

Matheus Moura Fernandes

# **Análise de Falhas em Sistemas Elétricos de Distribuição: Uma Abordagem Baseada em Redes Bayesianas.**

Divinópolis - Brasil

8 de fevereiro de 2025

Matheus Moura Fernandes

# **Análise de Falhas em Sistemas Elétricos de Distribuição: Uma Abordagem Baseada em Redes Bayesianas.**

Monografia apresentada ao Curso de Engenharia de Computação da UEMG Unidade Divinópolis, como requisito parcial para obtenção do título de Bacharel em Engenharia da Computação, sob a orientação do Prof. Itallo Guilherme Machado e coorientação do Prof. Marcos Alberto Saldanha

Universidade do Estado de Minas Gerais – UEMG

Unidade Divinópolis

Curso de Engenharia da Computação

Orientador: Itallo Guilherme Machado

Coorientador: Marcos Alberto Saldanha

Divinópolis - Brasil

8 de fevereiro de 2025

---

Matheus Moura Fernandes

Análise de Falhas em Sistemas Elétricos de Distribuição: Uma Abordagem Baseada em Redes Bayesianas./ Matheus Moura Fernandes. – Divinópolis - Brasil, 8 de fevereiro de 2025-

52 p.

Orientador: Itallo Guilherme Machado

Monografia – Universidade do Estado de Minas Gerais – UEMG

Unidade Divinópolis

Curso de Engenharia da Computação, 8 de fevereiro de 2025.

1. Palavra-chave1. 2. Palavra-chave2. I. Orientador. II. Universidade xxx. III. Faculdade de xxx. IV. Título

Número na biblioteca

---

Matheus Moura Fernandes

## **Análise de Falhas em Sistemas Elétricos de Distribuição: Uma Abordagem Baseada em Redes Bayesianas.**

Monografia apresentada ao Curso de Engenharia de Computação da UEMG Unidade Divinópolis, como requisito parcial para obtenção do título de Bacharel em Engenharia da Computação, sob a orientação do Prof. Itallo Guilherme Machado e coorientação do Prof. Marcos Alberto Saldanha

Trabalho aprovado. Divinópolis - Brasil, 8 de fevereiro de 2025:

---

**Itallo Guilherme Machado**  
Orientador

---

**Professor**  
Convidado 1

---

**Professor**  
Convidado 2

Divinópolis - Brasil  
8 de fevereiro de 2025

*Este trabalho é dedicado ao meu menino interior, que, tempos atrás, lia a revista Ciência Hoje e acreditava que se tornaria um cientista; que assistia Fullmetal Alchemist e via na ciência a maior de todas as magias.*

# Agradecimentos

O desenvolvimento deste trabalho foi um grande desafio e, ao longo dessa jornada, contei com o apoio inestimável de pessoas que contribuíram, direta e indiretamente, para a conclusão do curso de Engenharia da Computação. A todas elas, expresso minha mais profunda gratidão.

Aos meus pais, Sidene Moura Fernandes e Walter Fernandes Leite, por seu amor incondicional, pelo incentivo constante e pelos valores que me ensinaram, fundamentais para minha trajetória. Agradeço também pelo investimento em minha formação acadêmica e, principalmente, por acreditarem em mim em todos os momentos, especialmente nos mais difíceis, quando a insegurança parecia prevalecer.

Aos meus irmãos, Heider Moura Fernandes e Flaviane Moura Fernandes, pelo companheirismo e pelas palavras de motivação que sempre me impulsionaram a seguir em frente. Vocês são minha maior inspiração.

À UEMG Divinópolis e a seus servidores, especialmente aos professores do curso de Engenharia da Computação, pelo comprometimento com um ensino público de qualidade e pela dedicação em formar profissionais capacitados e críticos.

Ao CEFET-MG Campus X Curvelo, aos seus funcionários e, em especial, aos professores do Departamento de Eletrotécnica, pela excelência no ensino público e por me prepararem para os desafios do ensino superior. Um agradecimento especial aos auxiliares de limpeza e manutenção Evaldo, Elbert e Adélia, por sempre me acolherem com carinho, pelas conversas enriquecedoras e pelo cuidado com meus pertences durante minha permanência na instituição. Às cozinheiras do restaurante universitário, pelo compromisso em garantir uma alimentação saudável e de qualidade, o que refletiu diretamente no meu desempenho acadêmico entre 2016 e 2018.

Ao meu orientador, Itallo Guilherme Machado, pela paciência, dedicação e valiosos ensinamentos, fundamentais para a construção deste trabalho. Ao meu coorientador, Marcos Alberto Saldanha, por suas contribuições e pelo olhar crítico que ajudou a aperfeiçoar cada etapa da pesquisa. Ambos são exemplos de profissionais que desejo seguir em minha futura trajetória.

Aos meus amigos de faculdade, Isabela Maria Barros e Paulo Henrique Gonçalves Silva, pela parceria, pelo apoio mútuo nas horas de estudo e pelas conversas que tornaram essa caminhada mais leve e prazerosa.

Aos meus companheiros de viagem nos ônibus, Luiz Felipe, Cleyton, Hélio, João Paulo, Mayara, Rosângela e Jaques, por todas as conversas que ajudaram a tornar as

longas viagens mais suportáveis e, acima de tudo, pelo incentivo e apoio na busca por estágio, sempre compartilhando oportunidades e me motivando a persistir.

Por fim, a todos que, de alguma forma, contribuíram para que este trabalho fosse concluído, deixo aqui minha mais sincera gratidão.

*"Nada na vida deve ser temido, somente compreendido. Agora é hora de compreender  
mais para temer menos."*

**Marie Curie**



# Sumário

<b>1</b>	<b>INTRODUÇÃO</b>	<b>1</b>
<b>1.1</b>	<b>Objetivos do Trabalho</b>	<b>2</b>
1.1.1	Objetivo Geral	2
1.1.2	Objetivos Específicos	2
<b>1.2</b>	<b>Justificativa</b>	<b>3</b>
<b>1.3</b>	<b>Estrutura da Monografia</b>	<b>3</b>
<b>2</b>	<b>ESTADO DA ARTE</b>	<b>4</b>
	Estado da Arte	4
<b>3</b>	<b>FUNDAMENTAÇÃO TEÓRICA</b>	<b>7</b>
<b>3.1</b>	<b>Interrupções em redes de distribuição elétrica</b>	<b>7</b>
<b>3.2</b>	<b>Redes Bayesianas</b>	<b>8</b>
3.2.1	Definições de probabilidade condicional e independência condicional	9
3.2.2	Definição do Teorema de Bayes	10
3.2.3	Modelagem da Rede	11
3.2.4	Inferência	13
<b>3.3</b>	<b>Algoritmos de Aprendizagem de Estrutura</b>	<b>13</b>
<b>4</b>	<b>MATERIAIS E MÉTODOS</b>	<b>14</b>
<b>4.1</b>	<b>Classificação de pesquisa</b>	<b>14</b>
<b>4.2</b>	<b>Ferramentas e tecnologias</b>	<b>14</b>
4.2.1	<i>Python</i>	14
4.2.2	<i>Jupyter Notebook</i>	15
<b>4.3</b>	<b>Coleta e Pré-Processamento dos Dados</b>	<b>15</b>
4.3.1	Variáveis Consideradas	15
4.3.2	Biblioteca <i>NumPy</i>	16
4.3.3	Biblioteca <i>Pandas</i>	16
<b>4.4</b>	<b>Aplicação do Modelo Bayesiano</b>	<b>17</b>
4.4.1	Modelagem da Rede Bayesiana	17
<b>4.5</b>	<b>Implementação</b>	<b>18</b>
4.5.1	Modelagem Bayesiana	18
4.5.2	Manipulação de Grafos	18
4.5.2.1	Biblioteca <i>NetworkX</i>	18
4.5.2.2	Biblioteca <i>Matplotlib</i>	19

4.5.3	Inferência por Eliminação de Variáveis . . . . .	20
<b>5</b>	<b>DESENVOLVIMENTO . . . . .</b>	<b>21</b>
	<b>Desenvolvimento . . . . .</b>	<b>21</b>
<b>5.1</b>	<b>Pré-processamento e Tratamento dos Dados . . . . .</b>	<b>21</b>
5.1.1	Seleção de variáveis e dados . . . . .	21
5.1.2	Discretização de Dados . . . . .	22
5.1.3	Reconfiguração das colunas . . . . .	22
<b>5.2</b>	<b>Estruturação da Rede Bayesiana . . . . .</b>	<b>23</b>
5.2.1	Algoritmos de Aprendizado Estrutural . . . . .	24
5.2.2	Implementação dos Métodos de Inferência . . . . .	24
<b>5.3</b>	<b>Implementação gráfica de Grafos e resultados . . . . .</b>	<b>25</b>
<b>6</b>	<b>ANÁLISE DOS RESULTADOS . . . . .</b>	<b>27</b>
	<b>Análise dos Resultados . . . . .</b>	<b>27</b>
<b>6.1</b>	<b>Estruturas Apredidas . . . . .</b>	<b>27</b>
<b>6.2</b>	<b>Inferências Condicionais . . . . .</b>	<b>30</b>
<b>6.3</b>	<b>Resultados para estrutura <i>Hill Climbing</i> . . . . .</b>	<b>31</b>
6.3.1	Inferências Simples Método <i>Hill Climbing</i> . . . . .	31
6.3.2	Inferências por Evidências – Método <i>Hill Climbing</i> . . . . .	33
<b>6.4</b>	<b>Resultados <i>Max-Min Parents and Children (MMPC)</i> . . . . .</b>	<b>35</b>
6.4.1	Inferências Simples Método <i>Max-Min Parents and Children (MMPC)</i> . . . . .	35
6.4.2	Inferências por Evidências – Método <i>Max-Min Parents and Children (MMPC)</i> . . . . .	36
<b>7</b>	<b>CONCLUSÃO . . . . .</b>	<b>39</b>
	<b>Conclusão . . . . .</b>	<b>39</b>
	<b>REFERÊNCIAS . . . . .</b>	<b>40</b>
	<b>APÊNDICES . . . . .</b>	<b>43</b>
	<b>APÊNDICE A – TABELAS DE INFERÊNCIAS . . . . .</b>	<b>44</b>
	<b>APÊNDICE B – GRÁFICOS DE BARRAS . . . . .</b>	<b>45</b>
	<b>APÊNDICE C – GRÁFICOS BOXPLOT . . . . .</b>	<b>46</b>
	<b>APÊNDICE D – GRÁFICOS FUNÇÃO DE DISTRIBUIÇÃO ACUMULADA . . . . .</b>	<b>47</b>

<b>ANEXOS</b>	<b>48</b>
<b>ANEXO A – INTERRUPÇÕES DE ENERGIA ELÉTRICA NAS RE-</b> <b>DES DE DISTRIBUIÇÃO . . . . .</b>	<b>49</b>

# Lista de ilustrações

Figura 1 – Directed Acyclic Graph - DAG. Fonte: Elaborada pelo Autor através do site Miror. . . . .	11
Figura 2 – Modelo de uma Rede Bayesiana. Fonte: Elaborada pelo Autor através do site Miror. . . . .	12
Figura 3 – Logo <i>Python</i> . Fonte: (Python Software Foundation, 2024) . . . . .	15
Figura 4 – Logo Jupyter <i>Notebook</i> . Fonte: (Project Jupyter, 2024) . . . . .	15
Figura 5 – Logo <i>NumPy</i> . Fonte: (NumPy, 2024) . . . . .	17
Figura 6 – Logo <i>Pandas</i> . Fonte: (Pandas Development Team, 2024) . . . . .	17
Figura 7 – Logo <i>pgmpy</i> . Fonte: (pgmpy Development Team, 2024) . . . . .	19
Figura 8 – Logo <i>NetworkX</i> . Fonte: (NetworkX Development Team, 2024) . . . . .	19
Figura 9 – Logo <i>Matplotlib</i> . Fonte: (Matplotlib Development Team, 2024) . . . . .	19
Figura 10 – Fluxograma de Desenvolvimento. Fonte: Autoria própria. . . . .	26
Figura 11 – Estrutura Bayesiana <i>Hill Climbing</i> . Fonte: Autoria própria gerada pela biblioteca <i>Matplotlib</i> . . . . .	27
Figura 12 – Estrutura Bayesiana ( <i>Max-Min Parents and Children - MMPC</i> ). Fonte: Autoria própria gerada pela biblioteca <i>Matplotlib</i> . . . . .	29

# Lista de tabelas

Tabela 1 – Propriedades da Base de Dados . . . . .	16
Tabela 2 – Classificação da Duração das Interrupções . . . . .	22
Tabela 3 – Classificação do Nível de Tensão . . . . .	23
Tabela 4 – Descrição das Colunas e Correspondência com os Nós . . . . .	23
Tabela 5 – Distribuição de Probabilidades para Tipos de Interrupção . . . . .	31
Tabela 6 – Distribuição de Probabilidades para os Motivos das Interrupções . . . . .	31
Tabela 7 – Distribuição de Probabilidades para Níveis de Tensão . . . . .	32
Tabela 8 – Distribuição de Probabilidades para as Concessionárias de Energia . . . . .	32
Tabela 9 – Tabela de Probabilidades para a duração da interrupção . . . . .	32
Tabela 10 – Tabela de Probabilidades para o turno em que a interrupção ocorre. . . . .	32
Tabela 11 – Distribuição de Probabilidades de Ocorrência das Interrupções por Mês . . . . .	33
Tabela 12 – Tabela de Probabilidades para a Variável Inferida D considerando Diferentes Evidências de E . . . . .	34
Tabela 13 – Tabela de Probabilidades para a Variável Inferida D considerando Diferentes Evidências de B . . . . .	34
Tabela 14 – Tabela de Probabilidades para a Variável Inferida D considerando Diferentes Evidências de C . . . . .	34
Tabela 15 – Tabela de Probabilidades para a Variável A . . . . .	35
Tabela 16 – Tabela de Probabilidades para a Variável B . . . . .	35
Tabela 17 – Tabela de Probabilidades para a Variável C . . . . .	35
Tabela 18 – Tabela de Probabilidades para a Variável E . . . . .	35
Tabela 19 – Tabela de Probabilidades para a Variável G . . . . .	36
Tabela 20 – Tabela de Probabilidades para a Variável Inferida A considerando Diferentes Evidências de C . . . . .	37
Tabela 21 – Tabela de Probabilidades para a Variável Inferida C considerando Diferentes Evidências de B . . . . .	37
Tabela 22 – Tabela de Probabilidades para a Variável Inferida C considerando Diferentes Evidências de G . . . . .	38

# Lista de abreviaturas e siglas

ABNT	Associação Brasileira de Normas Técnicas
abnTeX	ABsurdas Normas para TeX
ANEEL	Agência Nacional de Energia Elétrica
DAG	<i>Directed Acyclic Graph</i> (Grafo Acíclico Dirigido)
DEC	Duração Equivalente de Interrupção por Unidade Consumidora
FEC	Frequência Equivalente de Interrupção por Unidade Consumidora
GMs	Modelos Gráficos
HMM	<i>Hidden Markov Models</i> (Modelos Ocultos de Markov)
IoT	<i>Internet of Things</i> (Internet das Coisas)
RBs	Redes Bayesianas

# Lista de símbolos

$\Gamma$	Letra grega Gama
$\Lambda$	Lambda
$\zeta$	Letra grega minúscula zeta
$\in$	Pertence
$D_1, D_2, D_3$	Representam doenças ou variáveis causais (Gripe, Resfriado, Covid-19).
$S$	Representa o sintoma ou efeito (Tosse).
$P$	Função de probabilidade.
$A, B$	Eventos genéricos em teoria de probabilidade.
$X, Y, Z$	Variáveis ou eventos em redes Bayesianas.
$n, n_A, n_B, n_{A \wedge B}$	Representam contagens em um experimento (como número de vezes que eventos ocorrem).
$A_i, A_j$	Partes do espaço amostral S em probabilidade.
$P(A)$	Probabilidade de evento A.
$P(A B)$	Probabilidade condicional de A dado B.
$P(A \wedge B)$	Probabilidade de ocorrência simultânea de A e B.
$P(A \vee B)$	Probabilidade de A ou B ocorrer (ou lógica).
$\Pi$	Produto de uma sequência de termos.
$\Sigma$	Soma de uma sequência de termos.
$\vee$	Operador lógico "ou".
$\wedge$	Operador lógico "e".
$=$	Igualdade matemática.

