não é capaz de indicar o seu desempenho. Sendo assim outras métricas precisam ser avaliadas, como a precisão e o *recall*. Cada uma dessas métricas possuem grau de importância diretamente associado à classe que elas descrevem e dependem do problema avaliado.

A precisão é especialmente utilizada em casos que falsos positivos são considerados mais prejudiciais que falsos negativos. Este é o caso das classes *Partial discharge*, *Low-temperature overheating*, *low/Middle-temperature overheating*, *Middle-temperature overheating*, pois uma predição que indica que a falha incipiente é deste tipo quando na verdade não é, implica uma continuidade da operação do equipamento com uma falha que pode ser muito mais severa do que a predita.

Já o *recall* pode ser usada em uma situação em que os Falsos Negativos são considerados mais prejudiciais que os Falsos Positivos. Sendo este o caso para as classes *High-temperature overheating* e *Arc discharge*. Por se tratarem de casos extremos de funcionamento do transformador, perder um caso positivo tem um custo muito maior do que classificar erroneamente como positivo.

Por fim o *f1-score* representa a média harmônica entre as duas outras medidas, dando assim uma visão geral para as classes avaliadas.

4.1 Execução 1: *Dataset* original

As tabelas 06, 07 e 08, contêm as métricas de classificação de cada algoritmo performado segregado por classe para o *dataset* original normalizado.

Tabela 06 – Métricas do *K-nearest neighbor* para *dataset* original

Classe	<i>precision</i>	<i>Recall</i>	<i>f1-score</i>
<i>Arc discharge</i>	0,64	0,44	0,52
<i>High-temperature overheating</i>	1,00	0,55	0,71
<i>Low-temperature overheating</i>	0,71	0,71	0,71
<i>Low/middle-temperature overheating</i>	0,33	0,50	0,40
<i>Middle-temperature overheating</i>	0,20	0,33	0,25
<i>Partial discharge</i>	0,60	1,00	0,75
<i>Spark discharge</i>	0,58	0,69	0,63

Fonte: elaborada pelo autor.

Quando observamos a performance do K-NN frente as classes possíveis para o *dataset* normalizado percebemos que a precisão foi superior a 0,7 em apenas duas classes, sendo estas *high-temperature overheating* (1,00) e *low-temperature overheating* (0,71). Já para as classes *middle-temperature overheating* e *low/middle-temperature overheating* obtivemos os menores valores de precisão, 0,33 e 0,20, respectivamente. O *recall* foi melhor para os casos de *low-temperature overheating* e *partial discharge*, respectivamente, 0,71 e 1,00, porém nas classes *arc discharge* (0,44), *middle-temperature overheating* (0,33) e *high-temperature overheating* (0,55), obtiveram-se valores muito baixos. Por se tratar da média harmônica, o *f1-score* acaba destacando positivamente os casos que foram muito bons nas duas métricas (*high-temperature overheating* e *low-temperature overheating* com 0,71 e 0,69 para a *partial discharge*) e negativamente naqueles que foram ruins, ou seja, *low/middle-temperature overheating* (0,40), *middle-temperature overheating* (0,25) e *arc discharge* (0,52).

Tabela 07 – Métricas do *Multilayer perceptron* para *dataset* original

Classe	<i>precision</i>	<i>Recall</i>	<i>f1-score</i>
<i>Arc discharge</i>	0,88	0,88	0,88
<i>High-temperature overheating</i>	0,90	0,82	0,86
<i>Low-temperature overheating</i>	0,80	0,57	0,67
<i>Low/middle-temperature overheating</i>	0,25	0,50	0,33
<i>Middle-temperature overheating</i>	0,33	0,33	0,33
<i>Partial discharge</i>	1,00	1,00	1,00
<i>Spark discharge</i>	0,82	0,88	0,85

Fonte: elaborada pelo autor.

