



UNIVERSIDADE FEDERAL DO PARÁ
INSTITUTO DE TECNOLOGIA
FACULDADE DE ENGENHARIAS ELÉTRICA E BIOMÉDICA
TRABALHO DE CONCLUSÃO DE CURSO

FABRÍCIO SILVA LIMA

**MODELO DE PREDIÇÃO DE TENSÃO CRÍTICA EM ISOLADORES DE
DISTRIBUIÇÃO.**

BELEM - PA

2021

FABRÍCIO SILVA LIMA

**MODELO DE PREDIÇÃO DE TENSÃO CRÍTICA EM ISOLADORES DE
DISTRIBUIÇÃO.**

Trabalho de Conclusão de Curso, apresentado ao curso de Graduação em Engenharia Elétrica da Universidade Federal do Pará, como requisito parcial para a obtenção do grau de Engenheiro Eletricista.

Orientador: Prof. Dr. Reinaldo Corrêa Leite.

BELÉM - PA

2021

AGRADECIMENTOS

Em primeiro lugar, a Deus, que fez com que meus objetivos fossem alcançados, durante todos os meus anos de estudos.

Agradeço aos meus pais e familiares em especial minha mãe, Núbia Cristina de Oliveira Silva, meu pai, Robson Pinheiro Lima júnior e ao meu padrastrô, Antônio Tadeu Garcia Reis que muito me incentivaram de todas as formas a trilhar esse caminho, seja na parte financeira como na parte mental.

Agradeço aos meus amigos Roberto Bruno de Oliveira Rocha e William Johnson Silveira Martins que estiveram em todos os momentos dessa caminhada me incentivando e apoiando para que pudesse chegar ao fim dessa jornada.

Agradeço à Ana Letícia Viera Braga e a Maria Lizi Vieira Soares pelos momentos de incentivos, inspiração e alegria que trouxeram ânimo e motivação para que pudesse chegar ao fim dessa fase.

E por fim, agradeço ao professor e Dr Reinaldo Correa Leite por ter me acolhido e incentivado a realização desse trabalho, sempre transmitindo, apoio, conhecimento e paciência.

RESUMO

Este trabalho trata-se de desenvolver um modelo em machine learning (ML) visando a melhora de técnicas de detecção preditiva de falhas de isoladores elétricos na rede de distribuição, buscando encontrar o melhor desempenho desses dispositivos no setor elétrico brasileiro.

Inicialmente é realizado um estudo teórico sobre os tipos de isoladores elétricos encontrados na rede de distribuição, suas características, formato físico e classe de tensão. Em seguida, trata-se dos diferentes tipos de poluição que esses isoladores podem sofrer durante sua vida útil e as consequências, como aparecimento de descargas parciais que podem se desenvolver para o fenômeno de flashover causando problemas como desligamento da rede e altos custos de manutenção.

Posteriormente, o trabalho trata de algoritmos de inteligência artificial, explicando os seus conceitos e a forma como atuam nos conjuntos de dados que recebem. O trabalho limita-se em algoritmos de aprendizado de máquina supervisionado. Com isso, detalha-se como podem ser aplicados para solução do trabalho em questão.

Por fim, aplica-se os algoritmos estudados no conjunto de dados referente as condições climáticas, corrente de fuga e tensão crítica para estabelecer resultados testes que possam prever futuros problemas causados pelo fenômeno de flashover nos isoladores de distribuição.

Palavras-Chave: Isoladores elétricos, Rede de distribuição, poluição, tensão crítica, machine learning, algoritmos de inteligência artificial.

ABSTRACT

This paper aims to develop a model in machine learning (ML) aiming to improve predictive detection techniques for failures of electrical insulators in the distribution network, seeking to find the best performance of these devices in the Brazilian electrical sector.

Initially, a theoretical study is carried out on the types of electrical insulators found in the distribution network, their characteristics, physical format and voltage class. Then, the different types of pollution that these insulators can suffer during their useful life are discussed, and the consequences, such as the appearance of partial discharges that can develop into the flashover phenomenon, causing problems such as networks shutdown and high maintenance costs.

Subsequently, the work deals with artificial intelligence algorithms, explaining their concepts and the way they act on the datasets they receive. The work is limited to supervised machine learning algorithms. Thus, it is detailed how they can be applied to solve the work in question.

Finally, the studied algorithms are applied in the dataset referring to weather conditions, leakage current and critical voltage to establish test results that can predict future problems caused by the flashover phenomenon in distribution insulators.

Key words: electrical insulators, distribution network, pollution, critical tension, machine learning, artificial intelligence algorithms

LISTA DE FIGURAS

Figura 2.1 – Isolador polimérico classe 35KV	15
Figura 2.2 – Isoladores LSR e HTV respectivamente	16
Figura 2.3 – Modelos de isoladores de porcelana	17
Figura 2.4 – processo de limpeza em isoladores de porcelana	18
Figura 2.5 – Isoladores disco de vidro.	18
Figura 2.6 – Isolador em poluição inerte: caulim, argila e óleo	22
Figura 2.7 – Presença de camada de poluição na saia do isolador	22
Figura 2.8 – Processo de formação de descargas parciais	23
Figura 2.9 - Microdinâmica de formação do Flashover	24
Figura 2.10 – Efeito corona em isoladores poliméricos	25
Figura 3.1 – Diagrama de modelo de classificação	29
Figura 3.2 – Verificação da complexidade do modelo treinamento e test (sub-ajuste e sobreajuste)	31
Figura 3.3 – Representação gráfica da regressão linear simples	32
Figura 3.4 – Representação gráfica da regressão linear múltipla	33
Figura 3.5 – Árvore de decisão para escolha de jogo de tabuleiro	34
Figura 3.6 – Modelo de funcionamento do algoritmo Random Forest	36
Figura 4.1 – Adição de bibliotecas na plataforma Jupyter notebook	41
Figura 4.2 – Importação dos dados para leitura	41
Figura 4.3 – leitura das cinco primeiras amostras de dados	42
Figura 4.4 – quantidade de valores lidos do arquivo	42
Figura 4.5 – Verificação de valores nulos	42
Figura 4.6 – Função de densidade dos dados	43
Figura 4.7 – Normalização e padronização das variáveis	44
Figura 4.8 – Cálculos de dados estatísticos	44
Figura 4.9 – Boxplots das variáveis do dataframe	45
Figura 4.10 – correlação das variáveis	46
Figura 4.11 – correlação entra as variáveis predictoras e a variável alvo (Uc)	46
Figura 4.12 – Aplicação do modelo Regressão linear e cálculo do seu score)	47
Figura 4.13 – – Acurácia do modelo com todas as variáveis predictoras	48
Figura 4.14 – score retirada das variáveis que possuem correlação menor que 0,5	48
Figura 4.15 – score retirada das variáveis que possuem correlação menor que 0,3	49
Figura 4.16 – score retirada das variáveis que possuem correlação menor que 0,1	49
Figura 4.17 – score do modelo com a retirada da variável corrente de fuga	49

Figura 4.18 – Comparativo dos scores utilizando a regressão linear	50
Figura 4.19 – separação de dados para treino e teste	50
Figura 4.20 – Inclusão da biblioteca <i>tree</i> e ajustes dos parâmetros para predição	51
Figura 4.21 - Cálculo do score para o modelo	51
Figura 4.22 – Modelo random forest e cálculo dos scores	51
Figura 4.23 – Criação de parâmetros a serem testados	52
Figura 4.24 – Execução do Grid	52
Figura 4.25 – Verificação dos melhores parâmetros	53
Figura 4.26 - Cálculo do score com os melhores parâmetros.	53
Figura 4.27 – Comparação entre os modelos simulados	54
Figura 4.28 – Comparação do valor real e valor previsto	55

SUMÁRIO

CAPÍTULO 01 – INTRODUÇÃO	10
1.1 – Justificativa	10
1.2 – Motivação	11
1.3 – Objetivos	11
1.3.1 – Objetivo Geral	11
1.3.2 – Objetivos Específicos	12
1.4 - Estrutura Organizacional do Trabalho	12
CAPÍTULO 2 – ISOLADORES DE DISTRIBUIÇÃO	14
2.1 – Introdução	14
2.2 – Tipos de isoladores de distribuição	15
2.3 – Poluição em isoladores	20
2.4 - Modelagem de descargas em isoladores poluídos	22
CAPÍTULO 3 – TÉCNICAS DE APRENDIZADOS DE MÁQUINA	27
3.1 – Introdução	27
3.2 – Aprendizagem supervisionada	27
3.2.1 – Classificação e regressão	28
3.2.2 – Generalização, sobre-ajuste e sub-ajuste	30
3.3 – Modelo de aprendizagem Supervisionada	31
3.3.1 – Regressão Linear	31
3.3.2 - Árvore de decisão;	33
3.3.3 – Random Forest	35
CAPÍTULO 4 – MODELAGEM DA DESCARGA SUPERFICIAL DE UM ISOLADOR POR MARCHINE LEARNING	37
4.1 – Cálculo da tensão crítica em função das condições climáticas	37
4.2 – Metodologia para criação de modelos utilizando aprendizado de máquina	40
4.2.1 – Tratamento de dados	40

4.2.2 – Análise e exploração de dados	42
4.3 – Descrição e comparação de modelos de aprendizados de máquina	47
4.3.1 – Modelo Regressão Linear	47
4.3.2 – Modelo Arvore de decisão	50
4.3.3 – Modelo Random Forest	51
4.3.4 – Random Forest com ajuste dos hiperparâmetros	52
4.4 – Resultados	54
REFERÊNCIAS BIBLIOGRÁFICAS	57

1 INTRODUÇÃO

1.1. JUSTIFICATIVA

A energia elétrica é um dos principais pilares que movem a sociedade atual e a partir disso deve ser tratada com tal importância desde a sua geração até a sua distribuição, passando pela sua transmissão. Um dos dispositivos importantes que fazem a energia chegar até o uso diárias das pessoas são os isoladores elétricos que tem grande funcionalidade principalmente nas etapas de transmissão e distribuição, com objetivos mecânicos (sustentação dos cabos condutores de energia) e elétricos (limitação de passagem de corrente entre equipamentos e cabos condutores energizados) esses isoladores desempenham um papel importante no sistema elétrico de energia, onde grande parte desses dispositivos ficam expostos ao tempo.

Uma variedade de fatores climáticos pode afetar o grau de suportabilidade elétrica do isolador, entre os principais parâmetros estão: chuva, vento, umidade, pressão entre outros que serão abordados durante o estudo feito para realização de diagnóstico antecipado de falha em isoladores elétricos de distribuição.

As falhas referentes ao isolador elétrico de distribuição representam pequenas cavidades na sua estrutura física interna, a quais são submetidas a pequenas descargas parciais que com o tempo levam a isolação ao colapso da sua capacidade dielétrica, dessa forma o isolador perde a sua funcionalidade no sistema. BEZERRA [2014]

Segundo STEFENON [2015] Atualmente, existem técnicas relacionadas a detecção de defeitos de isoladores no sistema de distribuição de energia, entre elas estão os instrumentos que detectam o ultrassom, técnica que é mais usada para realização dessas inspeções por serem mais eficazes. Porém, essa técnica depende do grau de conhecimento do operador, pois baseia-se na “tradução” dos sinais ultrassônicos em sinais audíveis, tornando assim essa técnica subjetiva e refém de técnicos experientes em trabalhos com isoladores em campo.

A exposição desses isoladores ao tempo, diminui a sua vida útil, sendo necessário a realização de reparos e manutenção ou em até casos mais prejudiciais uma troca da sua cadeia toda. Devido a esse tipo de defeito, o presente trabalho pretende desenvolver um modelo preditivo da Tensão crítica de flashover nos isoladores usando técnicas de Aprendizado de Máquina (Machine Learning).

A partir da combinação de algoritmos e big_data, a ferramenta realiza a análise de grandes volumes de dados, cria padrões, estabelece conexões e aprende a executar tarefas de forma inteligente, sem a intervenção do homem.

Nesse contexto, é importante destacar que a tecnologia pode ser dividida em duas categorias principais — a supervisionada e a não supervisionada. Classifica-se como algoritmos supervisionados aqueles que dependem da intervenção do homem para controlar a entrada e saída de dados, bem como o treinamento das máquinas. Já na modalidade não supervisionada, o processamento das atividades, desde as mais simples até as mais complexas é composta exclusivamente pelo algoritmo utilizado. CRONAPP [2019]

Com auxílio de alguns parâmetros que avaliam o seu bom funcionamento, esses métodos são algoritmos que preveem através de dados de treinamento e teste o grau da tensão crítica a que o isolador está sendo submetido e o impacto das condições climáticas atuais e do valor da corrente de fuga que ele está apresentando. Os algoritmos que serão apresentados ao longo deste trabalho são: Regressão Linear, Árvore de decisão e Radom Forest.

1.2. MOTIVAÇÃO

O isolador elétrico é um dispositivo de grande importância no sistema elétrico. Devido a isso requer um grau de proteção da sua vida útil mais eficiente, visto que os mesmos estão localizados em áreas externas, propícios a variações climáticas que afetam o seu funcionamento ao longo tempo por meio da poluição. A poluição nos isoladores acontece através de contaminantes que podem ter sido gerados por indústrias, regiões ambientais e/ou marítimas. Os depósitos, se aliados à umidade, propiciam o surgimento de descargas superficiais, e podem progredir para a ocorrência de arcos de potência ao longo da cadeia de isoladores, ocasionando interrupções no fornecimento de energia. Dessa forma, faz-se necessário a criação de métodos e técnicas que possam ser usados para monitoramento dos isoladores durante o seu funcionamento e com o aumento das pesquisas em inteligência artificial, a comunidade de engenharia de manutenção apropriou-se desse mecanismo para que se pudesse auxiliar na busca de agregar novas tecnologias para modernização, objetivando o crescimento do setor elétrico brasileiro.

1.3. OBJETIVO

Este presente trabalho tem como finalidade realizar um estudo computacional a partir das condições climáticas e da corrente de fuga de isoladores do tipo pilar polimérico do tipo High Temperature Vulcanizing (HTV). Dessa forma, obtém-se através de cálculo os valores

referentes a tensão crítica de flashover, criando assim um banco de dados para geração dos modelos em Machine Learning (ML). Em seguida, criam-se os modelos com algoritmos de classificação (regressão, árvore de decisão e Random Forest) com o objetivo de fazer uma análise preditiva referente às condições que o isolador apresenta em seu funcionamento para detectar futuras falhas nesse dispositivo através dos valores treinados e testados pelo modelo de algoritmos em de Aprendizado de Máquina.

