

**UNIVERSIDADE DE SÃO PAULO**

Instituto de Ciências Matemáticas e de Computação

## Predição de defeitos em reatores da Eletrobras Eletronorte utilizando regras de associação

**Walmor Vieira Gomes**

Monografia - MBA em Inteligência Artificial e Big Data



SERVIÇO DE PÓS-GRADUAÇÃO DO ICMC-USP

Data de Depósito:

Assinatura: \_\_\_\_\_

**Walmor Vieira Gomes**

## **Predição de defeitos em reatores da Eletrobras Eletronorte utilizando regras de associação**

Monografia apresentada ao Departamento de Ciências de Computação do Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo - ICMC/USP, como parte dos requisitos para obtenção do título de Especialista em Inteligência Artificial e Big Data.

Área de concentração: Inteligência Artificial

Orientador: Prof. Willian Dener de Oliveira

**Versão original**

**São Carlos**

**2023**

Ficha catalográfica elaborada pela Biblioteca Prof. Achille Bassi  
e Seção Técnica de Informática, ICMC/USP,  
com os dados inseridos pelo(a) autor(a)

G633p      Gomes, Walmor Vieira  
             Predição de defeitos em reatores da Eletrobras  
Eletronorte utilizando regras de associação / Walmor  
Vieira Gomes; orientador Willian Dener Oliveira. --  
São Carlos, 2023.  
             49 p.

             Trabalho de conclusão de curso (MBA em  
Inteligência Artificial e Big Data) -- Instituto de  
Ciências Matemáticas e de Computação, Universidade  
de São Paulo, 2023.

             1. SAGE. 2. Regras de associação. 3. Reatores. 4.  
Manutenção preditiva. I. Oliveira, Willian Dener,  
orient. II. Título.

**Walmor Vieira Gomes**

Monograph presented to the Departamento de Ciências de Computação do Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo - ICMC/USP, as part of the requirements for obtaining the title of Specialist in Artificial Intelligence and Big Data.

Concentration area: Artificial Intelligence

Advisor: Prof. Willian Dener de Oliveira

**Original version**

**São Carlos**

**2023**



## **AGRADECIMENTOS**

Aos meus pais, Manoel e Maria da Conceição, por todo amor devotado ao longo da minha vida,

A minha esposa, Gisele, por toda a compreensão ao longo dos meses, principalmente nos finais de semana em que precisei me dedicar ao trabalho,

Ao meu orientador, prof. Willian Dener de Oliveira, pelo aprendizado e paciência,

À Eletronorte, pela oportunidade e confiança,

Meus sinceros agradecimentos.





## RESUMO

GOMES, W. V. **Predição de defeitos em reatores da Eletrobras Eletronorte utilizando regras de associação.** 2023. 49p. Monografia (MBA em Inteligência Artificial e Big Data) - Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos, 2023.

A Eletrobras Eletronorte, concessionária no segmento de energia elétrica, possui uma grande rede de instalações que exige monitoramento contínuo. Dados são coletados em intervalos de milissegundos, portanto é necessário o uso de Inteligência Artificial para sua gestão eficiente. Este projeto aplicou o método de mineração de dados chamado Regras de Associação para encontrar relações entre alarmes e desligamentos de reatores, equipamentos essenciais no setor de energia, com o propósito de aperfeiçoar a manutenção preventiva. Após análise do banco de dados SAGE, seleção e preparação, o método encontrou regras para orientar a manutenção, porém não há força suficiente para explicar os defeitos dos reatores (Suporte de apenas 10% e Confiança pouco acima de 50%).

**Palavras-chave:** SAGE. Regras de associação. Reatores. Manutenção preventiva.



## **ABSTRACT**

GOMES, W. V. **Defect prediction in Eletrobras Eletronorte reactors using association rules.** 2023. 49p. Monograph (MBA in Artificial Intelligence and Big Data) - Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos, 2023.

Eletrobras Eletronorte, an electric power company, has an extensive network of facilities that requires continuous monitoring. Data is collected at millisecond intervals, making the use of Artificial Intelligence essential for its efficient management. This project applied the data mining method called Association Rules to find relationships between alarms and reactor shutdowns, essential equipment in the energy sector, with the purpose of improving preventive maintenance. After analyzing the SAGE database, along with selection and preparation, the method found rules to guide maintenance, but there is not enough strength to explain reactor defects (Support is only 10%, and Confidence is slightly above 50%).