Por sua vez, para o modelo com MLP, observamos que *high-temperature overheating* (0,90), *partial discharge* (1,00) e *arc discharge* (0,88) obtiveram valores muito altos de precisão, em detrimento de *middle-temperature overheating* (0,33) e *low/middle-temperature overheating* (0,25) que não performaram de forma satisfatória. Quando se observa os resultados obtidos da perspectiva do *recall*, temos bons resultados para *partial*

discharge (1,00), *spark discharge* (0,88), *arc discharge* (0,88) e *high-temperature overheating* (0,82), porém nas classes *middle-temperature overheating* (0,33), *low-temperature overheating* (0,57) e *low/middle-temperature overheating* (0,50) obtiveram-se valores consideravelmente baixos. Nesse caso, o *f1-score* penalizou consideravelmente a classe *low/middle-temperature overheating* e *middle-temperature overheating* (0,33), pois estas possuem resultados baixos de *recall* e *precision*.

Tabela 08 – Métricas do *Naïve Bayes* para *dataset* original

<i>Classe</i>	<i>Precision</i>	<i>Recall</i>	<i>f1-score</i>
<i>Arc discharge</i>	0,83	0,31	0,45
<i>High-temperature overheating</i>	0,33	0,09	0,14
<i>Low-temperature overheating</i>	0,27	0,43	0,33
<i>Low/middle-temperature overheating</i>	0,20	0,50	0,29
<i>Middle-temperature overheating</i>	0,00	0,00	0,00
<i>Partial discharge</i>	0,36	0,83	0,50
<i>Spark discharge</i>	0,44	0,44	0,44

Fonte: elaborada pelo autor.

O modelo performado com *Naïve Bayes*, apresentou um desempenho bastante insatisfatório em todas as métricas avaliadas. Para a precisão, obtivemos apenas para a *arc discharge* (0,83), um valor superior a 0,7; já para o *recall*, apenas o *partial discharge* (0,83) superou esse patamar; e no *f1-score* nenhuma classe conseguiu tamanha expressividade.

Quando analisamos a performance global dos classificadores, por meio da acurácia, percebemos que a MLP foi o modelo que obteve melhor resultado (0,80) em comparação ao KNN (0,60) e *Naïve Bayes* (0,36), como observado na a seguir.

Tabela 09 – Acurácia para cada classificador aplicado ao *dataset* original

<i>Classificador</i>	<i>Acurácia</i>
<i>K-nearest neighbor</i>	0,60
<i>Multilayer perceptron</i>	0,80
<i>Naïve Bayes</i>	0,36

Fonte: elaborada pelo autor.

Ikeshoji (2020) ao desenvolver uma metodologia para diagnóstico de falhas incipientes em transformadores de potência utilizando análise de gases e aprendizado de máquina, observou que aplicando KNN para os “dados brutos”, com K igual a 1, obteve-se uma taxa de acerto (acurácia) de 0,809. Já ao aplicar *Naïve Bayes*, obteve como resultado o valor de 0,738. Por fim, ao implementar uma MLP com 1 camada e 9 neurônios, obteve um valor de acurácia de 0,815. Os resultados foram obtidos para uma amostra (2048 registros)

muito maior que a utilizada nesse estudo, o que explica, em parte, as diferenças de resultados obtidos.

4.2 Execução 2: *Dataset* das razões

As tabelas 10, 11 e 12, contém as métricas de classificação de cada algoritmo performado segregado por classe para o *dataset* das razões.

Tabela 10 – Métricas do *K-nearest neighbor* para *dataset* das razões

<i>Classe</i>	<i>Precision</i>	<i>Recall</i>	<i>f1-score</i>
<i>Arc discharge</i>	0,79	0,94	0,86
<i>High-temperature overheating</i>	0,64	0,64	0,64
<i>Low-temperature overheating</i>	0,83	0,71	0,77
<i>Low/middle-temperature overheating</i>	0,00	0,00	0,00
<i>Middle-temperature overheating</i>	0,33	0,33	0,33
<i>Partial discharge</i>	0,67	0,33	0,44
<i>Spark discharge</i>	0,76	0,81	0,79

Fonte: elaborada pelo autor.