# Resumo

No contexto histórico, a energia elétrica é de suma importância para o desenvolvimento social moderno, além de se mostrar fator determinante para o desempenho econômico nacional. Tendo em vista a relevância, o setor elétrico brasileiro têm normativas apropriadas para resguardar a distribuição de energia elétrica com continuidade e qualidade aos setores residencial, industrial e agrícola. Tal regulamentação é estabelecida pela ANEEL (*Agência Nacional de Energia Elétrica*), impondo assim parâmetros de qualidade às concessionárias de distribuição de energia elétrica. Nesse cenário, observa-se que as interrupções elétricas estão diretamente relacionadas à eventos como falhas operacionais e em equipamentos, causas naturais e meteorológicas. Posto isso, este trabalho tem como objetivo apresentar uma análise das interrupções em redes de distribuição elétrica no Brasil, utilizando modelos bayesianos e técnicas de *Machine Learning*. Com base nos dados fornecidos pela *Agência Nacional de Energia Elétrica* (ANEEL), o estudo visa diagnosticar falhas e prever possíveis causas para interrupções nas redes de distribuição elétrica. A aplicação de Modelos Ocultos de Markov (*HMM*) e Redes Bayesianas (*BN*) permite a análise eficiente das redes de distribuição elétrica, considerando fenômenos naturais e a necessidade de confiabilidade no fornecimento de energia.

**Palavras-chaves:** Redes de Distribuição. Interrupções Elétricas. Modelos Bayesianos. *Machine Learning*.



# Abstract

In the historical context, electrical energy is of paramount importance for modern social development, in addition to being a determining factor for national economic performance. In view of the relevance, the Brazilian electrical sector has appropriate regulations to safeguard the distribution of electrical energy with continuity and quality to the residential, industrial and agricultural sectors. This regulation is established by ANEEL (National Electric Energy Agency), which imposes quality standards on electricity distribution concessionaires. In this scenario, it is observed that electrical interruptions are directly related to events such as operational and equipment failures, natural, and meteorological causes. That said, this work aims to present an analysis of interruptions in electrical distribution networks in Brazil, using Bayesian models and Machine Learning techniques. Based on data provided by the National Electric Energy Agency (ANEEL), the study aims to diagnose faults and predict possible causes of interruptions in electrical distribution networks. The application of Hidden Markov Models (HMM) and Bayesian Networks (BN) allows the efficient analysis of electrical distribution networks, considering natural phenomena and the need for reliability in energy supply.

**Keywords:** Distribution Networks. Electrical Outages. Bayesian models. Machine Learning.

# 1 Introdução

Sob a perspectiva histórica, o sistema de transmissão elétrica tornou-se indispensável para o desenvolvimento agrícola e industrial, tendo em vista que a energia elétrica é a segunda maior fonte energética consumida no Brasil, ficando atrás somente das energias derivadas da queima de petróleo, de acordo com a ([Empresa de Pesquisa Energética \(EPE\)](#), 2024). Segundo [Souza et al. \(2013\)](#) e [Neto, Corrêa e Perobelli \(2019\)](#), o crescente consumo de eletricidade é um movimento global associado ao desenvolvimento dos corpos sociais e à qualidade de vida dos indivíduos.

Apesar dos avanços tecnológicos, as falhas em sistemas elétricos de distribuição apresentam um alto grau de incerteza, tendo como principal causa as descargas atmosféricas. Observa-se que o desempenho das redes de energia elétrica está diretamente condicionado a eventos meteorológicos adversos. Assim, o crescimento dos núcleos populacionais reflete no aumento das redes de transmissão e distribuição no sistema elétrico brasileiro.

Diante desse cenário, destaca-se como desafio determinante para as concessionárias de energia o fornecimento sustentável de eletricidade, garantindo sua confiabilidade. O sistema elétrico brasileiro é composto por empresas independentes de geração, operadores de linhas de transmissão, distribuição e comercializadoras, formando um sistema interligado. O propósito final dessas empresas é atender consumidores, sejam eles pessoas físicas ou jurídicas, segundo a [Neoenergia \(2019\)](#).

Além disso, enfatiza-se que o setor elétrico brasileiro incentiva constantemente a competitividade entre as empresas integrantes. As concessionárias têm a obrigação de garantir a disponibilidade de eletricidade para seus consumidores, observando a qualidade e o crescimento da demanda. Nesse contexto, a confiabilidade e a sustentabilidade do fornecimento de energia elétrica, associadas à qualidade do serviço, são aspectos fundamentais para o funcionamento do sistema elétrico. Esses fatores se tornaram ainda mais relevantes devido à regulamentação da [Agência Nacional de Energia Elétrica \(ANEEL\) \(2007\)](#), que define penalizações para indisponibilidades nos sistemas de transmissão. Com isso, um diagnóstico eficiente de falhas permite a aplicação de ações corretivas que minimizem essas penalizações.

As práticas tradicionais de diagnóstico de interrupções baseiam-se nos julgamentos dos operadores, considerando sua experiência e conhecimento prático. As subestações de energia elétrica automatizadas operam por meio de sinais de alarme que indicam anormalidades no sistema durante a ocorrência de falhas. O estado dos equipamentos é definido pela atuação dos relés e disjuntores, sendo posteriormente avaliado pelos operadores, que determinam o status do circuito e a natureza do defeito (detecção,

classificação e localização).

Dado o aumento da complexidade dos sistemas de distribuição atuais, a atuação eficiente diante de falhas tornou-se um dos principais desafios das concessionárias, pois os procedimentos adotados dependem diretamente da análise dos operadores, tornando as avaliações suscetíveis a erros humanos. Conforme indicado por [Librantz et al. \(2020\)](#), metodologias baseadas em *Redes Bayesianas* são capazes de extrair conhecimento útil, sendo aplicadas em processos de alta complexidade, como o diagnóstico de falhas. No contexto do gerenciamento de *Big Data*, paradigma que envolve o processamento de grandes volumes de dados, diversas ferramentas têm sido desenvolvidas para a análise e extração de informações essenciais, aplicando-se em áreas variadas ([SAGIROGLU et al., 2016](#); [BHATTARAI et al., 2019](#); [GHORBANIAN; DOLATABADI; SIANO, 2019](#))

## 1.1 Objetivos do Trabalho

Nesta seção, são expostos o objetivo geral e os objetivos específicos almejados ao longo do trabalho proposto.

### 1.1.1 Objetivo Geral

Este trabalho tem como objetivo principal desenvolver uma aplicação de *machine learning*, por meio de técnicas de modelos de redes bayesianas, buscando identificar padrões que possam auxiliar na previsão e diagnóstico de falhas no sistema elétrico de distribuição, relacionando eventos probabilísticos com indicadores referentes às interrupções elétricas.

### 1.1.2 Objetivos Específicos

- a) Coletar e estruturar dados históricos sobre falhas em redes elétricas de distribuição, considerando fatores meteorológicos, operacionais e estruturais;
- b) Modelar a estrutura de uma *Rede Bayesiana* para representar as relações probabilísticas entre as variáveis associadas às interrupções;
- c) Avaliar a precisão dos modelos desenvolvidos por meio de métricas como *BIC Score*;
- d) Comparar os resultados obtidos com métodos tradicionais de diagnóstico de falhas em sistemas de distribuição;
- e) Testar métodos alternativos de controle de falhas elétricas e uma arquitetura viável baseada no diagnóstico de falhas por redes bayesianas.

## 1.2 Justificativa

O sistema energético brasileiro está sujeito ao mal funcionamento devido a diversas perturbações, causados em sua maioria por descargas elétricas, rompimentos periódicos de cadeias de isoladores, acedentes, incêndios e queimadas, o que geram faltas do tipo curto-circuito (monofásicos, bifásicos e trifásicos), aspectos avaliados por (CUNHA, 2010). Posto esse cenário, dispositivos capazes de detectar e diagnosticar são de suma importância para o funcionamento sustentável e eficiente das linhas de transições e distribuições, tendo em vista o grau de importância do fornecimento constante de energia elétrica, com isso, esses dispositivos devem operar em menor tempo, preservando a integridade física e mental dos sistemas e de seus operadores.

A identificação e diagnóstico de falhas é um entrave de característica complexas, tendo em vista a incertezas para possíveis causas das perturbações no sistema, a tipologia de rede ser altamente ramificada, variações de impedâncias, e descargas atmosféricas. Diante essas dificuldades, o desenvolvimento dos equipamentos para aquisições, processamento e transmissão de sinais elétrico possibilita o emprego de técnicas e algoritmos para o diagnóstico de equipamentos em redes de distribuições de energia elétrica.

Com o objetivo de diminuir o tempo de diagnóstico de faltas em linhas de distribuições, além de preservar a saúde dos operadores de linhas de distribuições, este trabalho propõem uma metodologia baseada em Redes Bayesianas (RBs) que busca modelar falhas relativas consoante à Resolução Normativa ANEEL 270/07 (Agência Nacional de Energia Elétrica (ANEEL), 2007), visando a qualidade dos serviços prestados.

## 1.3 Estrutura da Monografia

O presente trabalho está dividido em seis capítulos. O Capítulo 1 apresenta a Introdução. No Capítulo 2 são apresentados os trabalhos relacionados ao tema proposto. No Capítulo 3 é realizada uma revisão da literatura e fundamentação teórica com ênfase nos conceitos de Modelos Ocultos de Markov (Hidden Markov Models - HMM) e Redes Bayesianas (Bayesian Networks). No Capítulo 4 são apresentados os materiais e métodos utilizados no trabalho. No Capítulo 5 é apresentado o desenvolvimento do trabalho. No Capítulo 6 são apresentados os resultados obtidos. Por fim, no Capítulo 7 são apresentadas as conclusões e sugestões de trabalhos futuros.

## 2 Estado da Arte

Com o propósito de desenvolver análises de dados referente à base de dados sobre interrupções em redes de distribuições elétricas, fornecida pela Agência Nacional de Energia Elétrica (ANEEL), foram realizadas pesquisas relacionadas ao desenvolvimento de análises de dados e aplicação de redes neurais no *Google Scholar* e no banco de dissertações e teses da Capes, bem como no livro *Redes Neurais: Princípios e Práticas*, Simon Haykin. Com isso, foram selecionados apenas periódicos que utilizaram como métodos de *Machine Learning* redes bayesianas e de previsão de eventos probabilísticos. Além disso, a fim de entender os dados coletados, foram incluídos trabalhos referentes ao funcionamento e métodos de modelagem em sistemas elétricos de média e alta tensão.

A princípio, destacou-se no resultado de pesquisa os estudos de [Barth \(2023\)](#), que introduziu expressões matemáticas básicas de redes bayesianas, definindo-as como *Classes de Modelos Gráficos* cujo objetivo principal é apresentar em formato de grafos distribuições de probabilidades conjuntas, partindo do pressuposto das independências condicionais entre variáveis analisadas [Murphy \(2002\)](#), [Koller e Friedman \(2009\)](#), [Neapolitan \(2003\)](#). Em sua estrutura, os nós dos grafos representam *Variáveis Aleatórias* e a ausência ou presença de arestas representam relações estatísticas entre as variáveis [Neapolitan \(2003\)](#).

Além disso, trabalhos como os de [Machado et al. \(2021\)](#) demonstram a aplicabilidade de *Redes Bayesianas* aplicadas em grandes volumes de dados. Os avanços demonstrados possibilitaram o entendimento da aplicação dos algoritmos genéticos adaptado para aplicação de uma estrutura Bayesiana. Um dos enfoques mais promissores é o uso de Redes Bayesianas para o diagnóstico de falhas. Segundo [Kinjo et al. \(2022\)](#), as Redes Bayesianas oferecem uma estrutura robusta para lidar com incertezas, permitindo a modelagem de relações probabilísticas entre diferentes variáveis extraídas de um sistema *Internet of Things* direcionado para a área da saúde. Um dos enfoques mais promissores é o uso de Redes Bayesianas para o diagnóstico de falhas. Segundo [Kinjo et al. \(2022\)](#), as Redes Bayesianas oferecem uma estrutura robusta para lidar com incertezas, permitindo a modelagem de relações probabilísticas entre diferentes variáveis extraídas de um sistema *Internet of Things* direcionado para a área da saúde.

Ademais, o estudo de [Mendonça, Pereira e Aylmer \(2022\)](#) utilizou como parâmetro a duração equivalente de interrupção por unidade consumidora (DEC) e a frequência equivalente de interrupção por unidade consumidora (FEC) para a aplicação de análise de fronteira estocástica com a finalidade de determinar um *benchmarking* relacionado aos indicadores de continuidade de energia elétrica fornecida pelas concessionárias. Tal referência teve como função avaliar a eficiência do serviço de distribuição de energia,

considerando o panorama de expansão no mercado, processo que ocorre periodicamente.

Outrossim, este trabalho é corroborado pelas observações apresentadas por [Ferreira et al. \(2020\)](#), que utilizou a metodologia de *Machine Learning* Regressão com dados em painel com o propósito de estimar o impacto dos investimentos destinados à manutenção de distribuição de linhas aéreas pela concessionária *Light*. Foram utilizados como parâmetros de análises os históricos anuais de investimentos direcionados à manutenção nas linhas da concessionária, identificando ações que impactaram nos indicadores de qualidade adotados (DEC, FEC e compensações).