1.3.1. OBJETIVO ESPECÍFICOS

- Criar modelo usando técnica de aprendizado de máquina para prever futuras falhas em isolador elétrico, a partir de banco de dados disponibilizados de condições climáticas, corrente de fuga e tensão de flashover.
- Calcular tensão de flashover através dos valores de corrente de fuga disponibilizados no banco de dados
- Comparar os modelos de classificação baseados em ML, para verificação de qual se adequa melhor como solução para o problema proposto.
- Monitorar quais os parâmetros que mais possuem correlação com as falhas em isoladores
- Propor uma nova técnica a ser usada na prevenção contra falhas em isoladores

1.4. ESTRUTURA ORGANIZACIONAL DO TRABALHO

O presente trabalho de conclusão de curso, apresenta a seguinte estrutura:

No capítulo 1, apresenta-se a introdução sendo composta por: motivação, justificativas, objetivos gerais e específicos do trabalho e a estrutura organizacional.

No capítulo 2 demonstrou-se os tipos de isoladores utilizados na rede de distribuição, as condições de poluição que o isolador sofre durante sua vida útil e por fim o modelo de descargas parciais em isoladores poluídos.

No capítulo 3 trata-se das técnicas de aprendizado de máquina que serão utilizadas no modelo a ser criado, primeiramente descrevem-se alguns conceitos como: aprendizado supervisionado e não supervisionado, algoritmos de classificação e regressão linear, generalização de sub-ajuste e sobre ajuste. Em seguida descrevem-se os modelos a serem utilizados são: regressão linear, árvore da decisão e random forest.

O capítulo 4 apresenta primeiramente o cálculo de tensão de flashover do isolador a partir dos valores de corrente de fuga e em seguida realiza a modelagem dessa tensão por Machine Learning, apresentando a metodologia desenvolvida, descrição dos modelos, análise exploratória de dados e por fim a apresentação de resultados obtendo os valores reais dos valores previsto e realizando a comparação do modelo que melhor se adequa a esse tipo problema.

2. ISOLADORES DE DISTRIBUIÇÃO E AS DESCARGAS PARCIAIS

2.1. INTRODUÇÃO

A isolação é um fator importante em qualquer meio na qual é utilizada a energia elétrica. Os componentes do sistema de potência tais como: cabos, equipamentos, instrumentos de medição e dispositivos elétricos. Com a finalidade de confinar o campo elétrico gerado pela tensão aplicada aos condutores, a isolação necessita ser vista de uma forma mais cuidadosa, pois qualquer falha em sua composição pode gerar sérios problemas seja no funcionamento de equipamentos ou na passagem de corrente elétrica nos cabos. Portanto, define-se o isolante elétrico segundo ENERGY [2019] como sendo toda a substância de muito baixa condutividade elétrica, tal que a capacidade de conduzir a corrente elétrica se torna desprezível, permitindo isolar eletricamente os condutores elétricos entre si e em relação à terra, e assegurar de forma suficiente a isolação entre os condutores e as partes metálicas do sistema.

Dentre os sistemas isolantes, há os isoladores elétricos de rede de distribuição de energia, que possuem importância considerável no sistema, visto que possuem duas funções, seja no âmbito mecânico, em que é responsável pela sustentação dos cabos dos ramais e a outra e principal função, de isolação dos condutores com relação às partes do sistema de energia. Esses isoladores podem ser classificados de acordo com as suas principais características, que são: resistência de isolamento, rigidez dielétrica, propriedades mecânicas (resistência à tração, compressão e à flexão), resistência a choque, propriedades físico-químicas, químicas e térmicas.

Conforme BONOMO [2003] a forma dos isoladores no projeto de fabricação é de tais formas que as linhas de forças do campo elétrico dentro da sua superfície se distribuem com maior homogeneidade, garantindo assim ao isolador melhor desempenho quanto a sua rigidez dielétrica. Para MELLO [2017] Devido às características dos isoladores, existe uma variedade de tipos que podem ser usados em situações diferentes, dependentes da região onde serão instalados. Isso se deve ao fato de que cada região possui condições climáticas diversas, afetando dessa forma o grau de poluição ao qual o isolador pode ser submetido ao longo da vida útil, e conseqüentemente o aumento das descargas parciais sobre sua superfície ocasionando, em casos mais críticos, a ruptura da sua tensão crítica para formação do flashover.

Neste capítulo serão abordadas primeiramente as características dos isoladores de acordo com o seu tipo, característica, dimensionamento, desempenho, em seguida um estudo

acerca de poluição em isoladores, como essa camada poluída é formada na superfície dos isoladores, quais as consequências que isso pode causar no desempenho final e por fim será mostrado a situação mais crítica que o isolador pode chegar que é a presença de descargas parciais e a formação de flashover.

2.2. TIPOS DE ISOLADORES DE DISTRIBUIÇÃO

2.2.1. ISOLADOR POLIMÉRICO – 35 KV

O Isolador Polimérico, mostrado na figura 2.1, é utilizado em redes de distribuição de energia elétrica com cabos nus ou cobertos, classe de tensão 35 kV. É fabricado em polietileno de alta densidade na cor cinza, e foi desenvolvido exclusivamente para se obter excelentes características mecânicas e atender aos requisitos de resistência aos raios ultravioleta, ao trilhamento elétrico e às intempéries exigidas neste tipo de rede.



Figura 2.1 – Isolador polimérico classe 35KV

Fonte: PLP Brasil

Outro modelo muito utilizado de isolador polimérico para rede de distribuição é o do tipo pilar polimérico. Isolador que será utilizado no desenvolvimento deste trabalho. O isolador pilar polimérico 35 kV do tipo JS é formado por um bastão de fibra de vidro fixado às ferragens integrantes e em seguida encapsulado com borracha de silicone, concedendo assim, ao bastão um revestimento isolante e totalmente vedado contra entrada de umidade e partículas no isolador. Segundo CABOLT [2017] há dois tipos de borrachas de silicone em uso para fabricação desse modelo de isolador que são mostrados na figura 2.2. A borracha de silicone vulcanizada de alta temperatura (HTV) e a borracha de silicone líquida (LSR) possuem

excelentes propriedades mecânicas no espaço de produtos de elastômero para uso em moldagem por compressão, extrusão e moldagem por injeção.



Figura 2.2 – Isoladores LSR e HTV respectivamente

Fonte: Adantas comercial e Silicones SR

Atualmente existem alguns fabricantes brasileiros que utilizam cada vez mais esses tipos de borrachas na confecção de isoladores. Uma dessas fabricantes é a STC SILICONES que acredita que a borracha de silicone garante uma alta estabilidade térmica em diferentes tipos de ambiente. Ainda o ensaio em revestimento de borracha de silicone utilizado em isoladores poliméricos pode ser submetido a diferentes tipos de estresses, tais quais são: envelhecimento, névoa salina, alta e baixa temperatura, raios UV e imersão em água, garantindo através desses ensaios que o isolador pode ser usado nessas condições de estresse.

2.2.2. ISOLADORES DE PORCELANA

Segundo BONOMO [2003] O isolador de porcelana é hoje um dos tipos de isoladores mais utilizado no sistema de distribuição elétrico brasileiro, dentre eles estão os modelos tipo: Pino (linha primária), Suspensão (linha primária) e Roldana (linha secundária). Suas características principais se dão pelo fato do seu bom desempenho em relação à intempérie, e apesar de apresentarem um menor custo de fabricação ainda sim, possuem características elétricas e mecânicas semelhantes aos demais visto no mercado.

QUEIROZ [2013] cita que a porcelana para fins elétricos consiste na mistura de vários minerais para formação de uma pasta, dos quais se destacam o quartzo, feldspato e argila. Cada um destes componentes tem uma influência predominante em aspectos térmicos, dielétricos:

- Quartzo: influencia o comportamento térmico da mistura, sendo que, quanto maior for a sua quantidade relativa, maior será a temperatura suportada pela porcelana
- Feldspato: determina o comportamento dielétrico do volume de porcelana através do aumento da rigidez dielétrica.
- Argila: — atribui maior capacidade de resistência a esforços de compressão em função da quantidade percentual presente na massa.

Os isoladores de porcelana, mostrados na figura 2.3 apresentam grande solidez, robustez à compressão, resistência a ataques químicos, e geralmente são quebradiços. Possuem grande energia superficial devido à quantidade de minerais ionizados, o que se traduz numa grande aderência de água e contaminantes presentes na atmosfera.



Figura 2.3 – Modelos de isoladores de porcelana

Fonte: Grupo ATS Elétrica

As vantagens do isolador de porcelana estão principalmente nas suas características externas e do material procedente da sua confecção, bem como a flexibilidade do seu comprimento que permite ajustar o tamanho das cadeias das linhas de transmissão, com isso regulando a distância de escoamento, aumentando do caminho que a corrente de fuga deve percorrer e por consequência um aumento da resistência durante a trajetória. Outra vantagem importante do isolador cerâmico é o processo de lavagem, mesmo sendo necessário fazer esse tipo de serviço de forma frequente, torna esse tipo de isolador mais vantajoso devido ao fato do mesmo não se danificar ao serem submetidos ao jato de água. Desta forma a limpeza se torna rápida e mais segura. A figura 2.4 mostra um isolador de porcelana sendo submetido a processo de lavagem.



Figura 2.4 – processo de limpeza em isoladores de porcelana

Fonte: Coluna consultoria, 2011

2.2.3. ISOLADOR DE VIDRO

O Isolador de vidro, como os mostrados na figura 2.5, é um modelo que tem se destacado mais em linhas de transmissão de alta tensão, embora, ainda hoje o sistema elétrico brasileiro ainda utilize muito o isolador do tipo porcelana nas linhas de distribuição, subestações entre outras aplicações. Porém, existem aplicações desse modelo em sistemas de transmissão e distribuição de energia. Segundo YOSHIMURA [2009] esses isoladores são caracterizados na rede de distribuição por denotarem maior confiabilidade de resistir em climas e graus de poluição onde os mesmos são instalados.



Figura 2.5 – Isoladores disco de vidro

Fonte: Qualidade online's Blog

Uma vez que a rigidez dielétrica dos isoladores de porcelana está diretamente relacionada com a qualidade da matriz de vidro que une as partículas, encontram-se semelhanças no comportamento elétrico com os isoladores de vidro. Ambos possuem baixas permeabilidades elétricas e a resistência à perfuração diminui com o aumento da temperatura. Apesar disso, como o vidro não possui descontinuidades na sua estrutura, tende a ser mais resistente à perfuração. O aumento da temperatura provoca ainda o aumento da condutividade superficial, de tal forma que, em casos de poluição elevada, a corrente de fuga que se estabelece entre os terminais pode causar erosões fortes na estrutura do material. Com efeito, os isoladores de vidro são mais suscetíveis a danos devido a atividades elétricas na sua superfície, com especial ênfase em contornamentos.

Em um experimento realizado por RODRIGUES et al [2017] utilizando uma substância hidrofóbica conhecida como verniz nanotecnológico na superfície do isolador de vidro, demonstrou-se que esses, apesar de possuírem uma resistência alta à corrente fuga, teve seu desempenho melhorado ao ser revestido por tal verniz, apresentando menores valores de correntes de fuga e uma menor redução da impedância aparente quando comparados às amostras sem verniz aplicado. Esse ensaio teve como objetivo reduzir gastos com a manutenção desse tipo de isolador, visto que, precisam de uma limpeza mais cuidadosa.

Há no mercado alguns produtos que minimizam o grau de poluição que os isoladores enfrentam durante a sua vida útil e muito desses produtos são compostos de substâncias hidrofóbicas, pois como será visto no capítulo 4 deste trabalho, a umidade tem uma correlação muito alta para com a tensão crítica, dessa forma, estudos buscam minimizar esse parâmetro de forma que os isoladores não sofram com grau de poluição na região onde serão instalados. No Brasil, algumas empresas comercializam esse tipo de substâncias hidrofóbicas para isoladores de vidro, uma delas é a ELETROMECHAN, fabricantes do segmento de Isoladores de Vidro Temperado, Poliméricos e de Porcelana para Linhas de Transmissão e Distribuição. Fabrica um isolador de vidro temperado com a tecnologia SEDICOAT que é um revestimento de silicone ao longo do perímetro do isolador. A camada de silicone proporciona propriedades hidrofóbicas à superfície do isolador de vidro temperado, o que aumenta a sua capacidade de resistir à poluição extrema e elimina a necessidade de lavagem

2.3. POLUIÇÃO EM ISOLADORES

Um dos grandes problemas que envolvem isoladores no sistema elétrico, sem dúvidas é a poluição, grande causadora de desligamentos e interrupções não programadas das linhas de energia. Os depósitos externos provenientes das condições climáticas e partículas de poluição representam uma ameaça às condições de operação da linha de distribuição porque afetam o nível de isolamento ou fiabilidade ao longo do tempo. Segundo QUEIROZ [2013] O envelhecimento dos isoladores e consequente perda de propriedades dielétricas deve-se à ação combinada de agentes de várias naturezas que ao longo do tempo afetam a estrutura dos equipamentos. Estas naturezas podem ser categorizadas segundo influências elétricas, mecânicas e ambientais. Por vezes as relações entre estes agentes são pouco perceptíveis devido à complexidade de intervenientes, tal como o grau de severidade que certas condições representam. Além disso, estão dependentes da qualidade e formato dos equipamentos utilizados.

HERNANZ et al [2016] realizaram um estudo sobre os principais tipos de isoladores e fontes de contaminantes que podem afetar a isolação em cadeia de isoladores. A tabela 2.1 mostra os dados referentes aos contaminantes que são depositados na superfície do isolador e as fontes de poluição que geram esses contaminantes.

Os poluentes são predominantes de origem industrial ou marítima. Os poluentes industriais, principalmente os de indústrias que fazem o consumo de combustíveis fósseis, como carvão e o diesel, e os de indústrias que fabricam fertilizantes, cimento e indústrias de refinaria de petróleo são responsáveis pela emissão dos contaminantes mais comuns encontrados nas saias dos isoladores.

Tabela 2.1 – Contaminantes e fonte de poluição de isoladores

Contaminantes	Fontes de Poluição
Sal	<ul style="list-style-type: none"> ● Regiões costeiras ● Industriais
Cimento	<ul style="list-style-type: none"> ● Fábricas de cimento ● Locais de construções, obras ● Pedreiras

Terra	<ul style="list-style-type: none"> ● Locais de plantios ● Estrada de terras
Fertilizantes	<ul style="list-style-type: none"> ● Fábricas de fertilizantes ● Locais de plantios
Metálicos	<ul style="list-style-type: none"> ● Manipulação de minerais ● Tratamento de minerais
Carvão	<ul style="list-style-type: none"> ● Processo de fabricação de carvão ● Combustão de carvão
Fezes	<ul style="list-style-type: none"> ● Locais próximos a migração de aves
Químicos	<ul style="list-style-type: none"> ● Produção de substâncias químicas, resíduos industriais, refinarias e entre outros
Fumaça	<ul style="list-style-type: none"> ● Emissão dos automóveis e caminhões, queima de processos industriais e queima em práticas agrícolas

Fonte: HERNANZ [2016]

A presença destes contaminantes sobre a superfície do isolador gera uma baixa condutividade do mesmo e conseqüentemente diminuindo a sua vida útil e a presença dessa camada de poluição possibilita a formação de descargas elétricas como as provocadas pelo efeito corona, que será abordado em tópico posterior, causando danos ao material dielétrico que constitui o isolador, tais como: perfurações, gelo e entre outros tipos de poluição. A figura 2.6 mostra isoladores em processo de poluição.