O KNN frente ao *dataset* das razões, mostrou resultados satisfatórios de precisão, *recall* e conseqüentemente, *f1-score*, para as classes *arc discharge* (respectivamente, 0,79, 0,94 e 0,86), *low-temperature overheating* (respectivamente, 0,83, 0,71 e 0,77) e *spark discharge* (respectivamente, 0,76, 0,81 e 0,79). Porém, o classificador não acertou nenhuma das predições para a classe *low/middle-temperature overheating*, e obteve valores *baixos* para *middle-temperature overheating* e *Partial discharge*.

Tabela 11 – Métricas do *Multilayer perceptron* para *dataset* das razões

<i>Classe</i>	<i>precision</i>	<i>Recall</i>	<i>f1-score</i>
<i>Arc discharge</i>	0,81	0,82	0,81
<i>High-temperature overheating</i>	0,70	0,64	0,67
<i>Low-temperature overheating</i>	0,75	0,86	0,80
<i>Low/middle-temperature overheating</i>	0,20	1,00	0,33
<i>Middle-temperature overheating</i>	0,00	0,00	0,00
<i>Partial discharge</i>	1,00	0,67	0,80
<i>Spark discharge</i>	0,92	0,69	0,79

Fonte: elaborada pelo autor.

Para o MLP, as melhores precisões obtidas foram os das classes *arc discharge* (0,81), *partial discharge* (1,00) e *spark discharge* (0,92), já as piores se concentraram em *middle-temperature overheating* (0,00), *low/middle-temperature overheating* (0,20). Por sua vez, para

o *recall*, os melhores desempenhos observados foram das classes *low/middle-temperature overheating* (1,00), *spark discharge* (0,85) e *low-temperature overheating* (0,86); os piores foram *middle-temperature overheating* (0) e *high-temperature overheating* (0,64).

Tabela 12 – Métricas do *Naïve bayes* para *dataset* das razões

<i>Classe</i>	<i>precision</i>	<i>Recall</i>	<i>f1-score</i>
<i>Arc discharge</i>	0,64	1,00	0,78
<i>High-temperature overheating</i>	0,60	0,27	0,37
<i>Low-temperature overheating</i>	1,00	0,71	0,83
<i>Low/middle-temperature overheating</i>	0,00	0,00	0,00
<i>Middle-temperature overheating</i>	0,12	0,33	0,18
<i>Partial discharge</i>	1,00	0,50	0,67
<i>Spark discharge</i>	0,92	0,69	0,79

Fonte: elaborada pelo autor.

Por fim o modelo performado com NB para o *dataset* das razões obteve valores altos de precisão para *low-temperature overheating* (1,00), *partial discharge* (1,00) e *spark discharge* (0,92), porém para as classes *middle-temperature overheating*, *low/middle-temperature overheating*, *high-temperature overheating* os valores obtidos foram baixos, respectivamente, 0,12, 0,0, e 0,60. No *recall*, os maiores valores foram das classes *arc discharge* (1,00), *spark discharge* (0,69) e *low-temperature overheating* (0,71); já os menores, se concentraram nas classes *high-temperature overheating* (0,27), *low/middle-temperature overheating* (0,00) e *middle-temperature overheating* (0,33).

Quando analisamos a performance global dos classificadores, por meio da acurácia, percebemos que a MLP e o KNN foram os modelos que obtiveram os melhores resultados (0,70) em comparação ao *Naïve Bayes* (0,63), como disposto na tabela 13. Comparado com os resultados obtidos para o *dataset* normalizado (tabela 09), percebemos um incremento global considerável na acurácia dos métodos KNN e *Naïve Bayes*, porém, a MLP se manteve com o mesmo valor.