Com isso, através dos resultados apresentados pelos autores supracitados, foi possível identificar variáveis importantes que auxiliam na tomada de decisão do planejamento de manutenção das redes de distribuição elétrica. Ambos os autores apresentados utilizaram a linguagem *Python* e práticas de *Machine Learning* para realizar análises pertinentes em relação à base de dados fornecida pela ANEEL (Agência Nacional de Energia Elétrica). Observa-se que ambos os projetos abordados apresentam como objeto de análises de desempenho a frequência e o tempo de continuidade das interrupções elétricas, representadas pelos indicadores FEC e DEC, respectivamente.

Nas considerações finais referenciadas, os autores apresentam como resultados estimativas baseadas em *Machine Learning* aplicando princípios de redes neurais, levando em consideração testes que visam atestar a credibilidade do modelo bayesiano utilizado em uma base de dados. Em complemento aos trabalhos referidos, que utilizaram algoritmos de aprendizado de máquina para a análise de dados referentes à frequência e continuidade de distribuição elétrica, foram explorados também estudos sobre a aplicação de redes bayesianas em unidades geradoras de usinas elétricas e detecção de sistema *Internet of Things* (Internet das Coisas) na saúde. Esses trabalhos expressam, seguramente, suma importância para a aplicação de redes bayesianas em bases de dados. Porém, muitos desses são aplicados em segmentos distintos de dados, posto que se torna não pertinente referenciá-los neste estudo devido ao tema e à estrutura dos dados, que têm como foco as características e informações das redes de distribuição presentes em território nacional.

Nesse panorama, ressalta-se que a [Agência Nacional de Energia Elétrica \(ANEEL\) \(2024\)](#) fornece informações diversas referentes ao sistema de energia elétrica brasileiro, sendo importante definir as diferenças de segmentos de geração, transmissão e distribuição de energia elétrica. De acordo com a regulação, o sistema de distribuição é formado por subestações, redes elétricas e agrupamentos de instalações e equipamentos que operam em níveis de tensão inferiores a 230 kV, sendo pertencentes a concessionárias ou permissionárias de distribuição elétrica.

Embasando-se no presente Estado da Arte, este projeto consistiu em desenvolver a análise computacional da base de dados fornecida pela [Agência Nacional de Energia Elétrica \(ANEEL\) \(2024\)](#) referente às causas de interrupções em redes de distribuição e

à frequência e continuidade do fornecimento de energia elétrica, utilizando a linguagem Python através das bibliotecas Pgmpy, Pandas, Numpy e Matplotlib. Este projeto tem como diferencial a análise atualizada da eficiência das redes de distribuição elétrica, além da investigação das causas de interrupções elétricas, especialmente aquelas causadas por fenômenos naturais.

## 3 Fundamentação Teórica

Neste capítulo serão retratados conceitos referentes às interrupções em redes de distribuição, redes bayesianas, Técnicas Bayesianas, preparação de dados com foco em definições de limpeza de dados, seleção de variáveis e discretização, modelagem de redes bayesianas, estimação de parâmetros, implementação do modelo, inferência e métodos de avaliação do modelo proposto. Tais apresentações são fundamentais para a melhor compreensão do trabalho.

### 3.1 Interrupções em redes de distribuição elétrica

Segundo [Ferreira \(2018\)](#), entende-se redes de distribuição por conjunto de equipamentos que têm como finalidade levar a energia aos consumidores domésticos ou industriais, nos respectivos níveis de tensão: a baixa tensão na qual os aparelhos estarão ligados diretamente, a média tensão, que alimenta todos os postos de transformação, e a alta tensão, que tem como única função fornecer energia para as subestações. Ademais, essas redes podem receber a produção de energia independente, que utilizam como exemplo a energia solar, energia mini hídrica e centrais de cogeração.

Para [Camargo et al. \(1996\)](#), o valor econômico associado ao suprimento de energia elétrica está diretamente relacionado à qualidade do fornecimento, sendo observadas as seguintes características na rede:

- Disponibilidade: suprimento permanente de energia elétrica de acordo com a quantidade de demanda do usuário;
- Conformidade: abastecimento de energia limpa, ou seja, com níveis satisfatórios de flutuação de tensão e harmônicas;
- Restaurabilidade: Capacidade temporal do sistema de distribuição retornar o fornecimento de energia;
- Flexibilidade: Capacidade de adaptação do sistema em relação a mudanças planejadas ou acidentais.

Com isso, são adotados como indicadores de qualidade de fornecimento de energia por conjunto de unidades consumidoras pela [Agência Nacional de Energia Elétrica \(ANEEL\) \(2021\)](#) as seguintes variáveis:

- DEC: Está relacionado ao intervalo de tempo em que um conjunto de unidades consumidoras ficou sem o fornecimento de energia elétrica, sendo o tempo médio avaliado igual ou maior que três minutos.



- FEC: Está relacionado à quantidade de vezes que um conjunto de unidades consumidoras ficou sem o fornecimento de energia elétrica, considerando que se avaliam interrupções de três ou mais minutos.

Dado esses indicadores, este trabalho busca relacionar, por meio de métodos bayesianos, as causas das interrupções registradas, levando em consideração a localização de cada ocorrência e os motivos causadores de tal falha. Sob a visão de Mamede (2011), as ocorrências de falhas em redes de distribuição são atribuídas a fenômenos naturais, falhas humanas, operacionais, em equipamentos ou em proteção e medição. Dado esse cenário e definidas as variáveis que servem de embasamento para a construção de uma análise, seguem a partir do próximo subtópico definições importantes para a modelagem de *redes bayesianas*.

## 3.2 Redes Bayesianas

Conceitua-se *Redes Bayesianas* como uma classe de *Modelos Gráficos (GMs)*, que têm como objetivo representar em forma de grafos as distribuições probabilísticas, partindo do pressuposto da relação de *Independências Condicionais* entre as variáveis analisadas (MURPHY, 2002; NEAPOLITAN, 2003).

Nos dizeres de Charniak (1991), o raciocínio lógico tem como principal vantagem agentes que podem tomar decisões mesmo sem informações suficientes que provem que uma ação será pertinente. Posto isso, é necessária a definição de conceitos que envolvem a área de estatística e probabilidade, tendo como objetivo concluir o entendimento da teoria de Bayes, que se trata da base teórica deste trabalho. Como isso define-se, a partir do entendimento de Meyer (2000), os seguintes conceitos:

1. Espaço Amostral: se define como o conjunto de todos os resultados possíveis de um experimento aleatório;
2. Evento: É qualquer subconjunto pertencente ao espaço amostral;
3. Probabilidade: A partir de um experimento  $\in$  e  $S$  o espaço amostral de  $\in$ . A cada evento pertencente ao espaço amostral  $S$  associa-se um número real  $P(A)$  denominado como probabilidade  $A$  que satisfaça as seguintes condições:
  - a) A probabilidade  $P(A)$  deve se encontrar no intervalo de  $[0;1]$ ;
  - b) A probabilidade referente ao espaço amostral  $P(S)$  tem que ser igual a um, ou seja, o somatório da probabilidade de eventos presentes no espaço amostral  $S$  é igual a um;

- c) Dado um evento B, se A e B forem mutuamente exclusivos, temos que  $P(A \vee B) = P(A) + P(B)$ . Essa probabilidade é denominada de probabilidade incondicional, pois não depende de outro evento.

### 3.2.1 Definições de probabilidade condicional e independência condicional

Como o trabalho apresentado é baseado na relação probabilística de eventos, define-se que ao contrário da probabilidade incondicional, a probabilidade condicional é dependente de eventos anteriores. A probabilidade condicional de um espaço amostral B em relação a um espaço amostral A, colocada como  $P(B|A)$ , segundo Meyer (2000), pode ser conceituada como a probabilidade da ocorrência do evento B, dada a ocorrência de um evento A. Posto isso, o cálculo de  $P(B|A)$  retrata a probabilidade de B em relação a A.

Sob a perspectiva de Hazzan e Iezzi (2004), utiliza-se a definição de frequência relativa para a definição formal de probabilidade condicional. Dado um experimento repetido n vezes e seja nA, nB e  $nA \wedge nB$ , respectivamente o número de vezes que ocorreram os eventos A, B e  $A \wedge B$ . Sendo assim, pode-se afirmar que  $nA \wedge nB / nA$  representa a frequência relativa de B dada a ocorrência do evento A. Com isso, formula-se a seguinte expressão matemática Eq(3.1).

$$P(A | B) = \frac{P(A \wedge B)}{P(A)} \quad (3.1)$$

Essa formulação está condicionada à seguinte proposição,  $P(A)$ . Por conseguinte, existem duas maneiras de realizar o cálculo da probabilidade de  $P(B|A)$  (MEYER, 2000):

Diretamente, levando em consideração a probabilidade de B em relação ao espaço amostral de A reduzido.

Pode-se também aplicar a Eq(1), apresentada, onde  $P(A \wedge B)$  e  $P(A)$  são calculados em relação ao espaço amostral original. Observa-se como reflexo importante da probabilidade condicional o teorema da multiplicação (HAZZAN; IEZZI, 2004).

O Teorema da Multiplicação retrata que a probabilidade de dois eventos acontecerem simultaneamente é o produto da probabilidade do primeiro evento pela probabilidade do outro evento, dado o primeiro. Tal formulação é representada na equação Eq(3.2).

$$P(A | B) = P(BA).P(A) \quad (3.2)$$

Observa-se como aspecto importante da aplicação da teoria Bayesiana o conceito de independência de eventos. Ocorre quando um evento A não depende de um evento B, podendo ser apresentado na formulação Eq(3.3) a seguir.

$$P(A | B) = P(A)$$

(3.3)

Com isso temos que A independe de B, podendo-se concluir que B também independe de A. Tal afirmativa é proposta na equação Eq(3.4).

$$P(B | A) = \frac{P(A \wedge B)}{P(A)} = \frac{P(B) \cdot (P(A | B))}{P(A)} = \frac{P(B) \cdot P(A)}{P(A)} = P(B) \quad (3.4)$$

Por meio da aplicação do teorema da multiplicação apresentada na Eq (2), identifica-se a seguinte expressão Eq(3.5).

$$P(A \wedge B) = P(A) \cdot P(B | A) = P(A) \cdot P(B) \quad (3.5)$$

Com isso definimos a independência entre eventos a partir da expressão numérica Eq(3.6).

$$P(A \wedge B) = P(A) \cdot P(B) \quad (3.6)$$

Destaca-se também a definição de independência condicional entre eventos. Dado um evento X que é condicionalmente independente de um evento Y, posto um evento Z, se a distribuição probabilística que rege Z é independente de Y, dado o valor de Z, representa-se através da seguinte expressão Eq(3.7).

$$P(X | Y \wedge Z) = P(X | Z) \quad (3.7)$$

### 3.2.2 Definição do Teorema de Bayes

De acordo com [Bobbio et al. \(2001\)](#), as *Redes Bayesianas*, também conhecidas como *Redes Causais*, *Redes de Crenças* ou *Gráfico de Dependência Probabilística*, tiveram seu desenvolvimento na década de 1980, tendo como base o teorema de Bayes que descreve a relação de um evento dado a ocorrência de outro evento, sendo utilizada nas áreas de finanças, saúde, desenvolvimento de jogos, entre outros. Proposta uma partição de um espaço amostral S, um conjunto de eventos A1, A2, A3, An, sendo tais eventos mutuamente exclusivos e sua união igual ao espaço amostral S. Dado outro evento B sendo a probabilidade P(B) maior que zero, tem-se a seguinte expressão matemática Eq(3.8).

$$B = S \wedge B = (A \vee A2 \vee \dots \vee An) \wedge B \quad (3.8)$$

Sendo A e B mutuamente exclusivos, temos que, por consequência, a probabilidade da ocorrência de B pode ser representada na Eq(3.9).

$$P(B) = P(A_1 \wedge B) + \dots + P(A_n \wedge B) = \sum_i (A_i \wedge B) \quad (3.9)$$

Aplicando o Teorema da Multiplicação apresentado na equação Eq(2), obtém-se a expressão Eq(9).

$$P(B) = \sum_i (A_i \wedge B) = \sum_i (B | A_i).P(A_i) \quad (3.10)$$

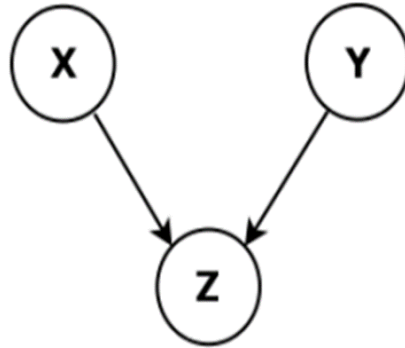
Usando de artifícios para a ordem  $P(A_i|B)$ , a resolução obtém o Teorema de Bayes (PAULINO; TURKMAN; MURTEIRA, 2003), apresentado na Eq(3.11).

$$P(A_i | B) = \frac{(P(B | A_i).P(A_i))}{P(B)} = \frac{(P(B | A_i).P(A_i))}{\sum_i (B | A_i).P(A_i)} \quad (3.11)$$

### 3.2.3 Modelagem da Rede

A análise das *redes bayesianas* está diretamente relacionada à teoria dos grafos, partindo do pressuposto de que as *redes bayesianas* são formadas por análises lógicas de *Grafo Acíclico Dirigido* (*Directed Acyclic Graph – DAG*), no qual os vértices são as representações dos nós e as arestas são as representações de dependências entre os nós interligados, sendo os nós independentes denominados nós pais. Como pode ser observado na Figura 1.

Figura 1 – Directed Acyclic Graph - DAG. Fonte: Elaborada pelo Autor através do site Miror.