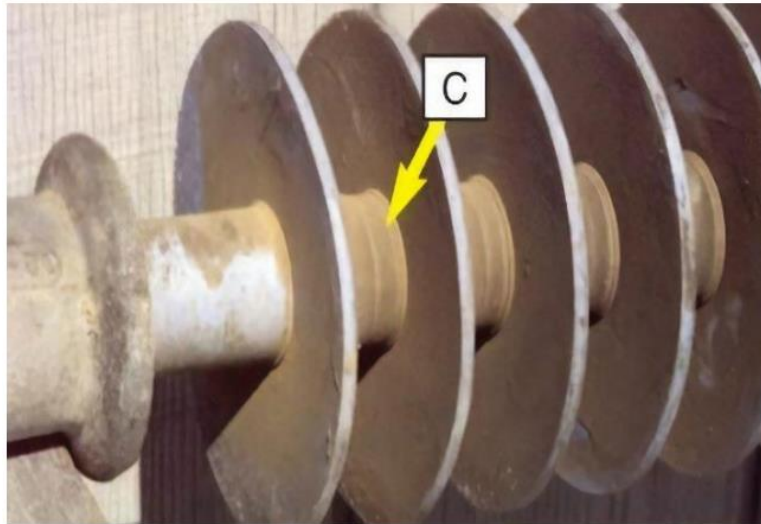


Figura 2.6 – Isolador em poluição inerte: caulim, argila e óleo

Fonte: EPRI (2006)



Figura 2.7 – Presença de camada de poluição na saia do isolador

Fonte: QUEIROZ (2013)

2.4. MODELOS DE DESCARGAS EM ISOLADORES POLUÍDOS

A alta concentração dos poluentes quando associados à umidade da região geram caminhos condutores que contribuem para o surgimento de descargas parciais (DP). As DP são definidas como uma descarga elétrica que ocorre em inclusões existentes em um isolante posicionado entre dois eletrodos condutores. Estas descargas não são capazes de causar disrupção total deste dielétrico. Nos isoladores é a DP é constituída da seguinte forma: com o

acúmulo de camada de poluentes na superfície dos dielétricos, essa se mantém seca, não ocorrendo alterações significativas na funcionalidade do isolador. Porém, quando a camada de poluição é umidificada de forma parcial pela umidade ou total pela chuva, neblinas ou até mesmo gelo, cria-se uma solução condutora, denominada de região úmida, conforme pode ser visto na figura 2.8.

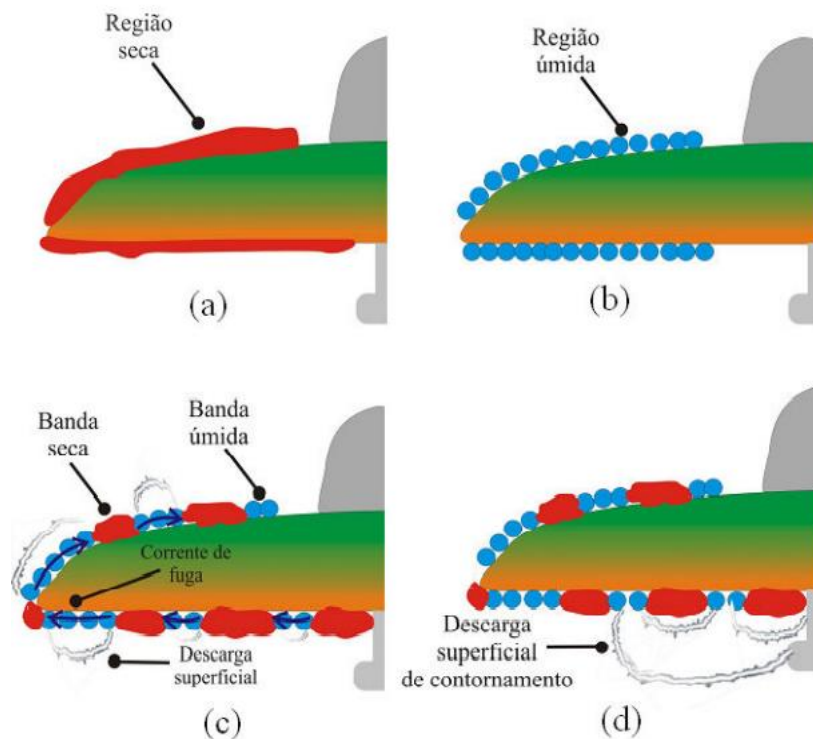


Figura 2.8 – Processo de formação de descargas parciais (a) Região de banda seca; (b) Região úmida; (c) descargas parciais na superfície; (d) descargas parciais contornando regiões secas e em contato com outros condutores

Fonte: Lima, 2010

. A formação destes arcos elétricos segue as seguintes etapas: primeiramente, correntes de ar levam poluição seca à superfície do isolador; a umidade, o orvalho, névoas e chuvas umedecem a superfície do isolador, que dissolve os contaminantes e forma um eletrólito condutor. Este eletrólito diminui a resistência na superfície e uma corrente de fuga é gerada; e a energia é dissipada pela corrente de fuga na forma de calor e o resultado é a evaporação da umidade. Devido à geometria do isolador a corrente tende a se concentrar em determinadas regiões (locais de menor secção reta), com isso, estas são secas mais rapidamente. Devido à secagem não uniforme na superfície do isolador são formadas bandas secas, interrompendo o fluxo de corrente ao longo da superfície. KARADY e SHAH [1995]; ABDELAZIZ et al., [2004].

As regiões de banda seca ficam submetidas a intensos campos elétricos, visto que a tensão de trabalho no isolador segundo LIMA [2010] é de ordem de KV e o comprimento da banda seca é de ordem de alguns centímetros, causando dessa forma uma ruptura dielétrica do ar na região ao redor da banda seca. Assim, as descargas parciais se manifestam contornando as regiões de banda seca, através do caminho iniciado no pino até as regiões de banda úmida ou entre bandas úmidas, estando eletricamente em paralelo com as bandas secas e eletricamente em série com a banda úmida da camada de poluição mais próxima. Essa microdinâmica dos isoladores pode ser vista na figura 2.9 a seguir, o processo pode se estender por toda a extensão da cadeia de isolador, causando o processo conhecido como flashover.

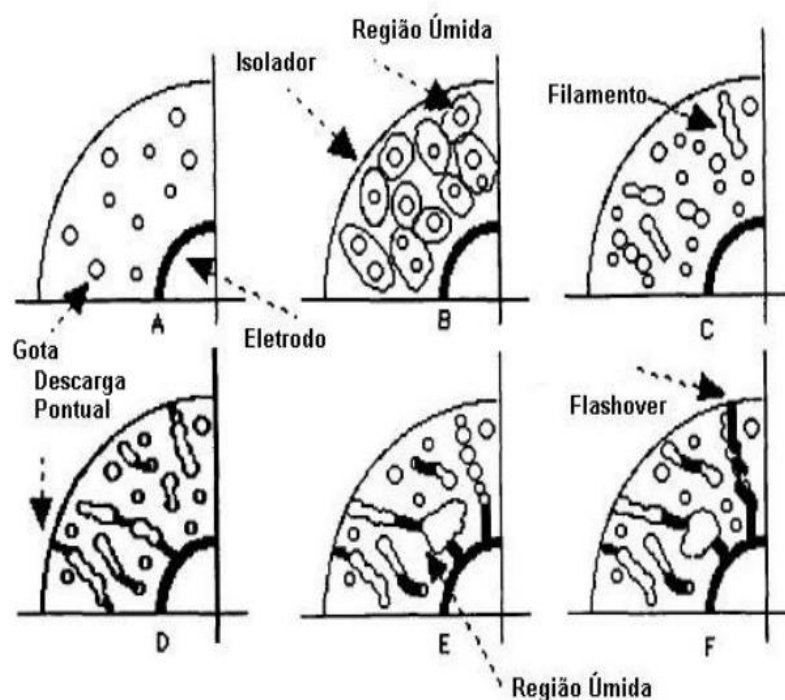


Figura 2.9 - Microdinâmica de formação do Flashover

Fonte: adaptado de Karady, 1995

Efeito Corona

Outro fenômeno a ser considerado quando se trata de descargas parciais é o efeito corona. Segundo SILVA [2011] o fenômeno mostrado na figura 2.10 é responsável pela origem de descargas disruptivas, este inicia-se quando há uma diferença de potencial aplicada entre eletrodos separados por um material dielétrico, acarretando um campo elétrico. À medida que este potencial aumenta, o campo elétrico por consequência aumenta, podendo de acordo com sua intensidade ultrapassar a rigidez dielétrica da isolação causando sua ruptura, ou disruptão.

Esta ruptura ocorre, normalmente no ar, devido à ação de um campo elétrico não uniforme, ou a não uniformidade presente na região do equipamento, e por fim dando o surgimento de descargas próximas ao intenso campo elétrico. Tecnicamente, em campo, as descargas ocorridas são arcos voltaicos entre condutores, ou partes energizadas, e partes aterradas da estrutura de sustentação das linhas de energia elétrica. Os estudos mostram que a energia da descarga elétrica necessária para causar a falha do isolador é diretamente proporcional ao grau de poluição em sua superfície, quando ocorre um arco voltaico. Portanto, o valor da corrente de fuga é um fator de grande influência no estado do arco, uma vez que, se o seu valor for muito elevado, a energia acumulada no processo de descarga será maior, necessitando de mais tempo para deionizar o meio. Neste caso, o arco não se extingue por completo quando a corrente atinge o valor nulo, formando um plasma que facilmente reacende a descarga no novo ciclo. Por outro lado, para valores mais reduzidos de corrente de fuga, mas suficientemente elevados para iniciar o processo, a descarga elétrica apresenta um comportamento intermitente, ou seja, extingue-se e reaparece.



Figura 2.10 – Efeito corona em isoladores poliméricos

Fonte: Gianelli B F. et al [2009]

Por fim, cada tipo de isolador apresenta características físicas e de operação distintas, cada um com suas vantagens e limitações. Sendo assim, cabe ao responsável de projetos de linhas de distribuição analisar qual será o melhor tipo de isolador para determinada região a ser instalada. Levando em consideração as condições climáticas, que como se pode verificar durante o capítulo são de extrema importância para o processo de degradação do isolador. Dessa forma, independentemente do tipo do isolador, sabe-se que todos estão susceptíveis a poluição e formação de descargas na sua superfície, assim cabe a ser verificado novos estudos que possam facilitar o processo de manutenção desses dispositivos sem causar tantos transtornos na rede.

3. TÉCNICAS DE APRENDIZADOS DE MÁQUINA

3.1 INTRODUÇÃO

Métodos baseados em inteligência de máquinas são usados atualmente para diversas situações do cotidiano, seja pela busca da simplificação do trabalho manual realizado pelo homem, como também a busca por soluções preventivas de problemas que poderiam ter um processo de manutenção bastante complexo. Os tipos de algoritmos de aprendizado de máquina de maior sucesso são aqueles que automatizam processos de tomada de decisão generalizando a partir de exemplos conhecidos. Neste cenário, que é conhecido como aprendizado supervisionado, o usuário fornece ao algoritmo pares de entradas e saídas desejadas, e o algoritmo encontra uma maneira de produzir a saída conhecida a partir de um dado valor de entrada. Em particular, o algoritmo é capaz de criar uma saída para uma entrada nunca vista antes sem qualquer ajuda de um humano.

Durante esse trabalho, serão apresentadas técnicas e definições de aprendizado de máquinas de classificação e regressão linear juntamente com seus modelos. Em seguida, será realizado um estudo sobre os ajustes de parâmetros que podem auxiliar na melhora da construção do modelo de machine learning (ML), que leve a uma melhoria no desempenho, que traga mais confiabilidade e uma precisão melhor nos resultados. E por fim, serão apresentados os algoritmos que vão ser testados no capítulo 4 como resultado do problema proposto. Esses algoritmos são métodos de aprendizado supervisionado e são chamados de regressão linear, árvore de decisão e random forest. Dessa forma, esse capítulo tem como finalidade mostrar as características dos algoritmos que por serem supervisionados, possuem características semelhantes no seu modelo, sempre com objetivo de fazer previsões precisas para dados novos nunca vistos.

3.2 – APRENDIZADO SUPERVISIONADO

Aprendizado de máquina supervisionado atualmente é o método mais conhecido no âmbito da linguagem de programação. Visto que, por ser um método capaz de gerar resultados futuros independentemente de como são gerados os valores de entrada, isso se torna ainda mais atrativo aos olhos do homem que sempre busca saber o resultado antes de colocar sua mão de obra ou investimento.

Para MARTINS [2017] o conceito de aprendizado supervisionado é um sistema que recebe um conjunto de dados com observações, sendo que cada uma das observações possui n características (do inglês, features) e uma etiqueta (do inglês, label) que determina o que significam aquelas características. Como analogia, pode-se comparar quando se estuda com livros-texto e tem exercícios para fazer e há uma página de soluções para esses exercícios no final do livro. Pode-se treinar com esses exercícios até que se torne apto para responder como o autor do livro nos ensina determinada disciplina. Depois disso, em algum momento há a necessidade de sermos testados para verificar se a “máquina humana” teve a capacidade de aprender. Ou seja, voltando à definição, o aprendizado de máquina tenta criar um modelo de dados de treinamentos e em seguida ser capaz de fazer previsões precisas sobre novos dados desconhecidos que possuem as mesmas características dos dados treinados. A partir disso, consegue-se generalizar o conjunto de treinamento para um conjunto de dados testes, sendo assim o grande objetivo do aprendizado de máquina supervisionado é prever com a máxima precisão possível resultados a partir de dados de entradas desconhecidas.