Tabela 13 – Acurácia para cada classificador aplicado ao *dataset* das razões

<i>Classificador</i>	<i>Acurácia</i>
<i>K-nearest neighbor</i>	0,70
<i>Multilayer perceptron</i>	0,70
<i>Naïve Bayes</i>	0,63

Fonte: elaborada pelo autor.

Ao aplicar os métodos de classificação para uma amostra de 2048 registros,

preprocessado como razões de Rogers, Ikeshoji (2020) identificou que utilizando KNN para K igual a 3, a acurácia era de 0,783. Para *Naïve Bayes*, obteve-se um valor de 0,753. Na MLP, obteve uma acurácia de 0,819. Valores próximos aos obtidos nesta pesquisa.

Já Liao et al (2021), ao comparar os resultados obtidos para diferentes metodologias frente ao GCN (*Graph Convolutional Network*), obteve acurácia de 0,672 para a implementação de uma MLP e 0,700 para o KNN, com K igual a 7, em uma amostra de tamanho igual a 718. Lima (2005), ao aplicar uma rede *multilayer perceptron*, com 440 amostras, classificando 4 tipos de falhas incipientes e casos normais obteve uma taxa de sucesso no diagnóstico de 0,9726.

Outros autores optaram por uma outra abordagem, utilizando redes neurais artificiais genéricas, como Huang (2003), Guardado et. al. (2001); ou ainda substituindo os dados comuns de uma análise de gases dissolvidos por dados físico-químicos, a exemplo de Barbosa et. al. (2009). As acurácias obtidas, se assemelham aos das pesquisas convencionais, respectivamente, 0,9354, 0,8788, e 0,9167.

4.3 Comparação entre os métodos computacionais e método tradicional

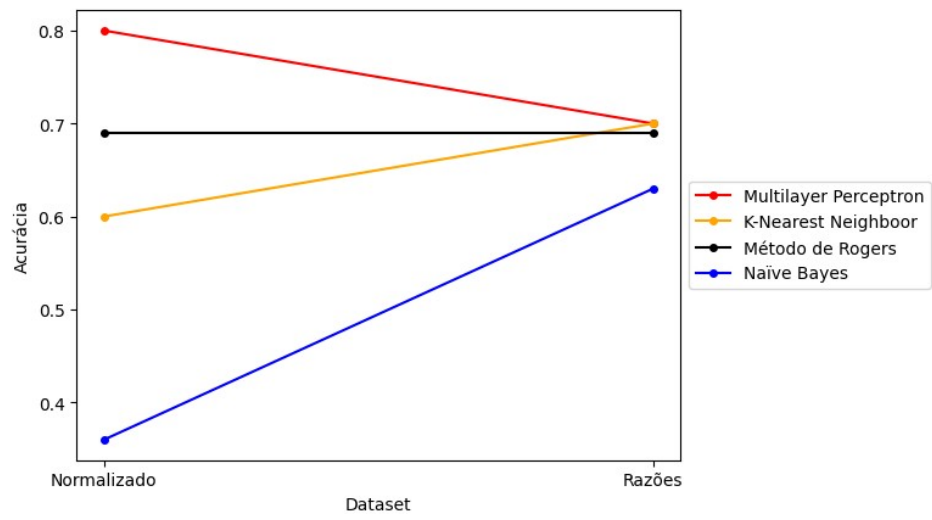
Quando comparamos os resultados obtidos por métodos computacionais, com a metodologia descrita por Rogers, observa-se que os métodos computacionais foram melhores em três das seis implementações, como descrito na tabela 14 e apresentado no gráfico 03.