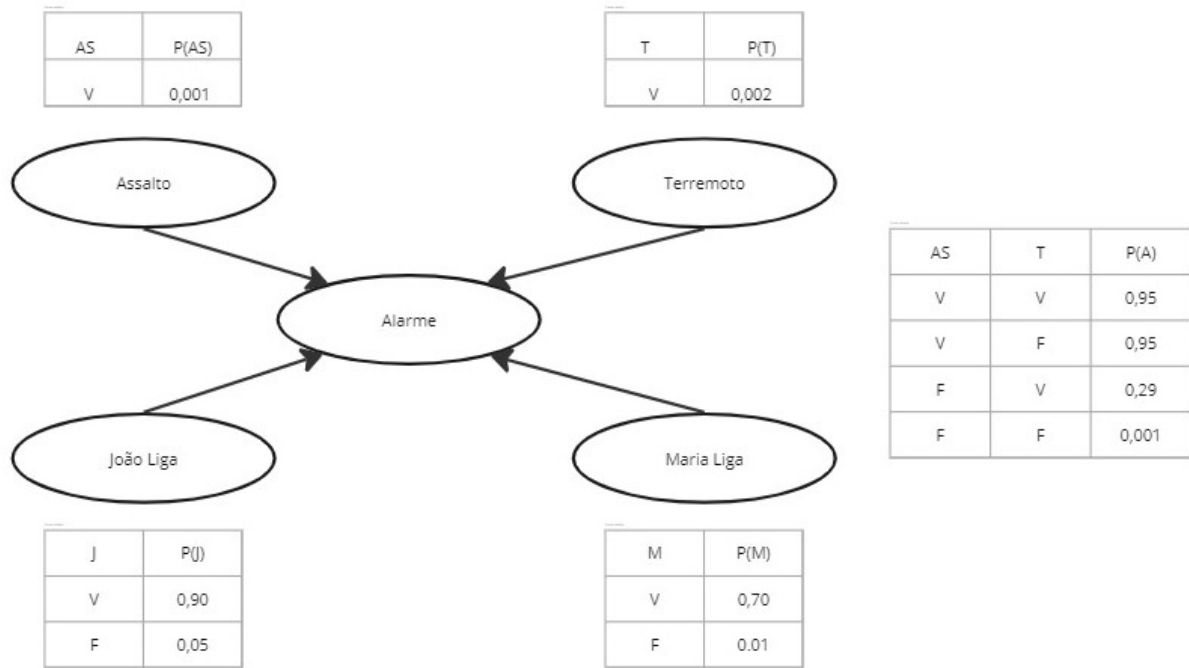
Além de se basear na Teoria dos Grafos, de acordo com Neapolitan (2003), as *redes bayesianas* são representações de domínio caso a condição de Markov seja satisfeita, posto isso a condição de Markov afirma que as variáveis não descendentes não fornecem informações adicionais à variável analisada. Segundo Marques e Dutra (2002), sendo definida a tipologia da rede, posteriormente deve-se determinar os valores probabilísticos referentes aos nós que participam a partir de dependências diretas e através dessas, computar as probabilidades das variáveis desejadas. Observa-se o exemplo explicativo extraído de (RUSSELL; NORVIG, 1995).

É representado na Figura 2 a seguinte situação hipotética, que serve para representação de uma modelagem para *redes bayesianas*:

- I. Considera-se a instalação de um novo alarme destinado a assaltos;

- II. O alarme pode ser acionado em caso de terremoto;
- III. Dois vizinhos, João e Maria, se disponibilizaram a telefonar em caso de acionamento do alarme;
- IV. João liga quando ouve o alarme, porém algumas vezes ele confunde com o telefone e também liga em caso de um telefonema;
- V. Já Maria ouve música com volume alto, posto isso, às vezes não ouve o alarme, deixando de ligar nesses casos. Observa-se como variáveis os eventos Assalto, Terremoto, Alarme, JoãoLiga (ligação do João, dadas as condições mencionadas) e MariaLiga (Ligação da Maria).

Figura 2 – Modelo de uma Rede Bayesiana. Fonte: Elaborada pelo Autor através do site Miror.



Conhecendo as probabilidades relativas aos eventos destacados no problema, caso a *rede bayesiana* satisfaça a condição de Markov, implica-se que sua distribuição de probabilidade conjunta é equivalente ao produto das probabilidades condicionais de todos os nós considerados os valores dos nós pais. Com isso, obtém-se a seguinte expressão numérica Eq(3.12).

$$P(X_1, X_2 \dots X_n) = \prod_{i=1}^n P(X_i | pa(X_i)) \quad (3.12)$$

Com isso, o exemplo proposto na Figura 2 resulta na seguinte expressão Eq(3.13).

$$P(A \wedge \neg AS \wedge \neg T \wedge J \wedge M) = P(J | A) \cdot P(M | A) \cdot P(A | \neg AS \wedge \neg T) \cdot P(\neg AS) \cdot P(\neg T) \quad (3.13)$$

### 3.2.4 Inferência

O processo de inferência bayesiana se trata do método de obtenção de probabilidades a posteriori a partir da relação com os valores probabilísticos a priori (NEAPOLITAN, 2003). Posto isso, atribui-se à inferência em *redes bayesianas* como o processo de obtenção de estimativas de probabilidade de eventos que associados aos dados, a partir do conhecimento de uma nova informação ou evidência.

## 3.3 Algoritmos de Aprendizagem de Estrutura

Define-se como aprendizagem estrutural, o processo necessário para realiza a estimativa de relações presentes entre o nós que compõem a estrutura DAG de uma rede Bayesianas. De acordo com Contaldi (2016), é utilizado três abordagens principais para resolução do aprendizado de uma estrutura:

1. **Constraint-based:** Os algoritmos que se baseiam nesse modelo, é fundamentado em teste estatísticos que tem como objetivo determinar independências condicionais entres os nós, contruindo a rede Bayesiana de acordo com essas relações.
2. **Search and Score (SS):** Utiliza-se da métrica de pontuação para pontuar a qualidade métrica de uma Rede Bayesiana, avalindo de acordo com a complexidade estrutural e com a verossimilhança com os dados.
3. **Híbrida:** Combina as abordagens *Constraint-based* e *Search and Score*, utilizando testes de independência para restringir o espaço de busca e, em seguida, aplicando um critério de pontuação para refinar a estrutura da rede.

## 4 Materiais e Métodos

No presente capítulo, são expostas as classificações das pesquisas, bem como as ferramentas e tecnologias empregadas no desenvolvimento do trabalho.

### 4.1 Classificação de pesquisa

A princípio, classifica-se a natureza de pesquisa aplicada. De acordo com [Santos e Nascimento \(2021\)](#), uma pesquisa aplicada tem como objetivo principal a geração de um produto, processo ou novas tecnologias de aplicação imediata, tendo como finalidade a resolução de entraves peculiares e concretos que se relacionam com interesses locais. Outrossim, os procedimentos adotados nessa pesquisa são tipificados como uma pesquisa-ação, tendo em vista o propósito de entender ou resolver o fenômeno de interrupções elétricas através de ações diretas.

Para [Gil \(2008\)](#), uma pesquisa exploratória tem como objetivo principal desenvolver, esclarecer e refinar conceitos e ideias, com o propósito de formular problemas mais precisos ou hipóteses passíveis de investigação em estudos futuros. Entre os diferentes tipos de pesquisa, a exploratória é a que apresenta menor rigidez em seu planejamento. Procedimentos de amostragem e técnicas quantitativas de coleta de dados foram aplicados a fim de determinar um modelo probabilístico aplicado em diagnósticos de interrupções elétricas. Com isso, embute-se a este projeto o caráter exploratório, tendo em vista o uso de métodos já existentes em uma base de dados.

### 4.2 Ferramentas e tecnologias

#### 4.2.1 *Python*

A linguagem de programação *open source Python*, criada em 1991, se trata de uma ferramenta de grande capacidade no desenvolvimento de projetos relacionados à análise de dados. Tal mecanismo evoluiu desde a data de sua criação e atualmente se encontra na versão 3.12.4 ([Python Software Foundation, 2024](#)). O desempenho eficiente da linguagem em análises de dados e modelagem bayesiana se deve ao vasto conteúdo de bibliotecas, capazes de auxiliar no desenvolvimento da pesquisa proposta tanto nas etapas de coleta e pré-processamento de dados quanto na implementação de modelagem de *redes bayesianas* aplicadas na base de dados coletada.

Figura 3 – Logo *Python*. Fonte: ([Python Software Foundation, 2024](#))

#### 4.2.2 Jupyter Notebook

O ambiente de desenvolvimento escolhido para aplicação do projeto proposto foi o Jupyter Notebook, que permite a execução de código e textos de forma interativa. Sua interface flexível possibilita aos usuários configurações e organizações distintas aos fluxos de trabalho relacionados a ciência de dados e aprendizado de máquina ([Project Jupyter, 2024](#)). Posto isso, justifica-se tal escolha devido à compatibilidade com a linguagem *Python* e outras ferramentas de *big data*. O logo do ambiente de desenvolvimento é apresentado na Figura 5.

Figura 4 – Logo Jupyter Notebook. Fonte: ([Project Jupyter, 2024](#))

### 4.3 Coleta e Pré-Processamento dos Dados

Neste subcapítulo, serão apresentadas as tecnologias que foram utilizadas em todo o processo relativo ao *data base*, de limpeza e transformação dos dados. As bibliotecas citadas a seguir foram selecionadas por fornecerem uma documentação abrangente, bem estruturada e já consolidada na área de *data science*, além de atenderem de forma eficiente aos objetivos deste trabalho.

#### 4.3.1 Variáveis Consideradas

A princípio, os dados serão coletados a partir das bases fornecidas pela ANEEL, observando informações sobre interrupções em redes de distribuição elétrica, com foco nos indicadores DEC e FEC, além de causadores de interrupções elétricas. Posto isso, foram destacadas as seguintes informações sobre a base de dados: Interrupções de Energia Elétrica nas Redes de Distribuição, coletada na ([Agência Nacional de Energia Elétrica \(ANEEL\), 2024](#)). Acesso em: 19 ago. 2024. Pode-se analisar as variáveis da base de dados consideradas para o desenvolvimento da estrutura do modelo Bayesiano na Tabela 1.



Tabela 1 – Propriedades da Base de Dados

Nome do Campo	Descrição
DatGeracaoConjuntoDados	Data do processamento de carga automática no momento da geração para publicação do conjunto de dados abertos.
IdeConjuntoUnidadeConsumidora	Código do conjunto da unidade consumidora.
DscAlimentadorSubestacao	Descrição do alimentador da subestação (Linha de transmissão que sai de uma subestação de distribuição e segue para a distribuição de energia).
DscSubestacaoDistribuicao	Descrição da subestação de distribuição (Instalação elétrica de alta potência, contendo equipamentos para transmissão e distribuição de energia elétrica, além de equipamentos de proteção e controle).
NumOrdemInterrupcao	Descrição da Ordem de Interrupção (Ofício ou normativo determinando interrupções de energia).
DscTipoInterrupcao	Tipo de interrupção (não programada ou programada).
IdeMotivoExpurgo	Identificador do motivo do expurgo, com valores variando de 0 a 8, representando diferentes causas.
DatInicioInterrupcao	Data e hora do início da interrupção.
DatFimInterrupcao	Data e hora do reestabelecimento (fim) da interrupção.
FatGeradorInterrupcao	Descrição do fato gerador da interrupção.
NumNivelTensao	Nível de tensão onde o fato gerador foi verificado.
NumUnidadeConsumidora	Número de unidades consumidoras atingidas.
NumConsumidorConjunto	Número de consumidores do conjunto.
NumAno	Ano de início da interrupção.
NomAgenteRegulado	Nome do agente do setor regulado.
SigAgente	Sigla da concessionária de energia elétrica.
NumCPFCNPJ	Número do CNPJ ou CPF do agente fiscalizado.

Fonte: Adaptado da AGÊNCIA NACIONAL DE ENERGIA ELÉTRICA (ANEEL).

### 4.3.2 Biblioteca *NumPy*

Destaca-se entre as bibliotecas que serão utilizadas no desenvolvimento do projeto a *NumPy*, que oferece suporte abrangente em operações rápidas e eficientes que envolvem *arrays* e matrizes multidimensionais, fatores determinantes em análise de dados. Tal ferramenta também possui uma ampla variedade de funções algébricas que oferecem resoluções matemáticas e estatísticas, permitindo cálculos complexos de forma otimizada. Além dessas funcionalidades, *NumPy* funciona como base de bibliotecas importantes no desenvolvimento de *data science*, como, por exemplo, a biblioteca *Pandas*, atuando na facilitação de manipulação de dados (NumPy, 2024). Pode-se visualizar o logo dessa ferramenta na Figura 6.

### 4.3.3 Biblioteca *Pandas*

Enfatiza-se também a utilização da biblioteca *Pandas*, uma ferramenta crucial em desenvolvimentos de *data science*, por oferecer importantes artifícios para manipulação e

Figura 5 – Logo *NumPy*. Fonte: (NumPy, 2024)

visualização de dados (Pandas Development Team, 2024). A escolha dessa biblioteca se deu pela necessidade de limpeza e manipulação dos dados obtidos. Observa-se o logo da biblioteca supracitada na Figura 7.

Figura 6 – Logo *Pandas*. Fonte: (Pandas Development Team, 2024)

## 4.4 Aplicação do Modelo Bayesiano

### 4.4.1 Modelagem da Rede Bayesiana

A modelagem da *Rede Bayesiana* consiste na definição de sua estrutura probabilística, permitindo a representação das dependências condicionais entre as variáveis do sistema, definindo a estimativa de relação entre as variáveis selecionadas. Para este trabalho, optou-se pela abordagem *MMHC* (*Max-Min Hill Climbing*), um método híbrido que combina técnicas baseadas em restrições e pontuação para a aprendizagem da estrutura da rede. Tal escolha se deu pela eficiência computacional e a capacidade de balancear precisão e escalabilidade demonstradas nos trabalhos de (TSAMARDINOS; BROWN; ALIFERIS, 2006). Ademais foi observado por Machado et al. (2021) a superioridade dessa abordagem em relação aos demais algoritmos de estruturação.

Destaca-se que por se tratar de uma abordagem híbrida o modelo *MMHC* (*Max-Min Hill Climbing*) é composto por fases de restrição e pontuação:

1. **Fase Constraint-based:** Determina independências condicionais entre as variáveis por meio do algoritmo *Max-Min Parents and Children* (*MMPC*), reduzindo o espaço de busca para a construção da rede.

2. **Fase Search and Score:** Aplica o algoritmo *Hill Climbing (HC)* para explorar a estrutura da rede, avaliando diferentes configurações com uma métrica de pontuação, como o *Bayesian Information Criterion (BIC)*, que calcula a versemelhança subtraído pela complexidade, aplicando penalidade à essa complexidade. Observa-se a modulação matemática na métrica BIC na equação eq(4.1) abaixo.

$$BIC(R | D) = \sum_{i=1}^n \sum_{j=1}^{q_i} \sum_{k=1}^{r_i} N_{ijk} \log \left( \frac{N_{ijk}}{N_{ij}} \right) - \frac{1}{2} \log(N) \sum_{i=1}^n (r_i - 1) q_i \quad (4.1)$$

- $n$ : Número de nós.
- $r_i$ : Número de estados do  $i$ -ésimo nó.
- $q_i$ : Número de pais do  $i$ -ésimo nó, onde  $q_i = 1$  se o  $i$ -ésimo nó não possuir pais.
- $N$ : Número de observações na amostra  $D$ .
- $N_{ijk}$ : Número total de observações na amostra em que o  $i$ -ésimo nó assume seu  $k$ -ésimo estado e seus pais assumem o  $j$ -ésimo estado.
- $N_{ij}$ : Quantidade de observações do  $i$ -ésimo nó quando seus pais assumem o  $j$ -ésimo estado.

## 4.5 Implementação

A seguir, são apresentadas as principais bibliotecas utilizadas no desenvolvimento da estrutura bayesiana deste trabalho.