3.2.1 CLASSIFICAÇÃO E REGRESSÃO

Existem dois tipos principais de problemas de aprendizado de máquina supervisionado, chamados de classificação e regressão. Na classificação, o objetivo é prever um rótulo de classe, que é uma escolha de uma lista de possibilidades. A classificação pode ser separada em classificação binária, que é o caso especial de distinguir exatamente entre duas classes e a multiclasse, que é a classificação entre mais de duas classes. MULLER [2016]

Para CAETANO [2015] A construção de um modelo de classificação envolve um processo de dois estágios, sendo o primeiro o de aprendizado, onde o modelo de classificação é construído (ou treinado) a partir de um conjunto de dados, denominado conjunto de treinamento e, um de classificação, onde o modelo construído é utilizado para classificar novos dados (predição). O processo está ilustrado na Figura 3.1:

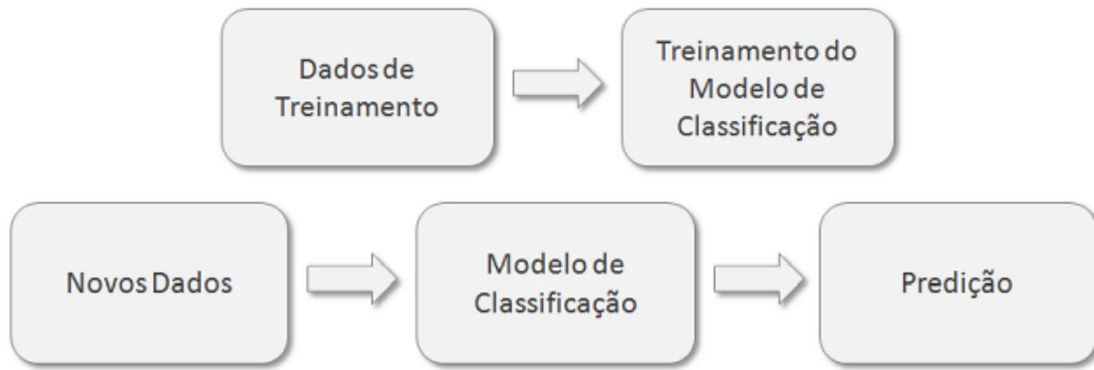


Figura 3.1 – Diagrama de modelo de classificação

Fonte – Fontana 2015

Em geral, para a construção do modelo é necessária uma amostra dos dados com todos os atributos que definem cada observação mais o atributo-classe que representa a resposta conhecida. O nome de classificação supervisionada vem justamente do fato de ter esse conjunto de dados para o qual a resposta já é conhecida.

Para o problema proposto neste trabalho, tem-se um conjunto de dados de variáveis referentes a condições climáticas, corrente de fuga e tensão crítica, dessas informações pode-se verificar os seguintes atributos e o atributo-classe, que são:

- Atributos: umidade, temperatura, pressão, ângulo do vento, velocidade do vento, chuva, corrente de fuga e tensão crítica
- Atributo-classe: atingiu ou não atingiu a tensão crítica

FONTANA [2015] afirma que um problema de classificação consiste, portanto, em encontrar uma relação entre um atributo-classe (atributo categórico ou quantitativo discreto) e demais atributos (categóricos ou quantitativos) através da construção de um modelo matemático ou estatístico, com base em dados contendo os valores dos atributos, e do atributo-classe (daí o nome aprendizado supervisionado).

Já para a regressão, segundo MULLER [2016] o objetivo é prever um número contínuo ou um ponto flutuante em termos de programação (ou número real em termos matemáticos). Por exemplo, prever uma renda anual da pessoa proveniente de sua educação, idade e onde vive é um exemplo de uma tarefa de regressão. Ao prever a receita, o valor previsto é uma quantidade e pode ser qualquer número em um determinado intervalo. Outro exemplo de regressão tarefa é prever o rendimento de uma fazenda de milho, dados atributos, como rendimentos anteriores,

clima e número de funcionários que trabalham na fazenda o rendimento novamente pode ser um número arbitrário. Uma maneira fácil de distinguir entre tarefas de classificação e regressão é perguntar se há algum tipo de continuidade na saída, se houver continuidade entre resultados possíveis, então o problema é um problema de regressão. Pense em prever rendimento anual. Há uma continuidade clara na saída. Se uma pessoa faz R\$40.000 ou R\$ 40.001 por ano não faz uma diferença tangível, mesmo que estes sejam diferentes quantias; se o algoritmo prevê R\$ 39.999 ou R\$ 40.001 quando deveria ter previsto R\$ 40.000, não importa muito. Em contrapartida, para a tarefa de reconhecer a linguagem de um site, não há questão de grau. Um site está em um idioma ou em outro. Não há continuidade entre as línguas, e não há linguagem que seja entre inglês e francês.

3.2.2 GENERALIZAÇÃO, SOBRE-AJUSTE E SUB-AJUSTE

MULLER [2016] O grande objetivo do aprendizado de máquina é construir um modelo de dados de treinamento e, em seguida, realizar previsões precisas sobre dados novos e invisíveis que têm as mesmas características do conjunto de treinamento usado. Se um modelo é capaz de fazer previsões precisas sobre dados desconhecidos, pode-se dizer que é capaz de generalizar do conjunto de treinamento para o conjunto de testes. Deseja-se construir um modelo que seja capaz de generalizar com a maior precisão possível.

Se os conjuntos de treinamento e teste tiverem o suficiente em comum, espera-se que o modelo também seja preciso no conjunto de novos dados de teste. No entanto, existem alguns casos em que este pode dar errado. Por exemplo, para a construção de modelos muito complexos, a generalização nem sempre pode ser tão preciso quanto quisermos no conjunto de treinamento. A única medida para saber se um algoritmo terá um bom desempenho em novos dados é a avaliação no conjunto de testes. No entanto, intuitivamente espera-se que modelos simples se generalizem melhor para novos dados.

Segundo MULLER [2016], construir um modelo muito complexo para certa quantidade de informações é chamado de sobre ajuste. Este ocorre quando ajusta-se um modelo muito próximo às particularidades do conjunto de treinamento e obtém-se um modelo que funciona bem no conjunto de treinamento, mas não é capaz de generalizar para novos dados, esse processo é conhecido como Overfitting. Por outro lado, se o modelo for muito simples, então pode ser que não seja capaz de capturar todos os aspectos e capacidade dos dados, e o modelo terá um desempenho ruim, mesmo no conjunto de treinamento. Escolher um modelo muito

simples é chamado de underfitting. Portanto, a ocasião ideal, vista na figura 3.2, seria achar um modelo entre o overfitting e o underfitting para melhor desempenho da generalização

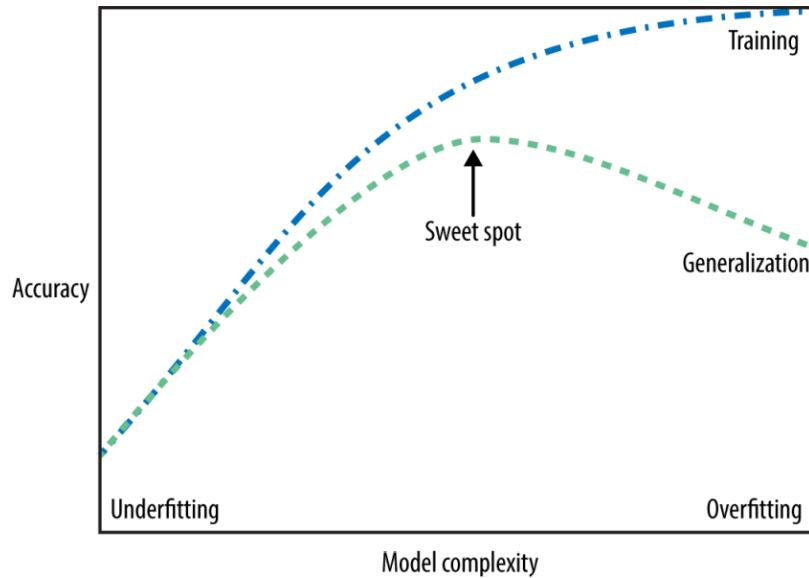


Figura 3.2 – Verificação da complexidade do modelo treinamento e teste (sub-ajuste e sobreajuste)

Fonte: MULLER, 2016

3.3 MODELOS DE APRENDIZADO SUPERVISIONADOS

3.3.1 REGRESSÃO LINEAR

AMPADU [2021] descreve a regressão linear como sendo, um algoritmo capaz de formular uma equação linear que capture a relação entre as características independentes e dependentes, ou seja, é encontrar a linha de melhor ajuste, que possa prever corretamente a saída das variáveis dependentes. A regressão linear pode ser usada para regressão linear simples e múltipla.

MULLER [2016] afirma que a regressão linear, é a linha mais simples e clássica regressão. A regressão linear encontra os parâmetros w e b que minimizam o erro quadrático médio entre as previsões e os verdadeiros alvos de regressão, y , no conjunto de treinamento. O erro quadrático médio é a soma das diferenças quadradas entre as previsões e os valores verdadeiros, dividido pelo número de amostras. A regressão linear não tem parâmetros, o que é um benefício, mas também não tem como controlar a complexidade do modelo.

Este modelo de machine learning cria e aprende os padrões dos dados matematicamente através de uma função para gerar previsões, portanto pode-se afirmar que em algoritmos supervisionados a função para gerar os valores de resposta está descrita na equação 3.1 é:

$$Y = f(x) = f(x_1, x_2, \dots, x_n) \quad (3.1)$$

Onde,

Y – Saída estimada com base na função.

$f(x)$ – Função que o algoritmo irá criar

x – Variável independente, ou seja, os atributos

Para DAMASCENO [2020] Existem 2 tipos de regressão linear: simples e múltipla. Regressão linear simples: refere-se quando temos somente uma variável independente (X) para fazermos a predição. Já na Regressão linear múltipla: refere-se a várias variáveis independentes (X) usadas para fazer a predição. E com isso a forma de representação de cada regressão varia conforme o tipo como pode ser vista nas figuras 3.3 e 3.4. Por exemplo a representação gráfica em uma regressão linear simples é uma reta em um plano de 2 dimensões, já em uma regressão linear múltipla, sua representação é feita em um plano que pode ser nD .

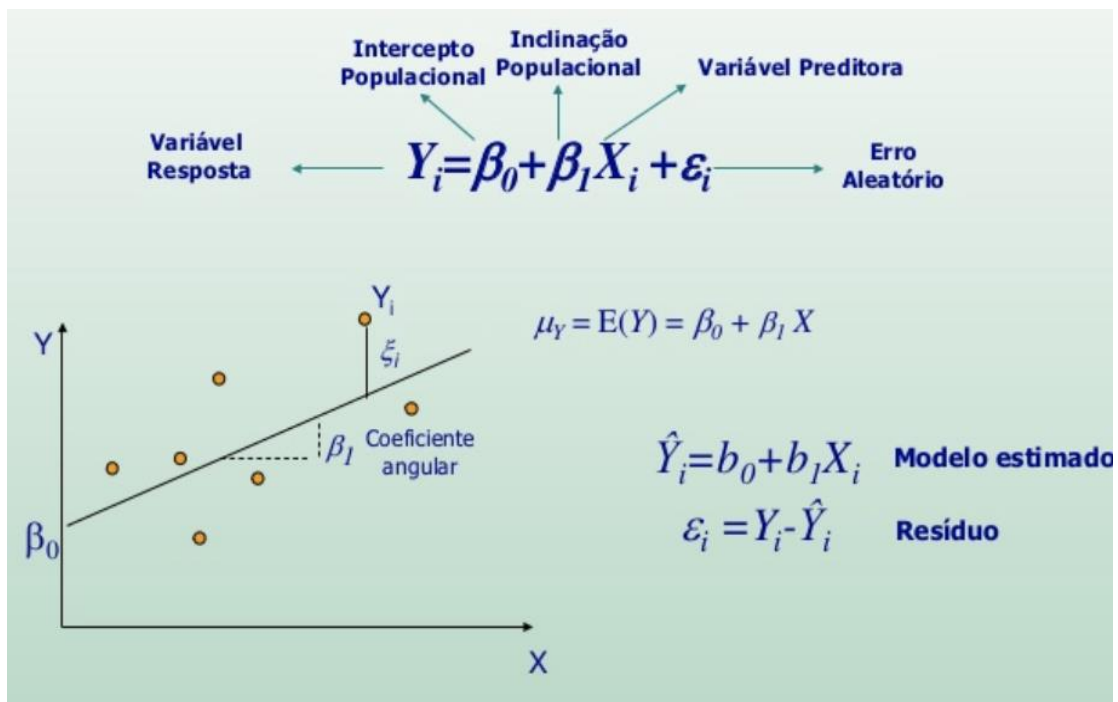


Figura 3.3 – Representação gráfica da regressão linear simples

Fonte: AMPADU 2021 – piscina da IA

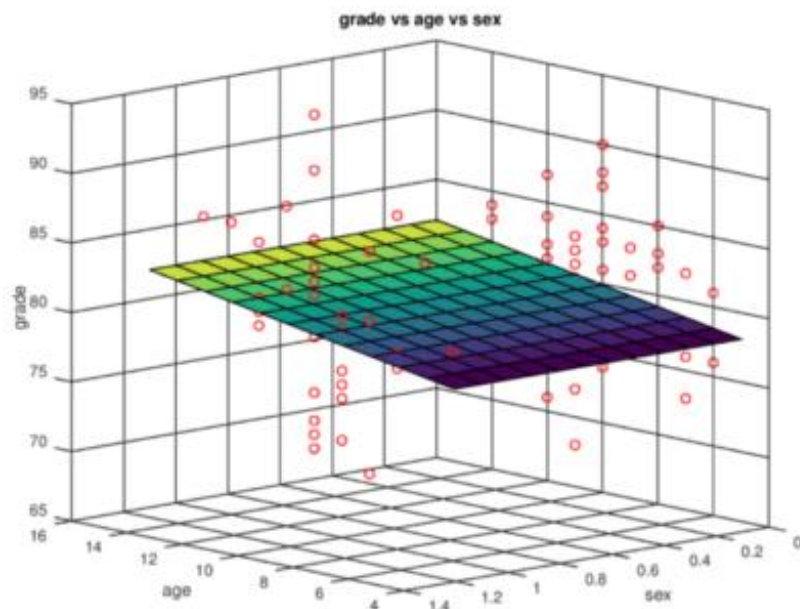


Figura 3.4 – Representação gráfica da regressão linear múltipla

Fonte: Damasceno, 2020

3.3.2 ÁRVORES DE DECISÃO

A indução de árvores de decisão é uma das formas mais simples, e ainda assim mais bem-sucedidas, de aprendizagem de máquina. Para CAMPOS [2017] Árvores de decisão são modelos amplamente usados para tarefas de classificação e regressão. Inicialmente, eles aprendem uma hierarquia de perguntas se/senão, levando a uma decisão.

RUSSEL [2010] Mostra que uma árvore de decisão representa uma função que toma como entrada um vetor de valores de atributos e retorna uma “decisão” — um valor de saída único. Os valores de entrada e saída podem ser discretos ou contínuos. Por ora vamos nos concentrar em problemas em que a entrada tem valores discretos e a saída tem exatamente dois valores possíveis; isto é: classificação booleana, em que cada exemplo é classificado como verdadeiro (positivo) ou falso (negativo). Uma árvore de decisão alcança sua decisão executando uma sequência de testes. Cada nó interno na árvore corresponde a um teste do valor de um dos atributos de entrada, A_i , e as ramificações dos nós são classificadas com os valores possíveis do atributo, $A_i = V_{ik}$. Cada nó de folha na árvore especifica o valor a ser retornado pela função. A representação de árvores de decisão parece ser muito natural para os seres

humanos; na realidade, muitos manuais do tipo “como fazer” (por exemplo, para consertos de automóveis) são inteiramente escritos como uma única árvore de decisão que se estende por centenas de páginas.