Tabela 14 – Acurácia dos classificadores computacionais contra método de Rogers

<i>Classificador</i>	<i>Acurácia (Original)</i>	<i>Acurácia (Razões)</i>
<i>K-nearest neighbor</i>	0,60	0,70
<i>Multilayer perceptron</i>	0,80	0,70
<i>Naïve Bayes</i>	0,36	0,63
<i>Método de Rogers</i>	0,69	0,69

Fonte: elaborada pelo autor

Gráfico 03 – Classificadores computacionais contra método de Rogers



Fonte: elaborada pelo autor

5 CONCLUSÕES

Com os resultados obtidos é possível concluir que os classificadores implementados por métodos computacionais, *Multilayer perceptron* para o *dataset* original, *dataset* das razões e *K-nearest neighbor*, obtiveram melhores taxas de acerto, respectivamente 0,80, 0,70 e 0,70, que o método tradicional descrito por Rogers (0,69), sendo este superior apenas ao *Naïve Bayes* (0,63).

Observa-se também que usar o método das razões no pré-processamento dos dados da DGA trouxe ganhos consideráveis de acurácia para os classificadores *K-nearest neighbor* e *Naïve Bayes*, respectivamente +0,10 e +0,27, quando comparado ao *dataset* normalizado. Porém, no caso da MLP os ganhos foram negativos (-0,10).

No caso dos classificadores que obtiveram melhor desempenho global é preciso avaliar como eles performaram tendo em vista o problema analisado para tanto as métricas precisão e *recall* mostram que a MLP nos dois *datasets* responde melhor que o modelo implementado com KNN.

Tal fato tem como justificativa a quantidade de altos valores de precisão, nas classes *partial discharge*, *low-temperature overheating*, e *spark discharge*, onde falsos positivos são considerados mais prejudiciais que falsos negativos. E a quantidade de altos valores de *recall*, nas classes *high-temperature overheating* e *arc discharge*, onde falsos negativos são considerados mais prejudiciais que os falsos positivos.

Por fim, mesmo que os resultados dos classificadores para as classes *low/middle-temperature overheating* e *middle-temperature overheating*, tenha sido inferior as outras classes avaliadas, pode-se afirmar que os objetivos estabelecidos no início dessa pesquisa foram alcançados de maneira satisfatória, uma vez que dois dos três classificadores avaliados foram melhores que o método tradicional descrito por Rogers.

Ademais, cabe a indagação em trabalhos futuros aspectos para além das concentrações de gases no processo de classificação de falhas incipientes, como as propriedades físico-químicas do óleo ou até mesmo o histórico de cromatografias para diferentes transformadores.

REFERÊNCIAS

- AMO, S.; ROCHA, A.R. **Mining Sequential Patter using Genetic Programming**. In: International Conference on Artificial Intelligence, Las Vegas, 2003 p.451-456
- ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS – ABNT. NBR 7274: **Interpretação da Análise dos Gases de Transformadores em Serviço**. Rio de Janeiro, RJ, 1982.
- AZEVEDO, C. H. B. **Metodologia para eficácia da detecção de descargas parciais por emissão acústica como técnica preditiva de manutenção em transformadores de potência imersos em óleo isolante**. Dissertação de Mestrado, GPT/BC/UFG – Programa de Pós-Graduação em Engenharia Elétrica e de Computação / Universidade Federal de Goiás, Goiânia, GO, 2009.
- BARBOSA, Fábio Rocha. **DIAGNÓSTICO DE FALHAS INCIPIENTES A PARTIR DAS PROPRIEDADES FÍSICO-QUÍMICAS DO ÓLEO ISOLANTE EM TRANSFORMADORES DE POTÊNCIA COMO MÉTODO ALTERNATIVO À ANÁLISE DE GASES DISSOLVIDOS**. 2013. 133 f. Tese (Doutorado) - Curso de Engenharia Elétrica, Departamento de Engenharia Elétrica, Universidade Federal do Ceará, Fortaleza, 2013.
- BARBOSA, F. R. **Monitoramento on-line e diagnóstico inteligente da qualidade dielétrica do isolamento líquido de transformadores de potência**, Dissertação de Mestrado, UFC/CT/PPGDEE –Programa de Pós-Graduação do Departamento de Engenharia Elétrica, Fortaleza, CE, 2008.
- BENGTSSON, C. **Status and Trends in Transformer Monitoring**. IEEE Transactions on Power Delivery, v. 11, n. 3, pp. 1379-1384, 1996.
- BETTA, G.; PIETROSANTO, A.; SCAGLIONE, A. **An Enhanced Fiber-Optic Temperature Sensor System for Power Transformer Monitoring**. IEEE Transactions on Instrumentation and Measurement, v. 50, n. 5, pp. 1138-1143, 2001.
- BUENO, Michel Ferreira; VIANA, Maury Reis; **Mineração de Dados: aplicação, eficiência e resultados**. In: Congresso de iniciação científica do INATEL – INCITEL, 2012, Santa Rita do Sapucaí. Anais..Minas Gerais: INATEL-INCITEL, 2012. p. 86-88.
- CAMILO, C. O.; SILVA, J. C. da. **Mineração de Dados: Conceitos, Tarefas, Métodos e Ferramentas**. Relatório Técnico – Universidade Federal de Goiás. Goiás: Universidade Federal de Goiás, 2009. 29 p.
- CASTILLO, E.V. **Aplicação de ontologia e sistema especialista para diagnóstico de falhas em transformadores de potência**. Dissertação (Mestrado em Engenharia Elétrica) – Universidade Federal de Santa Catarina, Florianópolis, 2003.