### 4.5.1 Modelagem Bayesiana

A biblioteca **pgmpy** foi utilizada para a modelagem e inferência em *Redes Bayesianas*. De acordo com [pgmpy Development Team \(2024\)](#), a biblioteca *pgmpy* se trata de uma aplicação pura para modularidade e extensibilidade de *redes bayesianas*, possibilitando o desenvolvimento de inferências causais. Com isso, a utilização dessa ferramenta se torna fundamental em uma aplicação de *data science* voltada para redes probabilísticas, como, por exemplo, *redes bayesianas* e de *Markov*, oferecendo uma base concreta de recursos para modelagem de incertezas, auxiliando na implementação de inferências e simulações, viabilizando sua utilização em projetos relacionados a diagnósticos probabilísticos. É apresentada na Figura 8 a logo da biblioteca mencionada.

### 4.5.2 Manipulação de Grafos

#### 4.5.2.1 Biblioteca *NetworkX*

O pacote *NetworkX* tem como finalidade a manipulação e estudo de estruturas, dinâmica e funções de redes complexas, sendo utilizado para geração de gráficos clássicos,

Figura 7 – Logo *pgmpy*. Fonte: ([pgmpy Development Team, 2024](#))

gráficos aleatórios e redes sintéticas ([NetworkX Development Team, 2024](#)). Essa ferramenta foi selecionada por ser integrada com a biblioteca *pgmpy* e pelo objetivo principal de aplicações em redes complexas de multigrafos. A visualização da logo dessa biblioteca está apresentada na Figura 9.

Figura 8 – Logo *NetworkX*. Fonte: ([NetworkX Development Team, 2024](#))

#### 4.5.2.2 Biblioteca *Matplotlib*

Destaca-se também a biblioteca *Matplotlib*, que, segundo [Matplotlib Development Team \(2024\)](#), é responsável pela visualização de dados, com a função de criação de gráficos variados. Com isso, tal biblioteca foi selecionada por oferecer facilidades na formulação de gráficos, histogramas e demais dispositivos de apresentação de dados estatísticos. Observa-se na Figura 10 a apresentação da logo da biblioteca *Matplotlib*.

Figura 9 – Logo *Matplotlib*. Fonte: ([Matplotlib Development Team, 2024](#))

### 4.5.3 Inferência por Eliminação de Variáveis

Para esse trabalho, foi adotado o método de inferência por eliminação de variáveis, procedimento de caráter exato, favorecendo a precisão e sendo utilizada para redes com números menores de variáveis, assim como proposto por (TRONCHONI, 2008). O funcionamento da inferência por eliminação de variáveis, de acordo com Murphy (2002), se trata de um método versátil para projetar inferência em *redes bayesianas*, entregando uma forma de lidar com a complexidade de realizar cálculos de distribuições marginais.

## 5 Desenvolvimento

Nesta seção, será apresentado o processo detalhado de desenvolvimento do modelo, incluindo as etapas de pré-processamento dos dados, estruturação da Rede Bayesiana, os algoritmos utilizados para aprendizado estrutural e os métodos de inferência aplicados para a previsão de interrupções elétricas decorrentes de falhas nas redes de distribuições.

### 5.1 Pré-processamento e Tratamento dos Dados

O pré-processamento dos dados é uma etapa de suma importância para o ramo da ciência de dados, tem como finalidade preparar os dados brutos para aplicações de métodos de aprendizado de máquina. É frequente que no processo de coleta de dados, muitas informações são registradas de maneira inconsistente, como valores ausentes, valores incondizentes com a realidade e formatos inadequados para a aplicação de algoritmos de machine learning. Posto isso o objetivo do pré-processamento é transformar esses dados em um formato estruturado e de qualidade, otimizando a precisão dos modelos e reduzindo possíveis de erros. Nesse sentido foram realizadas as seguintes etapas através da manipulação de dados.

Em síntese, os dados da base de interrupções de energia elétrica abrangem o período de 1º de janeiro de 2021 até 31 de dezembro de 2024, cobrindo um histórico de quatro anos de ocorrências registradas. Para realizar a importação dos arquivos CSV correspondentes a cada ano, foi utilizado o método *pd.read\_csv()*, uma função específica da biblioteca *Pandas*, com a especificação do *encoding* como "ISO-8859-1", a fim de lidar com caracteres especiais, e o delimitador ";" para garantir a correta separação das colunas. Essa abordagem permite carregar e estruturar os dados de forma padronizada, possibilitando a concatenação dos arquivos referentes ao período mencionado para posterior análise, processamento e aplicação da *estrutura Bayesiana*.

#### 5.1.1 Seleção de variáveis e dados

A priori realizou-se a seleção de dados, destacando as variáveis, permitindo a extração de informações relevantes de maneira eficiente pertinentes para a análise de interrupções elétricas. Esse processo foi feito através da seleção de colunas, atribuindo as colunas pertinentes a uma nova data frame. Posteriormente a esse processo, excluiu-se todos os valores nulos e duplicados da base de dados, considerando esses valores irrelevantes para a aplicação da estrutura Bayesiana. Junto a esse processo, observou-se a necessidade ampliar as informações através das colunas já existentes, com isso através das colunas

temporais, foi acenado a base de dados informações de duração da interrupção, o mês e turno em que a interrupção ocorreu.

### 5.1.2 Discretização de Dados

Destaca-se também a a realização da discretização, um processo fundamental que visa transformar variáveis contínuas em variáveis categóricas ou discretas. Nesse contexto, foram realizadas discretizações no intuito de simplificar e categorizar as informações referentes às colunas, de duração da interrupção e níveis de tensão, observando faixas de valores.

Para a discretização da duração das interrupções e dos níveis de tensão observa-se que a faixa de tempo em que uma rede permanecia interrompida variava de 0,01 a 3201,81 horas e o nível de tensão de 9 a 230000 volts. Com intuito de categorizar esses valores a coluna de interrupções submeteu-se as seguintes etapas:

1. Conversão dos valores da coluna para o tipo numérico;
2. Categorização dos valores numérico, agrupando e atribuindo característica a uma faixa de valor;
3. Atribuição da coluna ao dataframe;
4. Conversão das categorias para o tipo category.

Através da desses processos a categorização desses valores estão representadas nas Tabela 2 e Tabela 3, a seguir.

Tabela 2 – Classificação da Duração das Interrupções

Faixa (em horas)	Categoria
< 0.05	Ruído
0.05 - 0.25	Interrupção muito pequena
0.25 - 0.5	Interrupção pequena
0.5 - 3	Média
3 - 6	Grande em área urbana
6 - 10	Grande em área rural
10 - 24	Preocupante
$\geq 24$	Anomalia

Fonte: Dados da análise de interrupções.

### 5.1.3 Reconfiguração das colunas

Posteriormente às etapas supracitadas, as colunas do data frame foram modificadas, a fim de se adequar aos formatos permitidos pelas bibliotecas pmgy e Networx, além

Tabela 3 – Classificação do Nível de Tensão

Faixa (em Volts)	Categoria
$\leq 1.000$	Baixa Tensão
1.000 - 36.000	Média Tensão
$> 36.000$	Alta Tensão

Fonte: Dados da análise de níveis de tensão.

disso foram renomeadas para facilitar a plotagem dos grafos. Posto isso, as variáveis estão configuradas de acordo com a Tabela 4.

Tabela 4 – Descrição das Colunas e Correspondência com os Nós

Descrição da Coluna	Letra do Nó	Valores da Coluna	Tipo dos Valores
Tipo de Interrupção	A	Programada, Não Programada	Categoria
Motivo da Interrupção	B	Ideneticador do Motivo da interrupção	Numérico (Categoria)
Nível de Tensão	C	Baixa, Média, Alta	Categoria
Sigla do Agente Regulador	D	Siglas das Concessionárias	Categoria
Duração da Interrupção	E	Gravidade da interrupção	Float (horas)
Turno da Interrupção	F	Manhã, Tarde, Noite	Categoria
Mês da Interrupção	G	Meses	Categoria

Fonte: Dados do data frame analisado.

## 5.2 Estruturação da Rede Bayesiana

Como a finalidade de representar as relações de dependências entre as variáveis representadas no dataframe, estrutura-se uma RB através das seguintes etapas.

- Definição das variáveis e estados possíveis;
- Construção de um grafo direcionado acíclico (DAG) para modelar as relações entre as variáveis;
- Definição de variáveis dependentes;
- Aplicação do modelo de restrição, em que é mantido apenas as variáveis fortemente dependentes;
- Utilização de métricas de qualidade do modelo, como o *Bayesian Information Criterion* (BIC).



### 5.2.1 Algoritmos de Aprendizado Estrutural

Para aprender a estrutura da Rede Bayesiana a partir dos dados históricos de interrupções, foram utilizados os seguintes algoritmos:

- **Hill Climbing:** Algoritmo heurístico que busca a melhor estrutura da rede através da maximização de uma função de pontuação (BIC);
- **(Max-Min Parents and Children - MMPC):** Método que avalia diferentes estruturas de redes utilizando funções de restrições avaliando o grau de dependências dos nós pais e filhos.

Com intuito de reduzir custos computacionais, aplicou-se o processo de algoritmo de aprendizado de estrutura por pontuação no base de dados completa. Para aplicação do algoritmo de aprendizado de estrutura Bayesiana são utilizou-se o método *Hill-Climbing Search* para buscar um rede de estruturas e a métrica BIC, para avaliação da rede encontrada, para evitar uma rede de alta densidade restringiu-se cada nó a no máximo três pais e o máximo de ce interações heurísticas.

Logo após, foi utilizado o método híbrido para base de dados, omitindo as variáveis de sigla da concessionária e turno. A princípio utilizou-se a algoritmo *MMPC* para encontrar uma estrutura inicial através da redução de relações irrelevantes entre os nós e posteriormente aplicado o algoritmo *Hill-Climbing Search* para refinar a estrutura encontrada.

### 5.2.2 Implementação dos Métodos de Inferência

Após a estruturação da Rede Bayesiana, foi aplicado método de inferência probabilística para prever a ocorrência de interrupções. O métodos utilizado foi:

- **Inferência por Eliminação de Variáveis:** algoritmo de inferência exata para calcular probabilidades condicionais.

Com isso, é possível realizar consultas probabilísticas relacionadas ao modelo proposto. Para a visualização das *CPDs* (*Conditional Probability Distributions*), utilizam-se estruturas de repetição em conjunto com a manipulação de *DataFrames*, permitindo representar as distribuições probabilísticas das variáveis dependentes. Essas distribuições possibilitam a análise detalhada do comportamento do sistema, identificando padrões e tendências a partir dos dados disponíveis. Com essas informações, torna-se viável realizar previsões de interrupções e diagnósticos, considerando as relações entre os nós estruturados pelas *Redes Bayesianas*.

### 5.3 Implementação gráfica de Grafos e resultados

Para uma melhor representação dos resultados obtidos nos processos anteriores, optou-se pela visualização dos DAGs (Directed Acyclic Graphs), utilizando as bibliotecas Matplotlib e NetworkX. A Matplotlib foi empregada para a visualização gráfica da estrutura da rede, enquanto a NetworkX foi utilizada para converter o modelo Bayesiano, obtido por meio da biblioteca pgmpy, em um grafo direcionado, representando seus nós e arestas de maneira estruturada e intuitiva.

Com o objetivo de representar distribuições probabilísticas, utilizou-se o Gráfico de Barras, uma vez que ele permite a categorização de valores discretos e assume um conjunto finito de possibilidades. Considerando que os valores numéricos de uma probabilidade variam entre 0 e 1, e que os nós da Rede Bayesiana podem ser tratados como categorias, esse tipo de gráfico se mostra uma escolha eficiente para a visualização dos resultados, facilitando a interpretação das probabilidades associadas a cada variável do modelo.

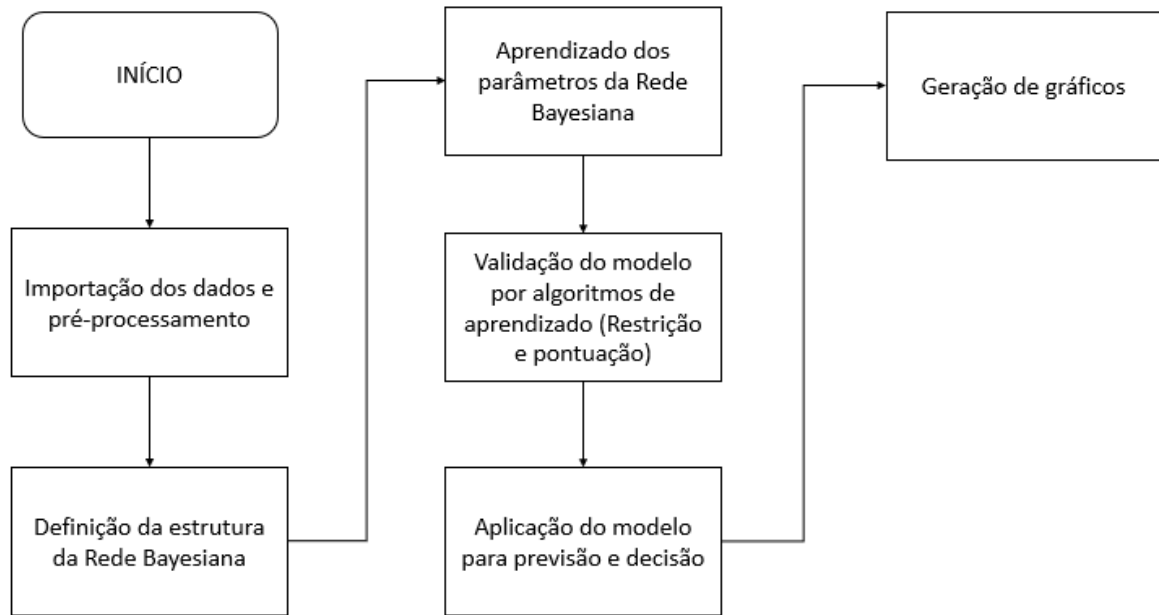
Outra representação gráfica utilizada foi o Boxplot, que exibe medidas estatísticas como a mediana, quartis e valores extremos. Esse tipo de abordagem é especialmente útil para identificar anomalias e padrões de dispersão nos dados. No contexto das interrupções de energia elétrica, o Boxplot permite detectar interrupções incomuns em diferentes concessionárias, possibilitando uma análise mais precisa das variações e tendências no fornecimento de energia.

Além disso, utilizou-se a Função de Distribuição Acumulada (Cumulative Distribution Function - CDF), que permite a visualização das probabilidades acumuladas, proporcionando uma análise mais detalhada sobre a ocorrência de eventos. No contexto das interrupções de energia, a CDF possibilita estimar a probabilidade de uma interrupção durar menos que um determinado tempo, facilitando a compreensão do comportamento dos tempos de falha e auxiliando na tomada de decisões estratégicas.