Um exemplo clássico de árvore de decisão é a escolha de um jogo de tabuleiro, conforme pode-se verificar na figura 3.5 abaixo.



Figura 3.5– Árvore de decisão para escolha de jogo de tabuleiro

Fonte: CAMPOS 2017

Árvores, de um modo geral em computação, são estruturas de dados formadas por um conjunto de elementos que armazenam informações chamadas **nós**. Na figura acima, os nós são representados pelos quadradinhos com as perguntas e as informações podem ser consideradas as perguntas e suas possíveis respostas. Além disso, toda árvore possui um nó chamado **raiz**, que possui o maior nível hierárquico (o ponto de partida) e ligações para outros elementos, denominados filhos. Esses filhos podem possuir seus próprios filhos que por sua vez também possuem os seus. O nó que não possui filho é conhecido como **nó folha** ou terminal (representado pelo símbolo arredondado na figura). Tendo essas definições esclarecidas, uma árvore de decisão nada mais é que uma árvore que armazena regras em nós, e as nossas folhas representam a decisão a ser tomada (no caso do exemplo, qual jogo escolher).

Em MULLER [2016] as árvores de decisão têm duas vantagens sobre muitos dos algoritmos, são elas: o modelo resultante pode ser facilmente visualizado e compreendido por não especialistas (em menos para árvores menores), e os algoritmos são completamente invariáveis para o dimensionamento dos dados. Como cada recurso é processado separadamente, as possíveis divisões dos dados não dependem de dimensionamento sem pré-processamento, como normalização ou padronização de recursos para algoritmos de árvore de decisão. Em particular, as árvores de decisão funcionam bem quando se tem recursos que estão em escalas completamente diferentes, ou uma mistura de binários e com recursos contínuos. A principal desvantagem das árvores de decisão é que, mesmo com o uso de pré-poda, elas tendem a super ajustar e fornecer desempenho de generalização pobre. Portanto, na maioria das aplicações, os métodos de conjunto de árvores de decisão que serão discutidos a seguir são geralmente usados no lugar de uma árvore de decisão geral.

3.3.3 RANDOM FOREST

Como visto anteriormente, a principal desvantagem das árvores de decisão é que elas tendem a super ajustar os dados de treinamento. Para BREIMAN [2001] as Random forest (florestas aleatórias) são uma forma de resolver esse problema. Uma floresta aleatória é essencialmente uma coleção de árvores de decisão, conforme pode ser visto na figura 3.6, onde cada árvore é ligeiramente diferente das outras. A ideia por trás das florestas aleatórias é que cada árvore pode fazer um bom trabalho de previsão, mas provavelmente superestimam parte dos dados. Se construirmos muitas árvores, que funcionem bem e se ajustem demais de maneiras diferentes, podemos reduzir a quantidade de overfitting calculando a média de seus resultados. Esta redução no overfitting, mantendo o poder preditivo das árvores, pode ser demonstrada por meio de matemática rigorosa. Para implementar essa estratégia, é necessário construir muitas árvores de decisão. Cada árvore deve fazer um trabalho aceitável de prever o alvo, e deve ser diferente da outra árvore. Florestas aleatórias recebem seu nome por injetar aleatoriedade na construção da árvore para garantir que cada árvore seja diferente. Existem duas maneiras em que as árvores de forma aleatória de floresta são randomizadas: selecionando os pontos de dados usados para construir uma árvore e selecionando os recursos em cada teste de divisão.

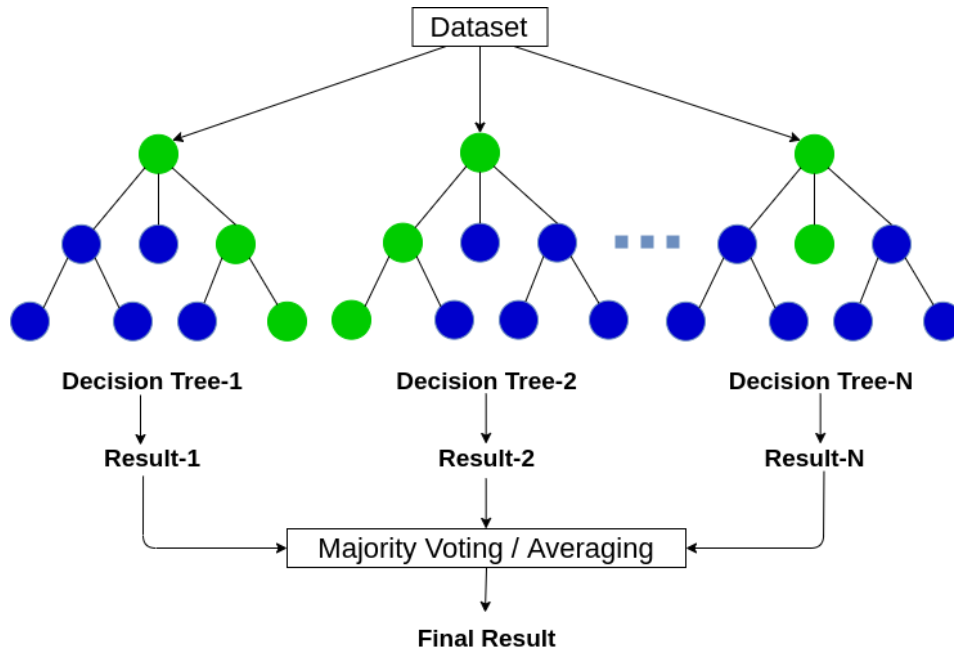


Figura 3.6 – Modelo de funcionamento do algoritmo Random Forest

Fonte: Piscina de IA, 2021

Como pode ser visto na figura acima, existem várias árvores de decisão, e cada uma trabalha independentemente para formar sua própria produção e dar uma previsão. A floresta aleatória então pega a previsão de cada árvore e seleciona a maioria da classe que cada árvore previu como a verdadeira classe prevista do conjunto de dados.

Para AMPADU [2021] o algoritmo do random Forest possui o funcionamento a seguir:

- Selecione aleatoriamente alguns pontos de dados do conjunto
- Construa árvores de decisão usando os pontos de dados selecionados
- Selecione o número de árvores de decisão a serem construídas
- Repita o primeiro e o segundo passos
- Para qualquer novo ponto de dados, calcule a previsão de cada árvore de decisão e, usando a maioria dos votos pelas árvores, atribua o novo data point(s) a essa classe.

4. MODELAGEM DA DESCARGA SUPERFICIAL DE UM ISOLADOR POR MACHINE LEARNING

4.1 – RELAÇÃO DA TENSÃO CRÍTICA COM PARÂMETROS DA DESCARGA SUPERFICIAL

Para a realização de simulação do trabalho proposto, com objetivo de antecipar a manutenção corretiva nos isoladores de distribuição, utilizou-se a técnica de inteligência artificial conhecida como aprendizado de máquina ("Machine Learning") com a programação em Python na plataforma Jupyter notebook para simulação do código e para verificar o algoritmo que melhor se adapta à resolução do problema, de forma que se mostre um resultado real e preciso diante das condições do isolador estudado.

Primeiramente foi realizado o tratamento dos dados fornecidos por PINOTTI [2009]. A tabela de resultados mostra os diferentes tipos de condições ambientais nas quais foram realizados os testes em um intervalo de 30 dias. Foram medidas as seguintes variáveis de condições ambientais: Umidade, temperatura, pressão, vento angular, velocidade do vento, chuva e corrente de fuga. Esses parâmetros são necessários para a realização da correlação com a corrente de fuga, que por fim tem alto grau de impacto na tensão crítica de flashover (TCFO) do isolador, ou seja, para cada combinação dos parâmetros selecionados existe uma tensão crítica que mostra a suportabilidade limite do isolador, antes de apresentar defeito ou até mesmo a causa do processo de flashover. Foram obtidos cerca de 40.320 dados de medição. Para exemplificar, a seguir, na tabela 4.1 serão apresentadas as cinco primeiras medições.

Umidade	Temperatura	Pressão	Vento_angular	Velocidade do vento	chuva	Corrente de fuga
70	29,6	1016	278	1,4	0	0,024992
69	29,7	1016	297	1,4	0	0,022767
70	29,8	1016	273	1,6	0	0,022607
68	29,8	1016	243	2,6	0	0,022426
68	29,7	1016	257	2,8	0	0,031797

Tabela 4.1 – Medição das condições climáticas e corrente de fuga

A partir dos dados obtidos em medição e do valor determinado para corrente de fuga, verificou-se por meio de cálculos, que serão mostrados a seguir, a tensão de disrupção ou tensão crítica para cada valor medido das condições climáticas e do valor da corrente de fuga. O aumento das falhas em isoladores, decorrente do alto valor da corrente de fuga, segundo GONOS et al [2012] tem se tornado crescente nos últimos anos acarretando sérios problemas na operação da rede de distribuição de energia através do fenômeno chamado Flashover, devido a esse crescimento, houve também o aumento nos estudos para minimizar essas ocorrências. Uma dessas pesquisas seria prever a ocorrência dessa tensão crítica. A aplicação de métodos de aprendizado de máquina torna-se fundamental para essa análise, onde por meio computacional pode-se antecipar o processo de detecção dos defeitos nos isoladores. Para a realização do cálculo dessa tensão crítica segundo GONOS et al [2012] tem-se a seguinte equação:

$$U = xAI^{-n} + (L - n)R_pI \quad (4.1)$$

Onde:

x – Comprimento do arco

A – Constante do arco

n – Constante do arco

I – Corrente de fuga

L – Caminho de vazamento do isolador

R_p – Resistência da zona úmida

Sabe-se que seria muito complexo realizar o cálculo para descobrir o valor da resistência da zona úmida. Sendo assim e por meio de alguns arranjos matemáticos pode substituir a equação (4.1) acima por (4.2):

$$U_c = \frac{A}{n+1} (L + \pi D_r F_i K_n) (\pi D_r \sigma_p A)^{\frac{-n}{n+1}} \quad (4.2)$$

Onde:

U_c – Tensão Crítica

D_r – Diâmetro do isolador

F_i – Fator de forma

K_n – Parâmetro de ajuste da resistência da camada de poluição

σ_p – Condutividade

Outra questão a ser esclarecida para realização do cálculo da tensão crítica, seria em decorrência dos parâmetros de A e n, visto que eles são constantes de arco e segundo o autor GENÇOGLU Et al [2008] essas constantes poderiam ser determinadas de acordo com o meio onde ocorre o processo da descarga. Para o presente trabalho, considera-se que os testes foram realizados em campo aberto, ou seja, no ar livre, GENÇOGLU et al [2008] para essas condições impostas o valor de A é igual a 63 e o de I igual a 0,7. Para uma análise mais profunda, segundo GONOS et al [2012] a equação da corrente de fuga do flashover é:

$$I_c = (\pi D_r \sigma_p A)^{\frac{1}{n+1}} \quad (4.3)$$

E para a tensão crítica é de:

$$U_c = \frac{A}{n+1} (L + \pi D_r F_i K_n) (\pi D_r \sigma_r A)^{-n/(n+1)} \quad (4.4)$$

Dessa forma, ao serem realizados alguns artifícios matemáticos nas equações acima, para que a tensão crítica possa depender fielmente da corrente de fuga, encontrou-se a seguinte equação:

$$U_c = \frac{A}{n+1} (L + \pi D_r F_i K_n) \frac{1}{I_c^n} \quad (4.5)$$

A adição do parâmetro K_n feito por WILKINS [1969].: tem como objetivo validar a equação 01 no instante que ocorre o flashover no isolador. Esse coeficiente tem a função de modificar a resistência da camada de poluição considerando a concentração atual no ponto de apoio do arco. Dado pela seguinte fórmula:

$$K = 1 + \frac{L}{2\pi F_i (L - X_c)} \ln \frac{L}{2\pi F_i \sqrt{\frac{I_c}{1.45\pi}}} \quad (4.6)$$

Sabe-se que L, F_i e I_c são parâmetros conhecidos para realização de cálculos, resta apenas descobrir X_c para achar o valor de K. Segundo WILKINS [1969]: o comprimento do arco no momento do flashover pode ser calculado pela seguinte equação:

$$X_c = \frac{1}{n + 1} L \quad (4.7)$$

A partir disso, tem-se todos os valores conhecidos para a determinar a tensão em que o isolador irá apresentar durante o processo de flashover, logo pode-se realizar o cálculo da tensão para cada valor medido de corrente de fuga, resultando nos seguintes valores demonstrados na tabela 4.2, vale ressaltar que são apenas uma amostra retirada do conjunto de dados que serão mantidos para aplicação do modelo em Machine Learnig.

Umidade	Temperatura	Pressão	Vento angular	Velo.do vento	chuva	Corrente de fuga	Tensão Crítica (KV)
70	29,6	1016	278	1,4	0	0,024992	24,11
69	29,7	1016	297	1,4	0	0,022767	25,80
70	29,8	1016	273	1,6	0	0,022607	25,93
68	29,8	1016	243	2,6	0	0,022426	26,03
68	29,7	1016	257	2,8	0	0,031797	20,25

Tabela 4.2 – Valores de tensão crítica para cada condição ambiental

Desta forma, com os valores encontrados para tensão na qual o isolador tende a entrar em processo de flashover, pretende-se encontrar um modelo em Machine Learning que possa identificar ou prever a falha do isolador antes dele chegar até a sua tensão crítica baseado nas variáveis ambientais obtidas. Com o auxílio da linguagem de programação em python pode-se aplicar o modelo para resolução do problema e mostrar os resultados claros e precisos.