DING, H.; HEYWOOD, R.; LPWORTH, J.; RYDER, S. **Learning from sucess and failure in transformer fault gas analysis and interpretation**. IET conference on reliablity of transmission and distribution Network (RTDN 2011), pp 1-6, 2011.

DÖRNENBURG, E.; GERBER, O. E. **Analysis of Dissolved and Free Gases for Monitoring Performance of Oil-Filled Transformers**. The Brown Boveri Review, v. 54, n. 213, pp. 104-111, 1967.

DUVAL, M. **Dissolved Gas Analysis: It Can Save Your Transformer**. IEEE Electrical Insulation Magazine, v. 5, n. 6, p. 22-27, 1989.

FIX, E.; HODGES, J. L. **Discriminatory analysis, nonparametric discrimination: consistency properties**. Randolph Field, Texas: Usaf School Of Aviation Medicine, 1951.

FONSECA, Jackson Farias. **Elementos de manutenção de transformadores de potência**. 2014. 97 f. Trabalho de conclusão de curso (Bacharelado Engenharia Elétrica) - Universidade Estadual Paulista, Faculdade de Engenharia de Guaratinguetá, 2014. Disponível em: <<http://hdl.handle.net/11449/123036>>.

GALLAGHER, C; MADDEN, M.G.; D'ARCY, B. **A Bayesian Classification Approach to Improving Performance for a Real-World Sales Forecasting Application**. In: 2015 IEEE 14th International Conference on Machine Learning and Applications, Florida, 2015.

GEORGILAKIS, P. S. **Environmental cost of distribution transformer losses**. Applied Energy, v.88, p.3146-3155, 2011.

GEORGILAKIS, P. **Spotlight on Modern Transformer Design**. London: Springer-Verlag, 2009.

GOMEZ, N.A.G. **Diagnóstico de falhas em transformadores isolados com óleo vegetal isolante a partir da análise de gases dissolvidos**. Dissertação (Mestrado em Engenharia dos Materiais) – Setor de Tecnologia. Universidade Federal do Paraná, Curitiba, 2013.

HAN, J.; KRAMER, M. **Data mining Concepts and Techniques**. 2nd ed. California: Morgan Kaufmann San Francisco, 2006.

IEEE Guide for the Interpretation of Gases Generated in Oil-Immersed Transformers, IEEE Std C57.104-1991, 2008.

INTERNATIONAL ELECTROTECHNICAL COMMISSION – IEC 60599: **Mineral Oil-Impregnated Electrical Equipment in Service – Guide to the Interpretation of Dissolved and Free Gases Analysis**. Geneva, Switzerland, 2. ed., 1999.