Em síntese, os resultados serão apresentados por meio de gráficos e tabelas interativas, desenvolvidos com o uso das bibliotecas Matplotlib e NetworkX, permitindo uma visualização clara e estruturada das análises realizadas

Os procedimentos mencionados nessa seção estão representados no fluxograma apresentado na Figura 10.

Figura 10 – Fluxograma de Desenvolvimento. Fonte: Autoria própria.



Os resultados obtidos serão discutidos na próxima seção, analisando o desempenho da Rede Bayesiana na previsão e no diagnósticos de eventos que influenciam a ocorrência de interrupções em redes de distribuição.

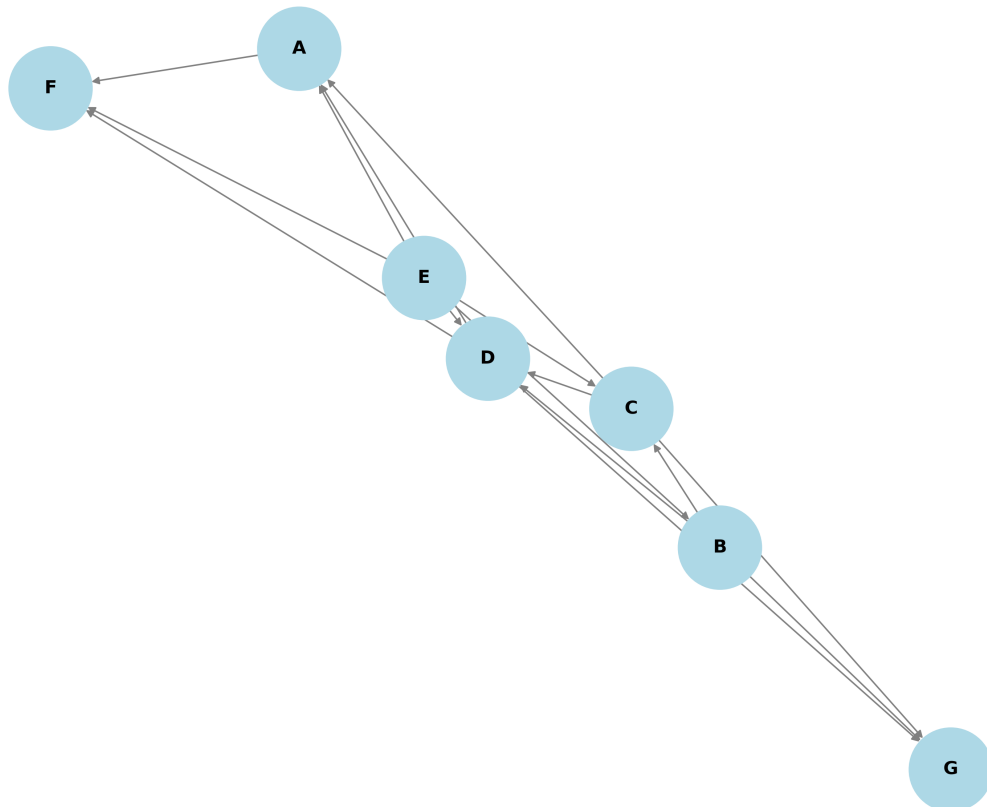
## 6 Análise dos Resultados

Neste capítulo, serão detalhados os resultados obtidos após a etapa de desenvolvimento do modelo para análise e diagnóstico de interrupções elétricas na rede de distribuição com objetivo principal desta análise é avaliar a eficácia do modelo implementado, compreender as principais causas das falhas e identificar padrões que possam auxiliar na mitigação de futuras interrupções, visando preservar o funcionamento da rede de distribuição elétrica.

### 6.1 Estruturas Apredidas

Os dados analisados foram extraídos de um conjunto representativo de registros históricos de falhas, abrangendo informações como concessionária responsável, tipo de interrupção, duração, turno e mês da interrupção, motivo da interrupção e nível de tensão da distribuição. A partir desses dados, foi construída *RBs* através do método *Hill Climbing*, como demonstrado na Figura 11, as letras representam os nós referenciados na Tabela 4 e no Anexo A.

Figura 11 – Estrutura Bayesiana *Hill Climbing*. Fonte: Autoria própria gerada pela biblioteca *Matplotlib*.



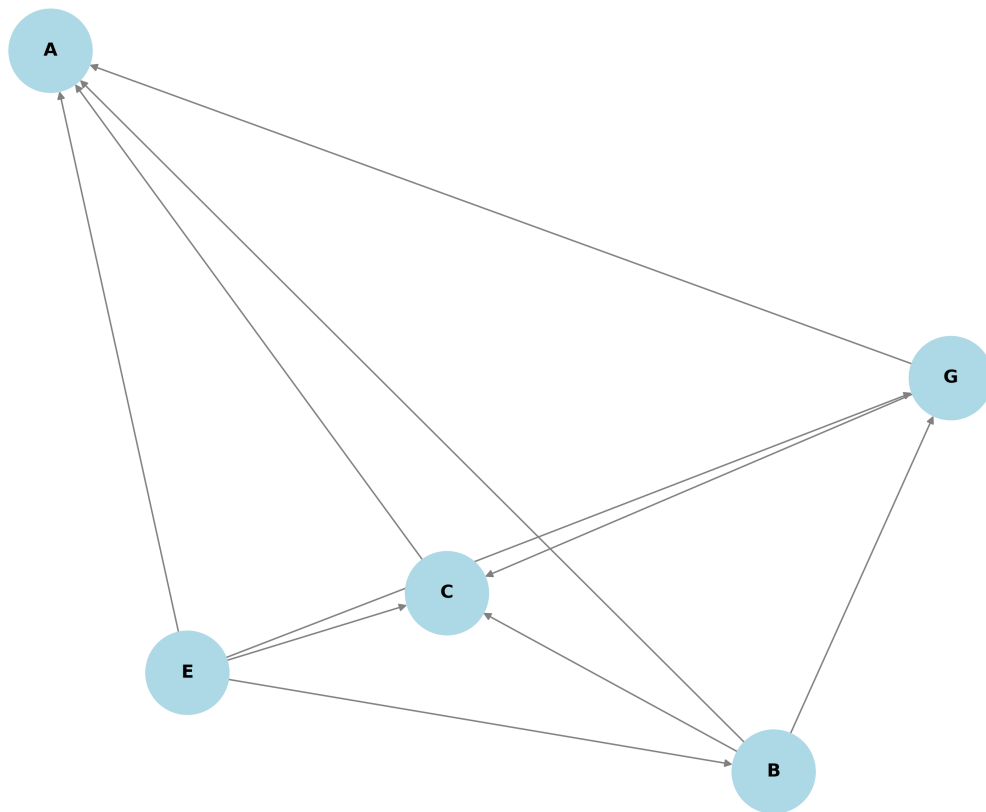
Através do modelo proposto após a aplicação do algoritmo de aprendizado *Hill Climbing*, é possível fazer as seguintes interpretações:

- 1º. ( $A \rightarrow F$ ): O tipo de interrupção, A, influencia o fator F, indicando que determinadas interrupções podem estar associadas a fatores específicos;
- 2º. ( $B \rightarrow D$ ): O nó B, referente ao motivo da interrupção, influencia a ocorrência do nó D, que representa a sigla da concessionária de energia elétrica. Isso ressalta que os tipos de falhas estão diretamente ligados a concessionárias específicas;
- 3º. ( $B \rightarrow C$ ): O nó B influencia a ocorrência de C, que representa o nível de tensão da rede, indicando que certos tipos de falhas são mais comuns em determinados níveis de tensão. Assim, falhas diferentes ocorrem em diferentes níveis de tensão;
- 4º. ( $B \rightarrow G$ ): O motivo da interrupção, B, varia ao longo dos meses representados por G. Pode-se interpretar que o tipo de interrupção está diretamente condicionado ao mês de ocorrência da interrupção;
- 5º. ( $D \rightarrow A$ ): A concessionária de energia, D, influencia diretamente o tipo de interrupção, A. Isso sugere que determinadas concessionárias estão mais associadas a interrupções programadas ou não programadas;
- 6º. ( $D \rightarrow G$ ): A concessionária, D, apresenta variações sazonais nas interrupções. Infere-se que algumas concessionárias têm falhas mais frequentes em determinados períodos do ano;
- 7º. ( $D \rightarrow F$ ): A concessionária, D, também afeta o fator F, sugerindo que agentes específicos podem estar associados a diferentes padrões de falha ou recuperação;
- 8º. ( $C \rightarrow D$ ): O nível de tensão, C, afeta a concessionária responsável, D. Isso implica que determinadas concessionárias operam predominantemente em baixa, média ou alta tensão;
- 9º. ( $C \rightarrow A$ ): O nível de tensão, C, influencia o tipo de interrupção, A. Redes de alta tensão podem apresentar falhas distintas das de baixa tensão;
- 10º. ( $C \rightarrow G$ ): Assim como as concessionárias, os níveis de tensão, C, também apresentam variações sazonais nas falhas do sistema. Isso indica que diferentes níveis de tensão estão mais suscetíveis a interrupções em determinadas épocas do ano;
- 11º. ( $E \rightarrow D$ ): A duração da interrupção, E, depende da concessionária responsável, D. Algumas concessionárias podem ter tempos de resposta mais rápidos para resolver interrupções;

- 12º.** ( $E \rightarrow F$ ): A duração da interrupção, E, também afeta o fator F, indicando que diferentes fatores podem estar associados a tempos de recuperação distintos;
- 13º.** ( $E \rightarrow C$ ): A duração da interrupção, E, é afetada pelo nível de tensão, C. Isso confirma a relação entre o tempo de reparo e os sistemas de baixa, média e alta tensão;
- 14º.** ( $E \rightarrow B$ ): O motivo da interrupção, B, influencia sua duração, E. Isso implica que a duração está diretamente relacionada ao motivo da interrupção;
- 15º.** ( $E \rightarrow A$ ): A duração da interrupção, E, depende do tipo de interrupção, A. Observa-se que a duração está relacionada ao fato de a interrupção ser programada ou não programada.

Além disso, foi realizada outra estrutura através dos métodos (*Max-Min Parents and Children - MMPC*) reduzindo os dados analisado, tal representação está visualmente apresentada na Figura 12.

Figura 12 – Estrutura Bayesiana (*Max-Min Parents and Children - MMPC*). Fonte: Autoria própria gerada pela biblioteca *Matplotlib*.



Observou-se as seguintes relações em comum da estrutura (*Max-Min Parents and Children - MMPC*) com o Método *Hill Climbing*, ( $B, C$ ), ( $B, G$ ), ( $C, A$ ), ( $E, C$ ), ( $E, B$ ), ( $E, A$ ).

As relações exclusivas das Análises em (*Max-Min Parents and Children - MMPC*) são:

- 1º.  $(B, A) \rightarrow$  O motivo da interrupção, B, afeta o tipo da interrupção, A. Ou seja, o motivo está diretamente se relaciona se a interrupção é programada ou não programada;
- 2º.  $(G, C) \rightarrow$  O mês influencia o nível de tensão, C, de operação;
- 3º.  $(G, A) \rightarrow$  O mês influencia o se tipo de interrupção, A, é programada ou não programada;
- 4º.  $(E, G) \rightarrow$  A duração da interrupção, E, pode variar conforme o mês do ano.

Nos tópicos a seguir, serão apresentados os principais achados do estudo, incluindo padrões identificados, estatísticas relevantes e implicações práticas para a operação da rede elétrica. O impacto das variáveis envolvidas será discutido detalhadamente, permitindo uma compreensão mais aprofundada dos fatores que contribuem para a ocorrência de interrupções e, conseqüentemente, fornecendo insights para estratégias de mitigação mais eficazes.

## 6.2 Inferências Condicionais

Após a construção da estrutura Bayesiana, foram elaboradas as tabelas correspondentes às inferências simples, que representam a probabilidade de ocorrência de uma determinada variável de forma isolada, sem levar em consideração qualquer dependência ou influência de outras variáveis no modelo. Esse tipo de inferência permite obter uma visão geral das distribuições de probabilidade dentro da rede, auxiliando na identificação de padrões e tendências globais das interrupções elétricas.

Além disso, foram realizadas inferências condicionais, nas quais se analisou o impacto da ocorrência de uma variável sobre outra, ou seja, a probabilidade de um determinado evento de interrupção ocorrer dado um cenário específico. Esse tipo de análise é essencial para compreender as relações de causa e efeito dentro do sistema, possibilitando a identificação de fatores críticos que influenciam diretamente as falhas na rede de distribuição.

Por meio das inferências condicionais, foi possível avaliar diferentes panoramas de interrupção elétrica, considerando cenários como diferentes níveis de tensão da rede, variações sazonais, agentes responsáveis e tipos de falhas. Essas análises permitem prever possíveis padrões de interrupções com base em condições previamente conhecidas, possibilitando uma melhor gestão da rede e uma tomada de decisão mais eficiente para minimizar impactos e melhorar a qualidade do fornecimento de energia.

Dessa forma, a aplicação de inferências simples e condicionais se mostrou fundamental para a extração de conhecimento a partir da rede Bayesiana, permitindo não apenas uma análise estatística das interrupções elétricas, mas também a modelagem probabilística de cenários futuros, contribuindo para a prevenção de falhas e otimização da operação do sistema elétrico.

## 6.3 Resultados para estrutura *Hill Climbing*

### 6.3.1 Inferencias Simples Método *Hill Climbing*

Para a determinação das inferências probabilísticas simples referentes à estrutura obtida pelo algoritmo *Hill Climbing*, foram consideradas as dez concessionárias com maior número de ocorrências de interrupção. Essa abordagem foi adotada devido à grande quantidade de valores distintos para essa variável, o que poderia dificultar a análise.

A Tabela 5 apresenta as probabilidades associadas aos diferentes tipos de interrupção.

Tabela 5 – Distribuição de Probabilidades para Tipos de Interrupção

Variável	Estado	Probabilidade
A	Não Programada	0.948128
A	Programada	0.051872

Já a Tabela 6 exibe as probabilidades para os possíveis motivos das interrupções.

Tabela 6 – Distribuição de Probabilidades para os Motivos das Interrupções

Variável	Estado	Probabilidade
B	0.0	8.054220e-01
B	1.0	4.734718e-02
B	2.0	3.036886e-03
B	3.0	8.624156e-02
B	4.0	4.726863e-04
B	5.0	1.447575e-07
B	6.0	5.657176e-02
B	7.0	1.534959e-04
B	8.0	7.543173e-04

A Tabela 7 detalha a distribuição de probabilidade para os diferentes níveis de tensão.

Além disso, a Tabela 8 apresenta a distribuição probabilística das dez principais concessionárias de energia em termos de ocorrências de interrupção.

Outrossim, a Tabela 9 demonstra a probabilidade da ocorrência de falhas de acordo com duração da interrupção.