4.2 - METODOLOGIA PARA CRIAÇÃO DE MODELOS USANDO APRENDIZADO DE MÁQUINA

4.2.1 TRATAMENTO DE DADOS

Primeiramente na plataforma no Jupyter notebook adicionaram-se as bibliotecas para a leitura e tratamento dos dados, conforme figura 4.1.

```
In [1]: import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import MinMaxScaler
from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_squared_error
from sklearn.tree import DecisionTreeRegressor
from sklearn.ensemble import RandomForestRegressor
from sklearn.model_selection import GridSearchCV
from sklearn.metrics import accuracy_score
from sklearn.preprocessing import StandardScaler
from sklearn.preprocessing import Normalizer
from IPython.display import display, HTML
%matplotlib inline
import warnings
warnings.filterwarnings('ignore')
```

Figura 4.1 – Adição de bibliotecas na plataforma Jupyter notebook

Os dados coletados das condições climáticas e a corrente de fuga medida nestas condições ambientais juntamente com os valores calculados de tensão crítica foram disponibilizados em arquivo CSV. e foram importados para um dataframe da biblioteca pandas, sendo executado pelo comando descrito abaixo.

```
In [9]: df = pd.read_csv('new_dados_isolador40.csv', sep=';', decimal=',')
```

Figura 4.2 – Importação dos dados para leitura

A leitura dos dados conforme é apresentada na figura 4.3, verificando que as medições e cálculos coincidem.

```
In [10]: df.head(5)
```

```
Out[10]:
```

	umidade	temperatura	pressao	vento_ang	vent_veloc	chuva	corrente_fuga	A	B	C	K	Uc
0	70	29.6	1016	278	1.4	0	0.024992	0.086084	0.074070	5.799708	1.499261	24119.27040
1	69	29.7	1016	297	1.4	0	0.022767	0.086084	0.070696	5.846330	1.503274	25805.31545
2	70	29.8	1016	273	1.6	0	0.022607	0.086084	0.070447	5.849856	1.503577	25937.52914
3	68	29.8	1016	243	2.6	0	0.022426	0.086084	0.070164	5.853876	1.503923	26089.05394
4	68	29.7	1016	257	2.8	0	0.031797	0.086084	0.083548	5.679300	1.488895	20256.54987

Figura 4.3 – leitura das cinco primeiras amostras de dados

Com a utilização do comando *shape* pode-se verificar a quantidade de linhas e colunas lidas pelo arquivo (dimensões do array), confirmando a informação exata dos valores que foram medidos e calculados

```
In [11]: df.shape
Out[11]: (40316, 12)

In [12]: display(df.columns.values)
array(['umidade', 'temperatura', 'pressao', 'vento_ang', 'vent_veloc',
       'chuva', 'corrente_fuga', 'A', 'B', 'C', 'K', 'Uc'], dtype=object)
```

Figura 4.4 – quantidade de valores lidos do arquivo

Para verificar se há valores nulos nos dados em questão, utiliza-se o seguinte comando descrito na figura 4.5. Vale ressaltar, que os valores que apresentaram valores nulos (ângulo do vento, velocidade do vento, chuva), são variáveis que dependem das condições climáticas da região no período dos dias que foram realizados os testes.

```
In [14]: (df == 0).sum()
Out[14]: umidade          0
         temperatura     0
         pressao         0
         vento_ang       92
         vent_veloc     14511
         chuva          38064
         corrente_fuga   0
         A               0
         B               0
         C               0
         K               0
         Uc              0
         dtype: int64
```

Figura 4.5 – Verificação de valores nulos

4.2.2 ANÁLISE E EXPLORAÇÃO DE DADOS

Tendo os dados já apresentados de forma satisfatória, busca-se agora uma análise mais refinada na qual procuram-se características que possam auxiliar na criação do modelo. Primeiramente, os dados foram inseridos em um gráfico de distribuição com histograma gerado pelo comando Seaborn, que tem como finalidade estimar uma curva KDE (curva de densidade estimada) da verdadeira função da densidade dos dados.

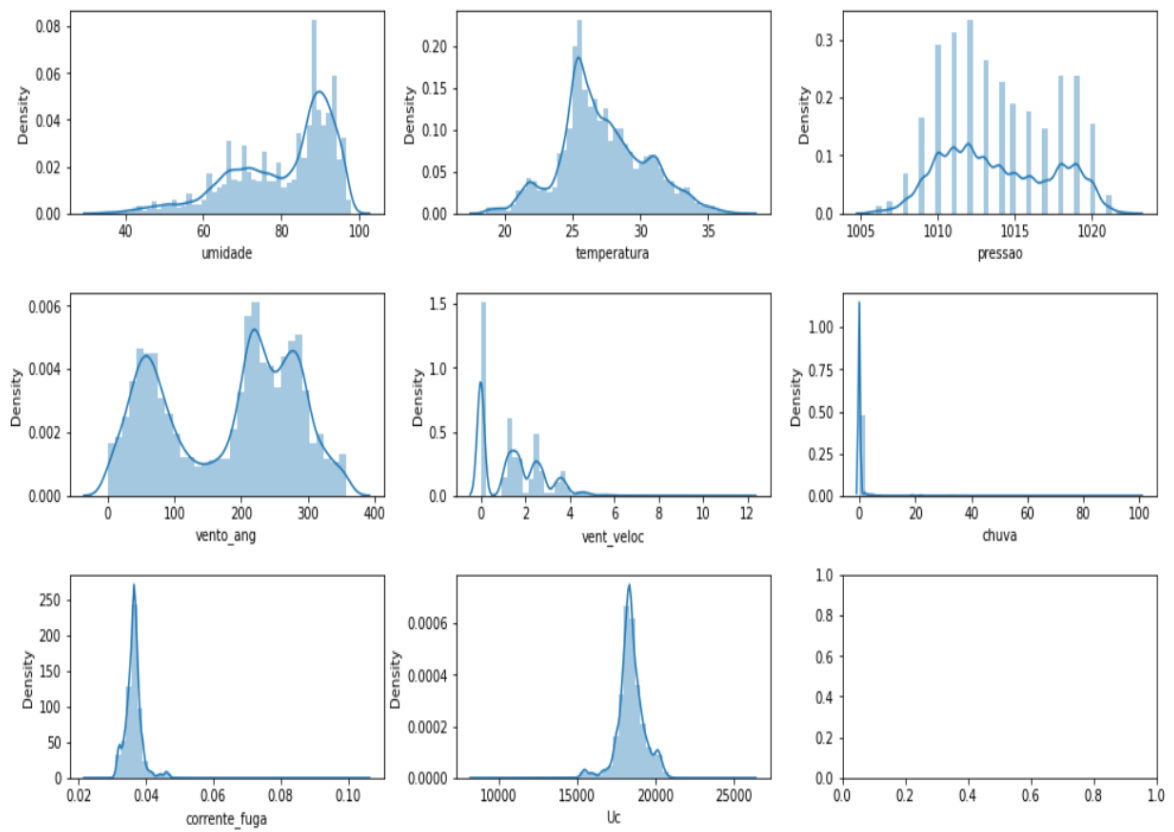


Figura 4.6 – Função de densidade dos dados

Visto que os dados apresentam escalas distintas, será necessário uma padronização e normalização nas suas variáveis, essas ações são necessárias para melhor eficiência e visualização do modelo.

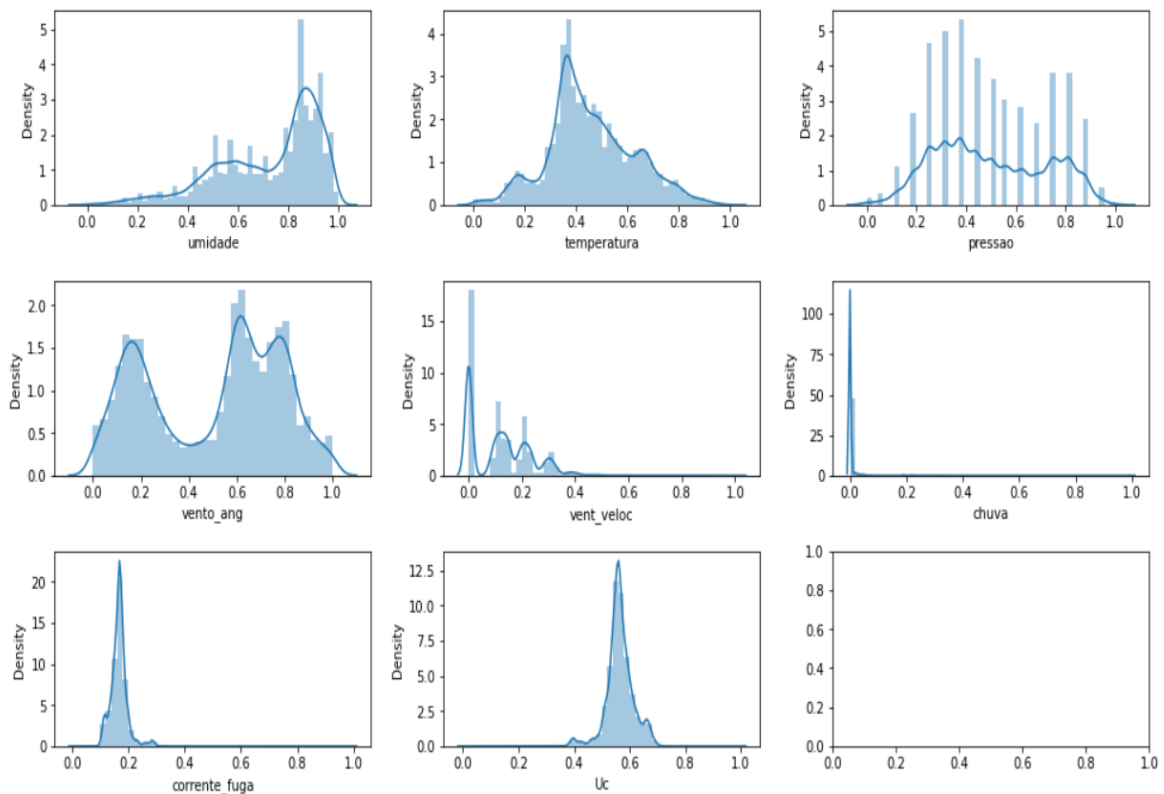


Figura 4.7 – Normalização e padronização das variáveis

Para explorarmos mais os dados em análise, verificaram-se cálculos estatísticos, visando a busca por características e tendências do conjunto. Os parâmetros calculados foram: média, desvio padrão, máximo e mínimo e os valores dos quartis.

```
In [21]: df = scaled_data
df.describe()
```

```
Out[21]:
```

	umidade	temperatura	pressao	vento_ang	vent_veloc	chuva	corrente_fuga	A	B	C	K
count	40316.000000	40316.000000	40316.000000	40316.000000	40316.000000	40316.000000	40316.000000	40316.0	40316.000000	40316.000000	40316.000000
mean	0.726673	0.456358	0.495887	0.503842	0.120464	0.003483	0.167389	0.0	0.232764	0.689886	0.689886
std	0.201368	0.166669	0.223528	0.278183	0.111802	0.027008	0.030226	0.0	0.036649	0.042686	0.042686
min	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.0	0.000000	0.000000	0.000000
25%	0.578125	0.351064	0.312500	0.210084	0.000000	0.000000	0.152262	0.0	0.214385	0.671046	0.671046
50%	0.796875	0.430851	0.437500	0.588235	0.117647	0.000000	0.167479	0.0	0.233494	0.688307	0.688307
75%	0.890625	0.563830	0.687500	0.742297	0.201681	0.000000	0.179329	0.0	0.248145	0.711173	0.711173
max	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	0.0	1.000000	1.000000	1.000000

Figura 4.8 – Cálculos de dados estatísticos

A função `describe()` disponibilizada pela biblioteca pandas gera os dados estatísticos que resumem a tendência central, a dispersão e a forma de distribuição de um conjunto de dados.

Com os cálculos estatísticos realizados, será apresentado o modelo de gráfico em boxplots conforme figura 4.9. Esses diagramas são construídos a partir de caixas que possuem a referência nos valores de dados estatísticos e tem como objetivo estudar as medidas estatísticas do conjunto de dados, como propriedade de locação, variabilidade, média e outliers.

```
In [17]: boxplots_dataframe(df)
```

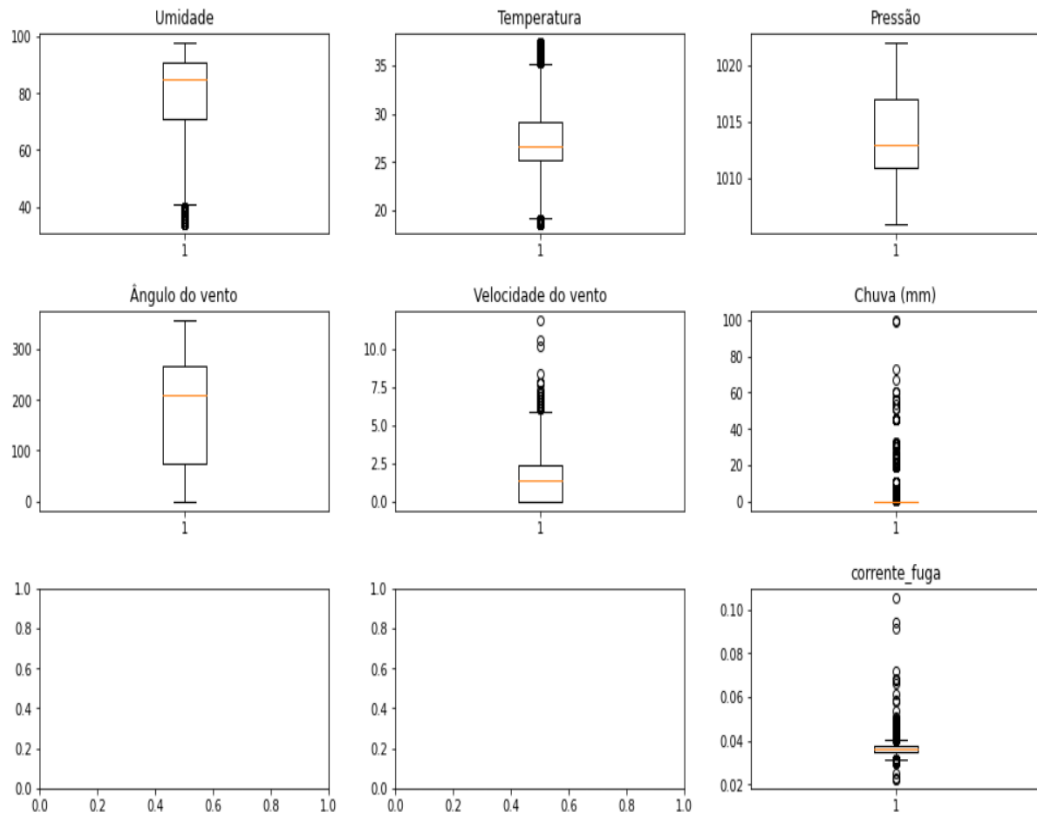


Figura 4.9 – Boxplots das variáveis do dataframe

A análise desses diagramas consiste no seguinte, o boxplot sempre começa pelo seu valor mínimo e termina no seu valor máximo (valores calculados de dados estatísticos), a caixa representa os valores centrais dos dados, tais como valores que aparecem com mais frequência. A linha onde começa cada caixa é representada pelo primeiro quartil, a linha contida no centro da caixa é o valor médio dos dados, e por fim, a caixa é finalizada no valor do terceiro quartil. Entretanto, esses diagramas são importantes também para a validação do modelo, pois são responsáveis pela verificação de overfitting e underfitting. Overfitting é um modelo que consegue detectar até os pontos mais discrepantes do boxplot, porém por possuir alta variância e conseqüentemente alta complexidade, gera resultados ruins. Ao contrário do overfitting, tem-se o underfitting que por ser o modelo mais simples, possui uma baixa variância, conseguindo detectar poucos pontos de testes e que também gera um resultado não satisfatório.