IKESHOJI, Marco Akio. **DESENVOLVIMENTO DE METODOLOGIA PARA DIAGNÓSTICO DE FALHAS INCIPIENTES EM TRANSFORMADORES DE POTÊNCIA UTILIZANDO ANÁLISE DE GASES E APRENDIZADO DE MÁQUINA**. 2020. 159 f. Tese (Doutorado) - Curso de Engenharia Elétrica, Faculdade de Engenharia, Universidade Estadual Paulista (Unesp), Bauru, 2020. Disponível em: <http://hdl.handle.net/11449/193553> . Acesso em: 16 ago. 2022.

KOSOW, Irving Lionel. **Máquinas Elétricas e Transformadores**. Porto Alegre: Globo, 1982. 689 p.

KOVACEVIC, S.; DOMINELLI, N. **Transformer Life Extension by Continuous On-Line Oil Conditioning**. Proceedings of Electrical Insulation Conference and Electrical Manufacturing & Coil Winding Technology Conference, pp. 23-25, 2003.

LI, Ewen. “**Dissolved gas data in transformer oil: Fault Diagnosis of Power Transformers with Membership Degree**”, IEEE Dataport, 14 de janeiro, 2019, Disponível em: <https://dx.doi.org/10.21227/h8g0-8z59>.

LIAO, Wenlong et al. Fault Diagnosis of power Transformers using graph convolutional network. **CSEE Journal Of Power And Energy System**. Beijing, p. 241-249. mar. 2021.

MILASCH, M. **Manutenção de Transformadores em Líquido Isolante**. Edgar Blücher, São Paulo, SP, 1984.

MITCHELL, T.M. **Machine Learning**. 1.ed. [S.1]: McGraw-Hill Science/Engineering/Math, 1997.

MORAIS, D. R. **Ferramenta Inteligente para Detecção de Falhas Incipientes em Transformadores Baseada na Análise de Gases Dissolvidos no Óleo Isolante**. Dissertação de Mestrado. UFSC, Florianópolis, SC, 2004.

MORAIS, D. R.; ROLIM, J. G. **A hybrid tool for detection of incipient faults in transformers based on the dissolved gas analysis of insulating oil**, IEEE Transactions on Power Delivery, v. 21, n. 2, pp. 673-680, 2006.

NIEUWENHOFF, Nathalia. **Uma comparação da aplicação de métodos computacionais de classificação de dados aplicados ao consumo de cinema no Brasil**. 2017. 131 f. Dissertação (Mestrado) - Curso de Sistemas da Informação, Escola de Artes, Ciências e Humanidades, Universidade de São Paulo, São Paulo, 2017.

PETERSON, Leif E. (comp.). K-nearest neighbor. **Scholarpedia**, [s. l.], v. 4, n. 2, p. 1883, 2009. Disponível em: [10.4249/scholarpedia.1883](https://doi.org/10.4249/scholarpedia.1883). Acesso em: 18 jun. 2022.

PUGH, P. S.; WAGNER, H. H. **Detection of Incipient Faults in Transformer by Gas Analysis**. AIEE Transactions, v. 80, p. 189-195, 1961.

ROGERS, R.R. **IEEE and IEC Codes to Interpret Incipient Faults in Transformers Using Gas in Oil Analysis**. IEEE Transactions Electrical Insulation, v. 13, n. 5, p. 349-354, 1978.

SENNA, S.L. **Computação evolucionária aplicada ao diagnóstico de falhas incipientes em transformadores de potência utilizando dados de cromatografia**. Dissertação (Mestrado em Engenharia Elétrica). Universidade Federal de Minas Gerais, Belo Horizonte, 2010.

SILVA, Ivan Nunes da; SPATTI, Danilo Hernane; FLAUZINO, Rogério Andrade. **Redes neurais artificiais para engenharia e ciências aplicadas: fundamentos teóricos e aspectos**