Tabela 7 – Distribuição de Probabilidades para Níveis de Tensão

Variável	Estado	Probabilidade
C	Alta Tensão	0.016264
C	Baixa Tensão	0.647533
C	Média Tensão	0.336203

Tabela 8 – Distribuição de Probabilidades para as Concessionárias de Energia

Variável	Estado	Probabilidade
D	CEMIG-D	0.104573
D	EQUATORIAL GO	0.069374
D	COPEL-DIS	0.061562
D	ENEL RJ	0.048998
D	ENEL CE	0.045510
D	Equatorial MA	0.041268
D	ELETROPAULO	0.040236
D	CPFL-PAULISTA	0.038328
D	RGE SUL	0.038065
D	Equatorial PA	0.035628

Tabela 9 – Tabela de Probabilidades para a duração da interrupção

Variável	Estado	Probabilidade
E	Anomalia	0.059067
E	Grande em área rural	0.116497
E	Grande em área urbana	0.245566
E	Interrupção muito pequena	0.020187
E	Interrupção pequena	0.020279
E	Média	0.370830
E	Preocupante	0.147090
E	Ruído	0.020484

Ademais, é mostrado as inferências condicionais referentes ao turno em que a interrupção ocorre pela Tabela 10.

Tabela 10 – Tabela de Probabilidades para o turno em que a interrupção ocorre.

Variável	Estado	Probabilidade
F	Manhã	0.379986
F	Noite	0.253068
F	Tarde	0.366946

Por fim, a Tabela 11 exibe a probabilidade de ocorrência de interrupções ao longo dos meses do ano.

Tabela 11 – Distribuição de Probabilidades de Ocorrência das Interrupções por Mês

Variável	Estado	Probabilidade
G	1	0.100821
G	2	0.088557
G	3	0.090778
G	4	0.077310
G	5	0.074642
G	6	0.067261
G	7	0.069908
G	8	0.075479
G	9	0.081419
G	10	0.101223
G	11	0.098093
G	12	0.074507

### 6.3.2 Inferências por Evidências – Método *Hill Climbing*

Com base nas relações aprendidas pela estrutura gerada pelo algoritmo *Hill Climbing*, observou-se que este modelo se diferencia do método *Max-Min Parents and Children (MMPC)* principalmente devido à exclusão de algumas variáveis durante o processo de aprendizado. Essa remoção ocorreu devido à grande variação apresentada por essas variáveis, o que poderia impactar a estabilidade do modelo e comprometer a precisão das inferências probabilísticas.

A presença de variáveis com alta variação pode dificultar a construção de um modelo mais interpretável e robusto, pois pode introduzir ruído nas relações probabilísticas e gerar inferências menos confiáveis. No *Hill Climbing*, todas as variáveis disponíveis inicialmente foram consideradas para a construção da *Rede Bayesiana*, resultando em uma estrutura mais ampla e detalhada. No entanto, o *MMPC*, ao aplicar critérios de filtragem baseados na relevância estatística, eliminou variáveis com grande dispersão de valores para garantir um modelo mais enxuto e focado nas conexões mais significativas.

Dessa forma, a estrutura gerada pelo *Hill Climbing* captura um maior número de interações entre os fatores que influenciam as interrupções elétricas na rede de distribuição. No entanto, essa abordagem também pode introduzir relações de menor impacto ou que sejam sensíveis a oscilações nos dados. Já o *MMPC*, ao reduzir a complexidade do modelo, prioriza a identificação de variáveis que apresentam um impacto mais estável e relevante dentro da rede probabilística.

Com isso, as inferências obtidas a partir desse modelo permitem uma análise aprofundada das relações entre as variáveis, possibilitando não apenas a identificação das causas prováveis das interrupções, mas também o desenvolvimento de estratégias mais eficazes para otimizar a operação da rede elétrica. As probabilidades inferidas podem ser utilizadas como suporte para ações preventivas, contribuindo para a melhoria da

confiabilidade do fornecimento de energia e para uma gestão mais eficiente dos recursos do sistema elétrico.

Diante disso, observam-se as através das relações encontradas, inferências condicionada à evidencia-se detcando-se as relações entre a ocorrência em concessionárias da CEMIG-D dada as durações catalogadas representada na Tabela 12, os motivos de falhas e os níveis de tensão que mais influenciam interrupções na concessionária mencionada, são apresentadas respectivamente nas Tabela 13 e Tabela 14.

Tabela 12 – Tabela de Probabilidades para a Variável Inferida D considerando Diferentes Evidências de E

Variável Inferida	Estado	Evidência	Probabilidade
D	CEMIG-D	E = Anomalia	0.089936
D	CEMIG-D	E = Grande em área rural	0.095410
D	CEMIG-D	E = Grande em área urbana	0.113142
D	CEMIG-D	E = Interrupção muito pequena	0.080755
D	CEMIG-D	E = Interrupção pequena	0.067518
D	CEMIG-D	E = Média	0.118247
D	CEMIG-D	E = Preocupante	0.086653
D	CEMIG-D	E = Ruído	0.037459

Tabela 13 – Tabela de Probabilidades para a Variável Inferida D considerando Diferentes Evidências de B

Variável Inferida	Estado	Evidência	Probabilidade
D	CEMIG-D	B = 0	1.071820e-01
D	CEMIG-D	B = 1	3.258330e-02
D	CEMIG-D	B = 2	6.383709e-02
D	CEMIG-D	B = 3	1.380435e-01
D	CEMIG-D	B = 4	6.552084e-07
D	CEMIG-D	B = 5	2.139495e-03
D	CEMIG-D	B = 6	8.061495e-02
D	CEMIG-D	B = 7	1.718431e-02
D	CEMIG-D	B = 8	5.489002e-02

Tabela 14 – Tabela de Probabilidades para a Variável Inferida D considerando Diferentes Evidências de C

Variável Inferida	Estado	Evidência	Probabilidade
D	CEMIG-D	C = Baixa Tensão	0.107171
D	CEMIG-D	C = Média Tensão	0.103413
D	CEMIG-D	C = Alta Tensão	0.025128

## 6.4 Resultados *Max-Min Parents and Children (MMPC)*

### 6.4.1 Inferências Simples Método Max-Min Parents and Children (MMPC)

As inferências simples consideradas pela análise (*MMPC*) estão visualmente apresentadas nas Tabela 15, Tabela 16, Tabela 17, Tabela 18 e Tabela 19.

Tabela 15 – Tabela de Probabilidades para a Variável A

Variável	Estado	Probabilidade
A	Não Programada	0.948128
A	Programada	0.051872

Tabela 16 – Tabela de Probabilidades para a Variável B

Variável	Estado	Probabilidade
B	0.0	8.054220e-01
B	1.0	4.734718e-02
B	2.0	3.036886e-03
B	3.0	8.624156e-02
B	4.0	4.726863e-04
B	5.0	1.447575e-07
B	6.0	5.657176e-02
B	7.0	1.534959e-04
B	8.0	7.543173e-04

Tabela 17 – Tabela de Probabilidades para a Variável C

Variável	Estado	Probabilidade
C	Alta Tensão	0.016264
C	Baixa Tensão	0.647533
C	Média Tensão	0.336203

Tabela 18 – Tabela de Probabilidades para a Variável E

Variável	Estado	Probabilidade
E	Anomalia	0.059067
E	Grande em área rural	0.116497
E	Grande em área urbana	0.245566
E	Interrupção muito pequena	0.020187
E	Interrupção pequena	0.020279
E	Média	0.370830
E	Preocupante	0.147090
E	Ruído	0.020484

Tabela 19 – Tabela de Probabilidades para a Variável G

Variável	Estado	Probabilidade
G	1	0.100821
G	2	0.088557
G	3	0.090778
G	4	0.077310
G	5	0.074642
G	6	0.067261
G	7	0.069908
G	8	0.075479
G	9	0.081419
G	10	0.101223
G	11	0.098093
G	12	0.074507

#### 6.4.2 Inferências por Evidências – Método Max-Min Parents and Children (MMPC)

Por meio do método *Max-Min Parents and Children (MMPC)*, analisou-se a probabilidade da causalidade de um determinado tipo de interrupção em função do nível de tensão da ocorrência, conforme representado na Tabela 20. Essa análise permite compreender a relação entre o nível de tensão e os diferentes tipos de falhas, fornecendo uma visão detalhada sobre como as interrupções variam de acordo com as características da rede elétrica.

Além disso, examinou-se a relação entre os motivos das falhas e os diferentes níveis de tensão, buscando compreender quais motivos de interrupções ocorrem com maior frequência em redes de baixa, média e alta tensão. Os achados dessa investigação estão apresentados na Tabela 21, permitindo identificar quais causas estão mais associadas a cada nível de tensão e como esses fatores influenciam a estabilidade da rede elétrica.

Ademais, investigou-se a influência de outros fatores na ocorrência de falhas no sistema. Em particular, analisou-se o impacto da variação sazonal ao longo dos meses, identificando padrões sazonais e tendências associadas a diferentes períodos do ano. Os resultados dessa análise estão organizados na Tabela 22, destacando a influência do mês sobre a frequência e a distribuição das falhas.

Essas análises, realizadas por meio do método MMPC, fornecem informações essenciais para a modelagem e compreensão das relações de dependência entre os fatores estudados, contribuindo para um diagnóstico mais preciso e fundamentado sobre o comportamento das falhas no sistema elétrico.

Tabela 20 – Tabela de Probabilidades para a Variável Inferida A considerando Diferentes Evidências de C

Variável Inferida	Estado	Evidência	Probabilidade
A	Não Programada	C = Baixa Tensão	0.976393
A	Programada	C = Baixa Tensão	0.023607
A	Não Programada	C = Média Tensão	0.892869
A	Programada	C = Média Tensão	0.107131
A	Não Programada	C = Alta Tensão	0.965104
A	Programada	C = Alta Tensão	0.034896

Tabela 21 – Tabela de Probabilidades para a Variável Inferida C considerando Diferentes Evidências de B

Variável Inferida	Estado	Evidência	Probabilidade
C	Alta Tensão	B = 0	0.018385
C	Baixa Tensão	B = 0	0.661604
C	Média Tensão	B = 0	0.320011
C	Alta Tensão	B = 1	0.000646
C	Baixa Tensão	B = 1	0.932094
C	Média Tensão	B = 1	0.067260
C	Alta Tensão	B = 2	0.004083
C	Baixa Tensão	B = 2	0.099320
C	Média Tensão	B = 2	0.896597
C	Alta Tensão	B = 3	0.006289
C	Baixa Tensão	B = 3	0.428016
C	Média Tensão	B = 3	0.565695
C	Alta Tensão	B = 4	0.000012
C	Baixa Tensão	B = 4	0.926969
C	Média Tensão	B = 4	0.073018
C	Alta Tensão	B = 5	0.040650
C	Baixa Tensão	B = 5	0.260163
C	Média Tensão	B = 5	0.699187
C	Alta Tensão	B = 6	0.009801
C	Baixa Tensão	B = 6	0.580472
C	Média Tensão	B = 6	0.409726
C	Alta Tensão	B = 7	0.379084
C	Baixa Tensão	B = 7	0.017635
C	Média Tensão	B = 7	0.603282
C	Alta Tensão	B = 8	0.342615
C	Baixa Tensão	B = 8	0.048284
C	Média Tensão	B = 8	0.609101

Tabela 22 – Tabela de Probabilidades para a Variável Inferida C considerando Diferentes Evidências de G

Variável Inferida	Estado	Evidência	Probabilidade
C	Alta Tensão	$G = 1$	0.012897
C	Baixa Tensão	$G = 1$	0.616085
C	Média Tensão	$G = 1$	0.371018
C	Alta Tensão	$G = 2$	0.021918
C	Baixa Tensão	$G = 2$	0.612916
C	Média Tensão	$G = 2$	0.365166
C	Alta Tensão	$G = 3$	0.018829
C	Baixa Tensão	$G = 3$	0.624364
C	Média Tensão	$G = 3$	0.356806
C	Alta Tensão	$G = 4$	0.015512
C	Baixa Tensão	$G = 4$	0.652386
C	Média Tensão	$G = 4$	0.332102
C	Alta Tensão	$G = 5$	0.017519
C	Baixa Tensão	$G = 5$	0.681045
C	Média Tensão	$G = 5$	0.301436
C	Alta Tensão	$G = 6$	0.018787
C	Baixa Tensão	$G = 6$	0.703915
C	Média Tensão	$G = 6$	0.277298
C	Alta Tensão	$G = 7$	0.018078
C	Baixa Tensão	$G = 7$	0.702329
C	Média Tensão	$G = 7$	0.279593
C	Alta Tensão	$G = 8$	0.032401
C	Baixa Tensão	$G = 8$	0.668040
C	Média Tensão	$G = 8$	0.299560
C	Alta Tensão	$G = 9$	0.011386
C	Baixa Tensão	$G = 9$	0.661973
C	Média Tensão	$G = 9$	0.326640
C	Alta Tensão	$G = 10$	0.009720
C	Baixa Tensão	$G = 10$	0.629057
C	Média Tensão	$G = 10$	0.361223
C	Alta Tensão	$G = 11$	0.010422
C	Baixa Tensão	$G = 11$	0.636665
C	Média Tensão	$G = 11$	0.352913
C	Alta Tensão	$G = 12$	0.012081
C	Baixa Tensão	$G = 12$	0.621390
C	Média Tensão	$G = 12$	0.366529

## 7 Conclusão

Portanto, analisando os elementos supracitados, torna-se evidente a relação entre as variáveis destacadas ao longo do processo e as interrupções identificadas por meio da estruturação Bayesiana. A modelagem probabilística comprovou a influência de diversos fatores na classificação das falhas em redes de distribuição elétrica, permitindo uma análise detalhada e embasada dos eventos que afetam a operação do sistema.

O desenvolvimento da aplicação de machine learning baseou-se em uma robusta coleta e estruturação de dados históricos sobre falhas elétricas, considerando fatores meteorológicos, operacionais e estruturais. Esse processo permitiu a construção de um conjunto de dados confiável e representativo da realidade do sistema elétrico, essencial para a modelagem das Redes Bayesianas.

Para garantir a qualidade e a precisão dos modelos desenvolvidos, foi utilizada a métrica BIC Score (Bayesian Information Criterion), um critério estatístico amplamente empregado na avaliação de modelos probabilísticos. Essa métrica possibilitou a seleção das melhores estruturas de Rede Bayesiana, assegurando que os modelos obtidos representassem adequadamente as relações probabilísticas entre os eventos de falha e os fatores condicionantes.

Com base nessa abordagem, a aplicação desenvolvida foi capaz de identificar padrões por meio de inferências probabilísticas, possibilitando a previsão de interrupções sazonais e a identificação dos turnos mais propensos a falhas. Além disso, permitiu diagnósticos precisos considerando variáveis como o motivo da falha, tipo de ocorrência, nível de tensão e duração das interrupções.