Na figura 4.9 com os boxplots das variáveis, pode-se notar que as variáveis: Umidade, temperatura, velocidade do vento, chuva e corrente de fuga apresentaram quantidades consideráveis de outliers, mas como são variáveis dependentes das condições climáticas na hora do teste, irão permanecer para a continuação de criação do modelo de Machine Learning.

Após toda a análise exploratória dos dados, deve-se correlacionar as variáveis predictoras com a variável alvo mostrado nas figuras 4.10 e 4.11. Com isso, pode-se analisar a dependência da variável da tensão crítica de flashover com as demais variáveis, sendo assim criar um grau de dependência que essas variáveis podem ter durante a criação do modelo em Machine Learning. Para usar a correlação utilizou-se o comando `corr()` e resultado comparativo dessa correlação aparece na figura a seguir:

	umidade	temperatura	pressao	vento_ang	vent_veloc	chuva	corrente_fuga	A	B	C	K	Uc
umidade	1.0	-0.44157	-0.20368	0.32259	-0.50396	0.11214	0.69658	nan	0.72056	-0.74008	-0.74008	-0.76278
temperatura	-0.44157	1.0	-0.39874	-0.43925	0.45762	-0.11037	-0.34919	nan	-0.35956	0.36748	0.36748	0.3759
pressao	-0.20368	-0.39874	1.0	-0.011539	-0.045943	0.044076	-0.064254	nan	-0.068527	0.072534	0.072534	0.077962
vento_ang	0.32259	-0.43925	-0.011539	1.0	-0.34307	0.0424	0.22667	nan	0.23102	-0.23374	-0.23374	-0.23568
vent_veloc	-0.50396	0.45762	-0.045943	-0.34307	1.0	-0.0043842	-0.3078	nan	-0.31741	0.32493	0.32493	0.33315
chuva	0.11214	-0.11037	0.044076	0.0424	-0.0043842	1.0	0.29679	nan	0.29242	-0.28624	-0.28624	-0.27544
corrente_fuga	0.69658	-0.34919	-0.064254	0.22667	-0.3078	0.29679	1.0	nan	0.99804	-0.99345	-0.99345	-0.98433
A	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan
B	0.72056	-0.35956	-0.068527	0.23102	-0.31741	0.29242	0.99804	nan	1.0	-0.99863	-0.99863	-0.99328
C	-0.74008	0.36748	0.072534	-0.23374	0.32493	-0.28624	-0.99345	nan	-0.99863	1.0	1.0	0.99794
K	-0.74008	0.36748	0.072534	-0.23374	0.32493	-0.28624	-0.99345	nan	-0.99863	1.0	1.0	0.99794
Uc	-0.76278	0.3759	0.077962	-0.23568	0.33315	-0.27544	-0.98433	nan	-0.99328	0.99794	0.99794	1.0

Figura 4.10 – correlação das variáveis

```
In [25]: df.corr()['Uc']
Out[25]: umidade      -0.762777
          temperatura  0.375896
          pressao     0.077962
          vento_ang   -0.235681
          vent_veloc  0.333150
          chuva       -0.275440
          corrente_fuga -0.984327
          A           NaN
          B           -0.993281
          C           0.997940
          K           0.997940
          Uc          1.000000
          Name: Uc, dtype: float64
```

Figura 4.11 – correlação entra as variáveis predictoras e a variável alvo (Uc)

Ao analisar os seguintes gráficos, chega-se à conclusão que as variáveis: umidade, corrente de fuga e K são as que possuem uma melhor correlação com a tensão crítica de flashover, esse resultado é satisfatório pois essas três variáveis são responsáveis diretamente pela ocorrência do processo de flashover do isolador, conforme explicado no capítulo 2.

4.3 – DESCRIÇÃO E COMPARAÇÃO DE MODELOS DE APRENDIZADO DE MÁQUINAS

4.3.1 – MODELO REGRESSÃO LINEAR

Conforme visto no Capítulo 3, a regressão linear tem como objetivo principal prever certo resultado, através da correlação entre os dados, partindo do princípio de que há um padrão entre eles e esse padrão pode ser aproximado por uma função linear do primeiro grau com uma variável. Sendo assim, testa-se o algoritmo, criando uma função para treinar o modelo e calcular o valor do score conforme feito na figura 4.12, e demonstra se o algoritmo é confiável ou não para modelar o comportamento do isolador frente a ocorrência de flashover causado por poluição.

```
In [88]: def calcula_regressao(columns):
X_train, X_test, y_train, y_test = train_test_split(df[columns], df['Uc'], test_size=0.2, random_state=42)
lr = LinearRegression(normalize = False, fit_intercept=True)

model_lr = lr.fit(X_train, y_train)
result_lr = model_lr.predict(X_test)

mse_lr = mean_squared_error(y_test, result_lr)

rmse_lr = (np.sqrt(mse_lr))

score_lr = model_lr.score(X_test, y_test)

print('Resultados do erro médio e score (acurácia do modelo) para o dataset de teste:\n MSE: {}'.format(mse_lr)
      + '\n RMSE: {}'.format(rmse_lr)
      + '\nScore: {:.24f}'.format(score_lr * 100) + '%')
return score_lr;
```

Figura 4.12 – Aplicação do modelo Regressão linear e cálculo do seu score

Esse algoritmo demonstra que os dados do modelo foram divididos em duas variáveis, uma de treino e outra de teste, na qual 20% dos dados foram reservados para teste e os 80% restantes foram destinados para a realização de treinamento. A função que faz o treinamento do modelo é a *fit* ().

Para verificação da acurácia do modelo, conforme figura 4.13, será removida a coluna da variável alvo (tensão crítica Uc) permanecendo todos os outros atributos, independentes do grau de correlação deles com a tensão crítica.

```
In [89]: cols = df.columns.drop(['Uc'])
        cols

Out[89]: Index(['umidade', 'temperatura', 'pressao', 'vento_ang', 'vent_veloc', 'chuva',
               'corrente_fuga'],
              dtype='object')
```

```
In [90]: score_cols = calcula_regressao(cols)

Resultados do erro médio e score (acurácia do modelo) para o dataset de teste:
MSE: 2.630505892469389e-05
RMSE: 0.005128845769244177
Score: 98.8772%
```

Figura 4.13 – Acurácia do modelo com todas as variáveis predictoras

Vale ressaltar que o alto valor de score se dá pela variável da corrente de fuga, visto que apresenta um grau de correlação muito alto, pois conforme descrito no capítulo 2, a corrente de fuga é o principal fator para que o isolador possa atingir a tensão crítica e entrar no processo de flashover, sendo assim, por meio dos cálculos realizados na seção 4.1 tem-se para cada valor de corrente de fuga, um valor exato da tensão na qual o isolador irá se romper.

A seguir será realizado teste retirando as colunas que possuem menor grau de correlação com a variável principal e para finalizar esse modelo, será realizado o teste com a retirada da coluna corrente de fuga, para comparar a importância dessa variável para o trabalho em questão.

Inicialmente retiram-se as colunas que possuem correlação menor que 0,5.

```
In [91]: cols_maior_05 = df.columns.drop(['temperatura', 'pressao', 'vento_ang', 'vent_veloc', 'chuva', 'Uc'])
        cols_maior_05

Out[91]: Index(['umidade', 'corrente_fuga'], dtype='object')
```

```
In [92]: score_cols_maior_05 = calcula_regressao(cols_maior_05)

Resultados do erro médio e score (acurácia do modelo) para o dataset de teste:
MSE: 2.720661552502498e-05
RMSE: 0.005215996120112148
Score: 98.8387%
```

Figura 4.14 – score retirada das variáveis que possuem correlação menor que 0,5

Em seguida, retiram-se as que possuem correlação menor que 0,3.

```
In [93]: cols_maior_03 = df.columns.drop(['pressao', 'vento_ang', 'chuva', 'Uc'])
cols_maior_03
```

```
Out[93]: Index(['umidade', 'temperatura', 'vent_veloc', 'corrente_fuga'], dtype='object')
```

```
In [94]: score_cols_maior_03 = calcula_regressao(cols_maior_03)
```

Resultados do erro médio e score (acurácia do modelo) para o dataset de teste:
MSE: 2.677769614235856e-05
RMSE: 0.0051747170108478934
Score: 98.8570%

Figura 4.15 – score retirada das variáveis que possuem correlação menor que 0,3

Depois, retira-se as que possuem correlação menor que 0,1.

```
In [95]: cols_maior_01 = df.columns.drop(['pressao', 'Uc'])
cols_maior_01
```

```
Out[95]: Index(['umidade', 'temperatura', 'vento_ang', 'vent_veloc', 'chuva',
'corrente_fuga'],
dtype='object')
```

```
In [96]: score_cols_maior_01 = calcula_regressao(cols_maior_01)
```

Resultados do erro médio e score (acurácia do modelo) para o dataset de teste:
MSE: 2.6348016145353567e-05
RMSE: 0.005133031866777525
Score: 98.8753%

Figura 4.16 – score retirada das variáveis que possuem correlação menor que 0,1

Por fim, retira-se a coluna de corrente de fuga.

```
In [100]: cols_corrente_fuga = df.columns.drop(['corrente_fuga', 'Uc'])
cols_corrente_fuga
```

```
Out[100]: Index(['umidade', 'temperatura', 'pressao', 'vento_ang', 'vent_veloc',
'chuva'],
dtype='object')
```

```
In [101]: score_cols_corrente_fuga = calcula_regressao(cols_corrente_fuga)
```

Resultados do erro médio e score (acurácia do modelo) para o dataset de teste:
MSE: 0.0008435596232469695
RMSE: 0.029044097907268
Score: 63.9930%

Figura 4.17 – score do modelo com a retirada da variável corrente de fuga.

Conforme previsto, com a retirada da variável de corrente fuga a acurácia do modelo cai para aproximadamente 64%, que ressalta cada vez mais a importância desta variável no modelo. Para uma melhor visualização dos scores. A figura 4.18 abaixo ilustra os testes que foram realizados referente a regressão linear.

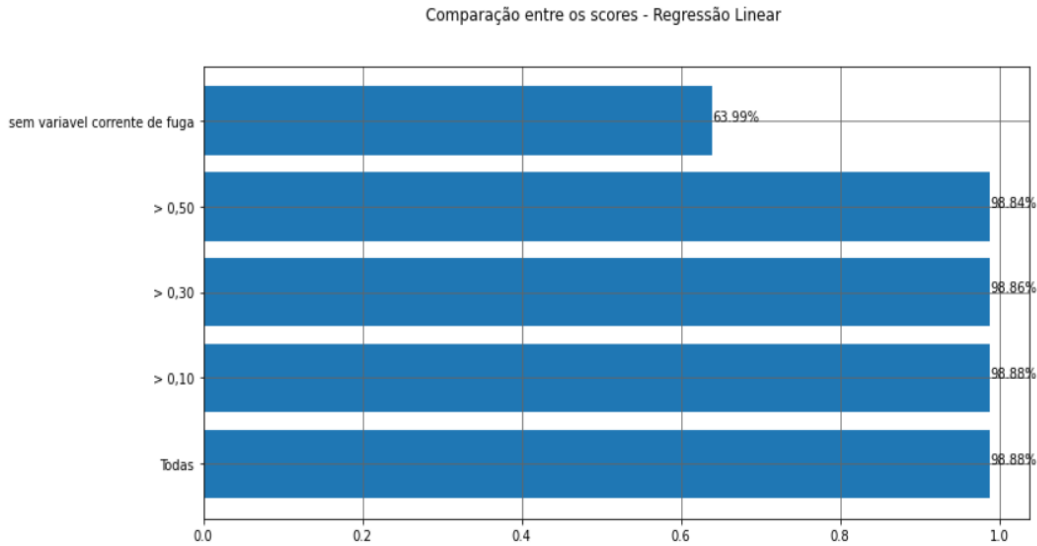


Figura 4.18 – Comparativo dos scores utilizando a regressão linear

4.3.2 – MODELO ÁRVORE DA DECISÃO

Outro algoritmo já estudado no capítulo 3 e seção 3.3.2, é o da árvore de decisão. Esse modelo é um tipo de método de aprendizado de máquina supervisionado e não paramétrico. Funcionando através de resultados de decisões tomadas durante determinado caminho.

De igual forma utilizou-se na regressão linear, 20% dos dados foram destinados para teste e os 80% para realização de treinamento. Com o auxílio da biblioteca *Tree* e ajustes nos parâmetros, foi realizado a predição e em seguida o cálculo dos scores para esse modelo, como mostram as figuras 4.19 e 4.20.

```
In [66]: colunas = df.columns.drop(['Uc'])
```

```
In [67]: X_train, X_test, y_train, y_test = train_test_split(df[colunas], df['Uc'], test_size=0.20, random_state=42)
```

Figura 4.19 – separação de dados para treino e teste

```
In [68]: tree = DecisionTreeRegressor()
```

```
In [69]: model_tree = tree.fit(X_train, y_train)
result_tree = tree.predict(X_test)
```

Figura 4.20 – Inclusão da biblioteca *tree* e ajustes dos parâmetros para predição

```
In [70]: mse_tree = mean_squared_error(y_test, result_tree)
rmse_tree = (np.sqrt(mse_tree))
score_tree = model_tree.score(X_test, y_test)
print('Resultados do erro médio e score (acurácia do modelo) para o dataset de treino:\n MSE: {}'.format(mse_tree)
      + '\n RMSE: {}'.format(rmse_tree)
      + '\nScore: {:.3f}'.format(score_tree * 100) + '%')
```

```
Resultados do erro médio e score (acurácia do modelo) para o dataset de treino:
MSE: 1.1417242860264717e-06
RMSE: 0.001068514991016257
Score: 99.951%
```

Figura 4.21 – Cálculo do score para o modelo

4.3.3 – MODELO RANDOM FOREST

O modelo random forest conforme foi estudado no capítulo 3, possui semelhanças com o modelo da árvore de decisão, visto que essa semelhança se dá pelo seu próprio nome que traduzido significa (floresta aleatória).

Baseado no princípio de funcionamento em *bagging*, esse algoritmo tem a capacidade de obter previsões mais acuradas tornando-se um modelo mais completo que os demais que já foram demonstrados, conforme veremos no resultado da simulação a seguir ilustrada na Figura 4.22.

```
In [71]: X_train, X_test, y_train, y_test = train_test_split(df[colunas], df['Uc'], test_size=0.20, random_state=42)
```

```
In [72]: tree_rf = RandomForestRegressor()
```

```
In [73]: model_rf = tree_rf.fit(X_train, y_train)
result_rf = model_rf.predict(X_test)
```

```
In [74]: mse_rf = mean_squared_error(y_test, result_rf)
rmse_rf = (np.sqrt(mse_rf))
score_rf = model_rf.score(X_test, y_test)
print('Resultados do erro médio e score (acurácia do modelo) para o dataset de treino:\n MSE: {}'.format(mse_rf)
      + '\n RMSE: {}'.format(rmse_rf)
      + '\nScore: {:.3f}'.format(score_rf * 100) + '%')
```

```
Resultados do erro médio e score (acurácia do modelo) para o dataset de treino:
MSE: 3.2853692524400067e-07
RMSE: 0.0005731814069245448
Score: 99.986%
```

Figura 4.22 – Modelo random forest e cálculo dos scores

Conforme previsto, o modelo apresentou uma acurácia melhor do que os demais e um ponto importante a ser destacado do modelo, é que por ser mais completo ainda sim, apresenta uma linguagem de programação básica, tornando ainda mais interessante o uso desse algoritmo em outros tipos de problemas e análises que desejam ser solucionados através de Machine Learning.

4.3.4 – MODELO RANDOM FOREST COM AJUSTE NOS HIPERPARÂMETROS.

O modelo de random forest tem a capacidade de possuir certas alterações que geram uma melhora significativa no algoritmo, tornando-o ainda mais preciso. Essas alterações se devem à inclusão de parâmetros para serem testados. Conforme descrito no capítulo 3, incluímos os parâmetros:

- *Min_samples_leaf*: número de amostras mínimas ao nível da folha,
- *Min_samples_split*: número de amostras mínimas para considerar um nó para divisão.
- *n_estimator*: número de árvores construídas pelo algoritmo antes de tomar uma votação ou fazer uma média de previsões:

```
In [75]: X_train, X_test, y_train, y_test = train_test_split(df[colunas], df['Uc'], test_size=0.20, random_state=42)

In [77]: param = {'min_samples_leaf':[1,3,5], 'min_samples_split':[2,3,5], 'n_estimators':[50,150,250]}
rf = RandomForestRegressor()
grid = GridSearchCV(rf, param, n_jobs=4)
```

Figura 4.23 – Criação de parâmetros a serem testados

Para a execução do treinamento do modelo, utilizou-se o comando *grid*. Esses ajustes no hiper parâmetro do modelo tornou a programação um pouco mais complexa e a obtenção do resultado mais demorada.

```
In [86]: grid.fit(X_train, y_train)

Out[86]: GridSearchCV(estimator=RandomForestRegressor(), n_jobs=4,
                      param_grid={'min_samples_leaf': [1, 3, 5],
                                   'min_samples_split': [2, 3, 5],
                                   'n_estimators': [50, 150, 250]})
```

Figura 4.24– Execução do Grid

Há a possibilidade do algoritmo de detectar qual a melhor combinação dos valores dos parâmetros que foram testados.

```
In [103]: grid.best_params_
Out[103]: {'min_samples_leaf': 1, 'min_samples_split': 2, 'n_estimators': 150}
```

Figura 4.25– Verificação dos melhores parâmetros

Em seguida, realizou-se o teste e cálculo da acurácia do modelo utilizando apenas os melhores parâmetros e verificou-se uma melhor acurácia do modelo.

```
In [105]: result_grid = rf_best.predict(X_test)

In [106]: mse_grid = mean_squared_error(y_test, result_grid)
rmse_grid = (np.sqrt(mse_grid))
score_grid = rf_best.score(X_test, y_test)
print('Resultados do erro médio e score (acurácia do modelo) para o dataset de treino:\n MSE: {}'.format(mse_grid)
      + '\n RMSE: {}'.format(rmse_grid)
      + '\nScore: {:.2f}'.format(score_grid * 100) + '%')
```

Resultados do erro médio e score (acurácia do modelo) para o dataset de treino:
MSE: 3.067882303130064e-07
RMSE: 0.0005538846723939979
Score: 99.987%

Figura 4.26 – Cálculo do score com os melhores parâmetros.

4.4 – RESULTADOS

O resultado que pode ser tirado das simulações realizadas em usando técnicas de Machine Learning é que todas possuem um score considerável para a solução do problema proposto, visto que a peça principal para a predição da tensão crítica de flashover está no valor da sua corrente de fuga, sendo essa a chave causadora do problema de falha no isolador e consequentemente o desligamento na rede de distribuição. Abaixo segue o gráfico comparativo entre os algoritmos simulados.

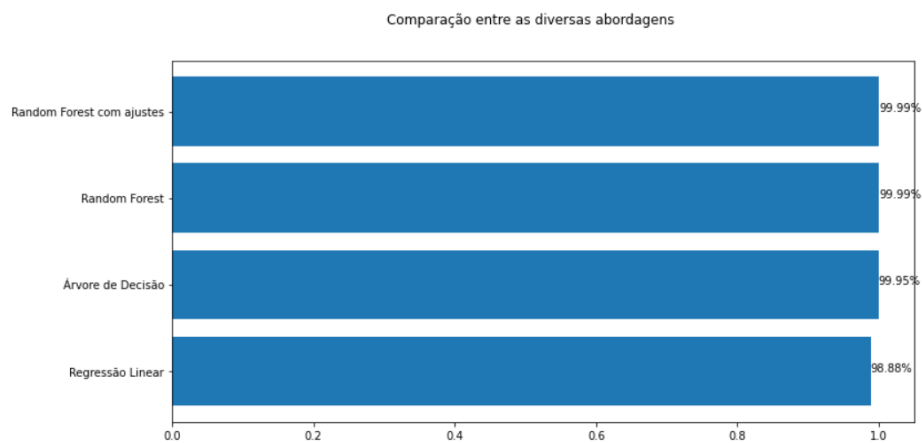


Figura 4.27 – Comparação entre os modelos simulados

O modelo Random Forest com ajuste no hiper parâmetros apresentou uma melhor acurácia para o modelo, porém não tão diferente dos demais, que também apresentaram um score positivo e que podem ser usados.

A seguir a comparação entre os valores que foram treinados do modelo com os valores possivelmente previstos dos dez primeiros registros.

```
In [108]: df_comparacao = pd.DataFrame({'Atual': y_test, 'Previsto': result_grid})
df_comparacao.head(10)
```

```
Out[108]:
```

	Atual	Previsto
4943	0.540900	0.540899
30083	0.507312	0.507311
25353	0.545586	0.545585
26959	0.535244	0.535245
18264	0.538270	0.538269
19244	0.568010	0.568010
30002	0.513142	0.513141
39360	0.574495	0.574495
2265	0.549878	0.549878
34139	0.634531	0.634540

Figura 4.28 – Comparação do valor real e valor previsto

O gráfico comparativo abaixo mostra o resultado dos valores reais dos valores previsto pelo modelo dos cem primeiros registros, reforçando cada vez mais a eficiência de modelo como solução para esse tipo de falhas em isoladores.

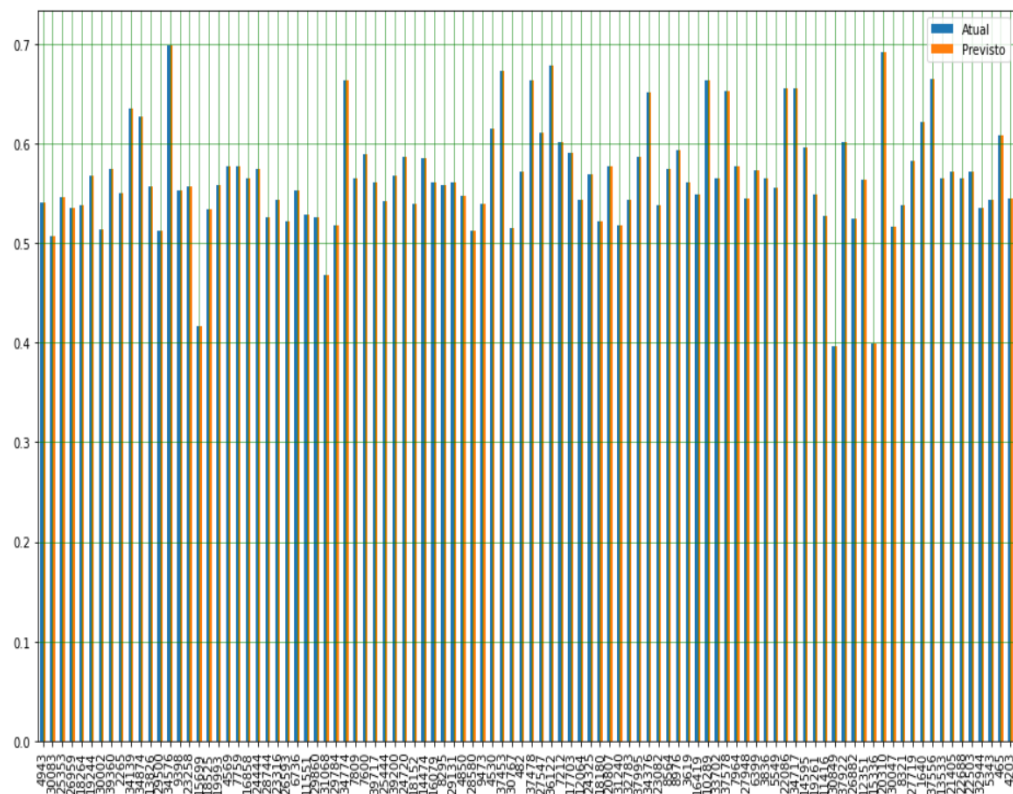


Gráfico 4.3 – Comparativo registros dos valores atuais e previstos

Por fim, o modelo obteve sucesso na sua simulação mostrando uma alternativa para a prevenção do defeito em isoladores da rede de distribuição e assim apresentando uma solução para possíveis outros problemas que possam vir a ocorrer. Porém,

analisando pelo lado real acaba se tornando uma solução de difícil acesso nos dias de hoje, pois para instalação sensores e equipamentos que possam determinar as condições reais do ambiente e o monitoramento da corrente de fuga desses isoladores, seria necessário um esforço financeiro muito alto e fazer o desligamento da rede de distribuição. Como proposta para trabalhos futuros, uma solução que possa vir a resolver problemas como esses e que possa automatizar e modernizar o sistema de distribuição de energia no geral.

REFERÊNCIAS BIBLIOGRÁFICAS

- ABDELAZIZ, E. Prevention of the Interruptions due to the Phenomena of the Electric Insulators Pollution. in First International Symposium on Control, Communications and Signal Processing, 2004.
- AMPADU, H. Ai pool. Random Forest Understanding. Disponível em: <https://ai-pool.com/a/s/random-forests-understanding>. 2021. Acessado em 29 de setembro de 2021
- BEZERRA, D. revestimento de isoladores elétricos utilizando borracha de silicone/alumina trihidratada/ nanosílica, 2014.
- BONOMO, A. Estudo sobre isoladores poliméricos à base de resinas poliuretanas para ambientes externos, 2003
- CABOT. Borrachas de silicone (HTV e LSR). Copyright© 1995-2019 Cabot Corporation. Disponível em: [https://www.cabotcorp.com.br/Borrachas de silicone \(HTV e LSR\) | Cabot Corporation](https://www.cabotcorp.com.br/Borrachas%20de%20silicone%20(HTV%20e%20LSR)%20|%20Cabot%20Corporation). Acesso em: 17 setembro de 2021.
- CAETANO. M, Modelos de classificação - Aplicações no setor bancário. 2015
- CAMPOS, R. Árvores de decisão. Machine Learning Beyond Deep Learning. 28 de novembro de 2017. Disponível em: <https://medium.com/machine-learning-beyond-deep-learning/árvores-de-decisão-3f52f6420b69>. Acessado em 02 de setembro de 2021
- DAMASCENO, L. Regressão Linear, 2020. Disponível em: <https://medium.com/@lauradamaceno/regressão-linear-6a7f247c3e29>. Acessado em 20 de setembro de 2021.
- ENERGY, L. Instalações elétricas internas – Condutores elétricos, 2019
- FONTANA, E. Introdução aos algoritmos de aprendizado supervisionados. 2020
- GONOS, I. F; TOPALIS, V.; STATHOPULOS, A. Genetic algorithm approach to the modeling of polluted insulators, 2012
- KARADY, G. G. e SHAH, M. Flashover Mechanism on Sillicone Rubber Insulators Used for Outdoor Insulation – I. IEEE Transactions on Power Delivery, v.10, n.4, 1995, p. 1965-1971.

LIMA, R.A. Contribuições para o desenvolvimento de um sistema de monitoração via satélite do estado de poluição de cadeias de isoladores de alta tensão. Dissertação de pós-Graduação em engenharia elétrica, Universidade Federal de Pernambuco, 2010.

MELLO, Darcy R. de. Dimensionamento de isoladores sob condições de poluição. Curso de isoladores. Comitê Nacional Brasileiro de Produção e Transmissão de Energia Elétrica. CIGRÉ - Brasil. Rio de Janeiro, 20 e 21 jun. 2017.

MULLER, Andreas C; GUIDO, Sarah. Introduction to machine learning with python. 1ª edição. Outubro de 2016.

PEELO, D. F. Current interruption using high voltage air-break disconnectors. Technische Universiteit Eindhoven, 2004

PEREIRA, A. A. estudo da aplicação de isoladores elétricos em linhas de transmissão e subestações de acordo com o grau de poluição do ambiente, 2020.

PINOTTI, R. H. modelo para predição de correntes de fuga em isoladores de distribuição classe 25 kv. Trabalho de conclusão de curso, pós-graduação, faculdade pontifícia universidade católica de minas gerais, 2020.

QUEIRÓS, José Pedro M. Análise comparativa do comportamento dielétrico dos isoladores compósitos com isoladores em materiais convencionais. Dissertação de Mestrado, Faculdade de Engenharia da Universidade do Porto. Porto, 2013

REDAÇÃO CRONAPP. Algoritmos de machine learnig: conheça os 6 principais e descubra os impactos na sua empresa. 8 de outubro de 2019. Disponível em: <https://blogs.cronapp.io/algoritmos-de-machine-learning/>. Acessado em 08 de agosto de 2021

RORDIGUES et al. Avaliação do Desempenho Elétrico de Isoladores de Vidro Revestidos por Verniz Nanométrico. Publicações Cgti.org. Acessado em 19 de setembro de 2021.

RUSSEL, S; NORVIG, P. Inteligência artificial. 3ª edição. Elsevier Editora LTDA. 2013

STC SILICONES. Transmissão e distribuição de energia. Disponível em:<https://www.stcsilicones.com.br/transmissão-e-distribuição-de-energia>. Acessado em 28 de agosto de 2021

STEFENON, S. estudo das técnicas de inspeção de redes de distribuição para identificação e classificação de defeitos em isoladores através do uso do ultrassom, 2015

WILKINS, R.: 'Flashover voltage of high voltage insulators with uniform surface pollution films', Proc. IEE, 1969, 116, (3), pp. 457–465