Os indicadores extraídos do modelo demonstraram-se fundamentais para a otimização dos processos de manutenção e aprimoramento da confiabilidade do sistema elétrico de distribuição. Dessa forma, a abordagem proposta não apenas fortalece a capacidade de diagnóstico do sistema, mas também contribui significativamente para a gestão eficiente das redes elétricas, permitindo ações preventivas e corretivas mais assertivas, alinhadas às necessidades operacionais do setor.



# Referências

Agência Nacional de Energia Elétrica (ANEEL). *Resolução Normativa ANEEL 270/07*. 2007. <<https://www.aneel.gov.br>>. Acesso em: 25 ago. 2024. Citado 2 vezes nas páginas 1 e 3.

Agência Nacional de Energia Elétrica (ANEEL). *Resolução Normativa ANEEL N° 956, de 29 de junho de 2021*. 2021. <[https://www2.aneel.gov.br/cedoc/aren2021956\\_2\\_7.pdf](https://www2.aneel.gov.br/cedoc/aren2021956_2_7.pdf)>. Acesso em: 25 ago. 2024. Citado na página 7.

Agência Nacional de Energia Elétrica (ANEEL). *Dados Abertos ANEEL*. 2024. <<https://dadosabertos.aneel.gov.br/>>. Acesso em: 25 ago. 2024. Citado 2 vezes nas páginas 5 e 15.

BARTH, V. B. d. O. *Um método Bayesiano orientado a dados para o aprendizado estrutural de Redes Bayesianas*. Tese (Doutorado) — Universidade de São Paulo, 2023. Citado na página 4.

BHATTARAI, B. P. et al. Big data analytics in smart grids: state-of-the-art, challenges, opportunities, and future directions. *IET Smart Grid*, Wiley Online Library, v. 2, n. 2, p. 141–154, 2019. Citado na página 2.

BOBBIO, A. et al. Improving the Analysis of Dependable Systems by Mapping Fault Trees into Bayesian Networks. *Reliability Engineering & System Safety*, Elsevier, v. 71, n. 3, p. 249–260, 2001. Citado na página 10.

CAMARGO, C. C. d. B. et al. Gerenciamento pelo lado da demanda: metodologia para identificação do potencial de conservação de energia elétrica de consumidores residenciais. 1996. Citado na página 7.

CHARNIAK, E. Bayesian Networks Without Tears: Making Bayesian Networks More Accessible to the Probabilistically Unsophisticated. *AI Magazine*, American Association for Artificial Intelligence, Menlo Park, CA, USA, v. 12, n. 4, p. 50–63, 1991. ISSN 0738-4602. Disponível em: <<http://portal.acm.org/citation.cfm?id=122716>>. Citado na página 8.

CONTALDI, C. *Bayesian Network Hybrid Learning Using a Parent Reducing Site-specific Mutation Rate Genetic Algorithm*. Tese (Doutorado) — University of Illinois at Chicago, 2016. Citado na página 13.

CUNHA, L. V. Desempenho de linhas de transmissão frente a descargas atmosféricas: influência do efeito corona na ruptura a meio de vão. Universidade Federal de Minas Gerais, 2010. Citado na página 3.

Empresa de Pesquisa Energética (EPE). *Relatório Anual de Energia 2024*. 2024. <<https://www.epe.gov.br>>. Acesso em: 25 ago. 2024. Citado na página 1.

FERREIRA, J. T. V. *Redes de distribuição de energia elétrica de média e baixa tensão: estágio na Helenos, SA*. Tese (Doutorado), 2018. Citado na página 7.

- FERREIRA, V. H. et al. Análise do impacto das ações de manutenção nos indicadores de continuidade em redes de distribuição utilizando machine learning e regressão com dados em painel. *Simposio Brasileiro de Sistemas Elétricos-SBSE*, v. 1, n. 1, 2020. Citado na página 5.
- GHORBANIAN, M.; DOLATABADI, S. H.; SIANO, P. Big data issues in smart grids: A survey. *IEEE Systems Journal*, IEEE, v. 13, n. 4, p. 4158–4168, 2019. Citado na página 2.
- GIL, A. C. *Métodos e Técnicas de Pesquisa Social*. São Paulo: Atlas, 2008. Citado na página 14.
- HAZZAN, S.; IEZZI, G. *Fundamentos de Matemática Elementar, Vol. 5*. [S.l.]: Atual, 2004. Citado na página 9.
- KINJO, E. M. et al. Modelagem bayesiana aplicada para cálculo da probabilidade de falha em sistemas de saúde iot. *RISTI-Revista Ibérica de Sistemas e Tecnologias de Informação*, Associação Ibérica de Sistemas e Tecnologias de Informação (AISTI), n. 47, p. 87–108, 2022. Citado na página 4.
- KOLLER, D.; FRIEDMAN, N. *Probabilistic Graphical Models: Principles and Techniques*. Cambridge, MA: MIT Press, 2009. Citado na página 4.
- LIBRANTZ, A. F. H. et al. Risk assessment in software supply chains using the Bayesian method. *International Journal of Production Research*, v. 59, n. 22, p. 6758–6775, 2020. Citado na página 2.
- MACHADO, I. G. et al. Aprendizagem estrutural de redes bayesianas utilizando algoritmo genético multi-agente. Universidade Federal de Minas Gerais, 2021. Citado 2 vezes nas páginas 4 e 17.
- MAMEDE, J. F. *Proteção de Sistemas Elétricos de Potência*. 3. ed. Rio de Janeiro, RJ: LTC, 2011. 605 p. Citado na página 8.
- MARQUES, R. L.; DUTRA, I. Redes bayesianas: o que são, para que servem, algoritmos e exemplos de aplicações. *Coppe Sistemas-Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brasil*, 2002. Citado na página 11.
- Matplotlib Development Team. *Matplotlib: Visualization with Python*. 2024. <<https://matplotlib.org/>>. Acesso em: 25 ago. 2024. Citado 2 vezes nas páginas iv e 19.
- MENDONÇA, M. J.; PEREIRA, A. O.; AYLMER, R. R. *Decomposição estrutural do consumo de energia elétrica no Brasil*. [S.l.], 2022. Citado na página 4.
- MEYER, P. L. *Probabilidade: Aplicações à Estatística*. 2. ed. [S.l.]: LTC, 2000. Citado 2 vezes nas páginas 8 e 9.
- MURPHY, K. P. *Dynamic bayesian networks: representation, inference and learning*. [S.l.]: University of California, Berkeley, 2002. Citado 3 vezes nas páginas 4, 8 e 20.
- NEAPOLITAN, R. E. *Learning Bayesian Networks*. Upper Saddle River, NJ: Pearson Prentice Hall, 2003. Citado 4 vezes nas páginas 4, 8, 11 e 13.
- NEOENERGIA. *Relatório Anual de Sustentabilidade 2019*. 2019. <<https://www.neoenergia.com>>. Acesso em: 25 ago. 2024. Citado na página 1.

- NETO, A. B. F.; CORRÊA, W. L. R.; PEROBELLI, F. S. Consumo de energia e crescimento econômico: uma análise do Brasil no período 1970-2009. *Análise Econômica*, v. 34, n. 65, p. 181–204, sep 2019. Citado na página 1.
- NetworkX Development Team. *NetworkX: Network Analysis in Python*. 2024. <<https://networkx.org/>>. Acesso em: 25 ago. 2024. Citado 2 vezes nas páginas iv e 19.
- NumPy. *NumPy: The fundamental package for numerical computing in Python*. 2024. <<https://numpy.org/pt/>>. Acesso em: 22 ago. 2024. Citado 3 vezes nas páginas iv, 16 e 17.
- Pandas Development Team. *Pandas Documentation*. 2024. <<https://pandas.pydata.org/docs/index.html>>. Acesso em: 25 ago. 2024. Citado 2 vezes nas páginas iv e 17.
- PAULINO, C. D.; TURKMAN, M. A. A.; MURTEIRA, B. *Estatística Bayesiana*. Lisboa: Fundação Calouste Gulbenkian, 2003. Citado na página 11.
- pgmpy Development Team. *pgmpy: Python Library for Probabilistic Graphical Models*. 2024. <<https://pgmpy.org/>>. Acesso em: 25 ago. 2024. Citado 3 vezes nas páginas iv, 18 e 19.
- Project Jupyter. *Jupyter: An Open-Source Project for Interactive Computing*. 2024. <<https://jupyter.org/>>. Acesso em: 21 ago. 2024. Citado 2 vezes nas páginas iv e 15.
- Python Software Foundation. *About Python Programming Language*. 2024. <<https://www.python.org/about/>>. Acesso em: 25 ago. 2024. Citado 3 vezes nas páginas iv, 14 e 15.
- RUSSELL, S. J.; NORVIG, P. *Artificial Intelligence: A Modern Approach*. New Jersey: Prentice Hall, 1995. Citado na página 11.
- SAGIROGLU, S. et al. Big data issues in smart grid systems. In: *2016 IEEE International Conference on Renewable Energy Research and Applications (ICRERA)*. [S.l.]: IEEE, 2016. p. 1007–1012. Citado na página 2.
- SANTOS, M. A. dos; NASCIMENTO, G. N. L. do. *Metodologia Científica: A Pesquisa como Compreensão da Realidade*. Palmas: Programa de Mestrado em Ciências da Saúde, 2021. 93 p. Citado na página 14.
- SOUZA, E. C. d. et al. Impactos das mudanças climáticas sobre o bem-estar relacionado à saúde no Brasil. *Pesquisa e Planejamento Econômico*, v. 43, n. 1, p. 49–87, 2013. Citado na página 1.
- TRONCHONI, A. B. *Identificação de Causas de Desligamentos Não Programados em Redes de Distribuição*. Dissertação (Dissertação de Mestrado) — Pontifícia Universidade Católica do Rio Grande do Sul, Porto Alegre, 2008. Orientador: Prof. Flávio Antonio Becon Lemos, Dr. Citado na página 20.
- TSAMARDINOS, I.; BROWN, L. E.; ALIFERIS, C. F. The max-min hill-climbing bayesian network structure learning algorithm. *Machine Learning*, v. 65, n. 1, p. 31–78, 2006. Citado na página 17.

## Apêndices

## APÊNDICE A – Tabelas de Inferencias

Texto

## APÊNDICE B – Gráficos de Barras

Texto

## APÊNDICE C – Gráficos Boxplot

Texto

## APÊNDICE D – Gráficos Função de Distribuição Acumulada

Texto



## Anexos

# ANEXO A – Interrupções de Energia Elétrica nas Redes de Distribuição

O documento a seguir apresenta informações detalhadas sobre as interrupções de energia elétrica no Brasil. O conteúdo completo pode ser encontrado a seguir.

## Dicionário de Metadados do Conjunto de Dados

### Conjunto de Dados

Interrupções de Energia Elétrica nas Redes de Distribuição

### Visão Geral

Dados de todas as interrupções de energia elétrica ocorridas nas redes de concessionárias de distribuição do país. Não constam as interrupções das permissionárias (cooperativas).

### Metadados

#### Nome do arquivo:

interrupcoes-energia-eletrica-[ANO] (csv, xml e json)

#### Resumo descritivo do arquivo:

Relação, por ano, das informações dos registros de interrupções nas redes elétricas de distribuição com a identificação da unidade consumidora, qual linha de transmissão e instalação elétrica bem como dados do tipo de interrupção, a data de início e fim da mesma e demais informações necessárias para entendimento dos registros dessas ocorrências.

#### Catálogo origem

<https://dadosabertos.aneel.gov.br>

#### Órgão responsável

ANEEL – Agência Nacional de Energia Elétrica

E-mail institucional da área responsável  
[monitoramento.sfe@aneel.gov.br](mailto:monitoramento.sfe@aneel.gov.br)

Categorias no VCGE  
Energia elétrica

Palavras-chave do conjunto de dados.  
interrupção, distribuidora, conjuntos elétricos, expurgos, indicadores de continuidade

Frequência de atualização  
Mensal

Detalhamento dos campos:

Nome do Campo	Tipo do dado	Tamanho do Campo	Descrição
DatGeracaoConjuntoDados	Data simples		Data do processamento de carga automática no momento da geração para publicação do conjunto de dados abertos.
IdConjuntoUnidadeConsumidora	Númerico	10	Código do conjunto da unidade consumidora.
DscAlimentadorSubestacao	Cadeia Caracteres	255	Descrição do alimentador da subestação. (Linha de transmissão que sai de uma subestação de distribuição e segue para a distribuição de energia)
DscSubestacaoDistribuicao	Cadeia Caracteres	255	Descrição da subestação de distribuição (Instalação elétrica de alta potência, contendo equipamentos para transmissão e distribuição de energia elétrica, além de equipamentos de proteção e controle)
NumOrdemInterrupcao	Cadeia Caracteres	255	Descrição da Ordem de Interrupção. (Ofício ou normativo determinando interrupções de energia)
DscTipoInterrupcao	Cadeia Caracteres	14	Armazenar os tipos de interrupções existentes que se relacionarão com os arquivos enviados pelos Agentes via DutoNet: - Não Programada - Programada

			<p>Identificador referenciando o Motivo do Expurgo.</p> <p>0 - Não houve expurgo</p> <p>1 - Falha nas instalações da unidade consumidora que não provoque interrupção em instalações de terceiros</p> <p>2 - Interrupção decorrente de obras de interesse exclusivo do consumidor e que afete somente a unidade consumidora do mesmo</p> <p>3 - Interrupção em situação de emergência</p> <p>4 - Suspensão por inadimplemento do consumidor ou por deficiência técnica e/ou de segurança das instalações da unidade consumidora que não provoque interrupção em instalações de terceiros, previstas em regulamentação</p> <p>5 - Vinculadas a programas de racionamento instituídos pela União</p> <p>6 - Ocorridas em dia crítico</p> <p>7 - Oriundas de atuação de esquemas de alívio de carga solicitado pelo ONS</p> <p>8 - Interrupção de origem externa ao sistema de distribuição</p>
IdeMotivoExpurgo	Numérico	2	
DatInicioInterrupcao	Data Hora		Data e hora do início da interrupção.
DatFimInterrupcao	Data Hora		Data e hora do reestabelecimento (Fim) da interrupção.
FatGeradorInterrupcao	Cadeia Caracteres	255	Descrição do fato gerador da interrupção.
NumNivelTensao	Numérico	10	Nível de tensão onde o fato gerador foi verificado.
NumUnidadeConsumidora	Numérico	10	Número de unidades consumidoras atingidas.
NumConsumidorConjunto	Numérico	10	Número de consumidores do conjunto.
NumAno	Cadeia de caracteres	4	Ano de início da Interrupção
NomAgenteRegulado	Cadeia de Caracteres	255	Nome do agente do setor regulado.
SigAgente	Cadeia de caracteres	20	Sigla que abrevia o nome dos Agentes regulados pela ANEEL
NumCPFCNPJ	Cadeia de caracteres	14	Número do CNPJ ou CPF do agente fiscalizado.