**Keywords:** SAGE. Association Rules. Reactors. Preventive maintenance.



## LISTA DE FIGURAS

Figura 1 – COI-GT Eletronorte (Fonte: Autor) . . . . .	18
Figura 2 – Reator de 230 kV da Eletronorte (Fonte: Autor) . . . . .	19
Figura 3 – Modelo de Dados da Configuração do SCD (MANUAL SAGE, 2020) . . . . .	21
Figura 4 – Tela de Alarmes/Eventos do SAGE (Fonte: autor) . . . . .	22
Figura 5 – Diagrama elétrico de linhas de transmissão com uso de reatores para compensação (CIGRÉ, 2023) . . . . .	23
Figura 6 – Tipos de Manutenção conforme Criticidade (MARQUES, 1996) . . . . .	24
Figura 7 – Visão geral das etapas do processo KDD (FAYYAD; PIATETSKY- SHAPIRO; SMYTH, 1996) . . . . .	28
Figura 8 – Atributos da Entidade PAS_R (Fonte: Autor) . . . . .	30
Figura 9 – Amostra da Entidade PAS_R (Fonte: Autor) . . . . .	30
Figura 10 – Atributos da Entidade PDS_R (Fonte: Autor) . . . . .	30
Figura 11 – Amostra da Entidade PDS_R (Fonte: Autor) . . . . .	31
Figura 12 – Atributos da Entidade EVE_H (Fonte: Autor) . . . . .	31
Figura 13 – Amostra da Entidade EVE_H (Fonte: Autor) . . . . .	31
Figura 14 – Contagem dos Atributos mais Frequentes (Fonte: Autor) . . . . .	32
Figura 15 – Exemplo de uso da função <i>apriori</i> : Suporte mínimo ajustado em 20% (Fonte: Autor) . . . . .	32
Figura 16 – Exemplo de uso da função <i>association_rules</i> : Confiança mínima ajustada em 50% para <i>itemsets frequentes</i> com Suporte mínimo ajustado em 20% (Fonte: Autor) . . . . .	33
Figura 17 – Concatenação dos arquivos em um dataframe (Fonte: Autor) . . . . .	43
Figura 18 – Dataframe modificado (Fonte: Autor) . . . . .	45
Figura 19 – Dicionário de medidas (Fonte: Autor) . . . . .	46
Figura 20 – Dataframe de medidas analógicas, padronizado pelo dicionário de medi- das (Fonte: Autor) . . . . .	46
Figura 21 – Dataframe de medidas digitais, padronizado pelo dicionário de medidas (Fonte: Autor) . . . . .	47
Figura 22 – Exemplo de transposição com linhas do dataframe de medidas digitais (Fonte: Autor) . . . . .	48
Figura 23 – Dataframe final: todas as medidas (transações) analógicas e digitais de reatores (Fonte: Autor) . . . . .	49
Figura 24 – Dataframe final: detalhamento (Fonte: Autor) . . . . .	49



## SUMÁRIO

1	INTRODUÇÃO . . . . .	17
1.1	Contextualização . . . . .	17
1.2	Motivação . . . . .	17
1.3	Objetivo . . . . .	18
2	FUNDAMENTAÇÃO TEÓRICA E TRABALHOS RELACIONADOS	21
2.1	Sistema SAGE . . . . .	21
2.2	Equipamento Reator . . . . .	23
2.3	Regras de Associação . . . . .	24
2.4	Algoritmo Apriori . . . . .	25
2.5	Processo KDD . . . . .	27
2.6	Trabalhos Relacionados . . . . .	27
3	APLICAÇÃO DE REGRAS DE ASSOCIAÇÃO NA PREDIÇÃO DE DESLIGAMENTOS DE REATORES DA ELETROBRAS ELETRO- NORTE . . . . .	29
3.1	Extração de dados . . . . .	29
3.2	Pré-Processamento . . . . .	29
3.3	Mineração de dados . . . . .	31
3.4	Resultados . . . . .	33
4	CONCLUSÃO . . . . .	35
	REFERÊNCIAS . . . . .	37
	APÊNDICE A – EXTRAÇÃO DE DADOS DO SAGE . . . . .	39
	APÊNDICE B – PRÉ-PROCESSAMENTO DOS DADOS . . . . .	43





# 1 INTRODUÇÃO

## 1.1 Contextualização

A Eletrobras Eletronorte é uma empresa brasileira, subsidiária da Eletrobras e concessionária no segmento de energia elétrica. Possui capacidade de geração instalada de 9.049,95 MW, opera 57 subestações e mais de onze mil quilômetros de linhas de transmissão (HISTÓRIA DA ELETRONORTE, 2023).

Para desenvolver o processo de gestão de seu conjunto de instalações, a empresa inaugurou em 2022 o Centro de Operação Integrada de Geração e Transmissão - COI-GT (Figura 1) - em Brasília. O uso das técnicas de Inteligência Artificial (IA) em centros de controle como este é considerado a única alternativa para tratar de modo eficiente o crescente volume de dados do sistema (PEREIRA *et al.*, 2018).

Para executar as tarefas de supervisão, controle e gestão dos equipamentos, a empresa utiliza o software SAGE (Sistema Aberto de Gerenciamento de Energia), desenvolvido pela empresa Cepel, da qual a Eletrobras Eletronorte é associada fundadora. O SAGE implementa as funções de aquisição, tratamento e distribuição de dados do sistema elétrico (SCADA). Dentre suas ferramentas, o Subsistema de Aquisição e Controle (SCD) implementa as funções de pré-processamento dos dados de tempo real (alarmes, eventos e cálculos), cujo resultado pode ser acessado imediatamente ou por meio do Gerenciador de Arquivos Históricos e de Tendência (PEREIRA *et al.*, 2018).

## 1.2 Motivação

No COI-GT, um evento é a condição de um equipamento indisponível, seja por desligamento ou impossibilidade de manobra, e os alarmes registram condições eletromecânicas de operação de um equipamento indisponível ou fora da faixa de operação desejada. O SAGE monitora simultaneamente os eventos e alarmes dos equipamentos em todas as instalações da Eletrobras Eletronorte.

As equipes de Operação e Manutenção (O&M) da Eletrobras Eletronorte têm como meta reduzir a frequência e a duração dos eventos. Conforme legislação brasileira, a qualidade do serviço de transmissão de energia elétrica está associada à disponibilidade dos equipamentos (AGÊNCIA NACIONAL DE ENERGIA ELÉTRICA – ANEEL, 2022), portanto há restrições e penalidades para os eventos, principalmente quando intempestivos.

Além disso, também é uma meta das equipes de O&M a otimização dos desligamentos para reduzir custos operacionais, principalmente relacionados aos recursos humanos e ao aluguel de equipamentos.

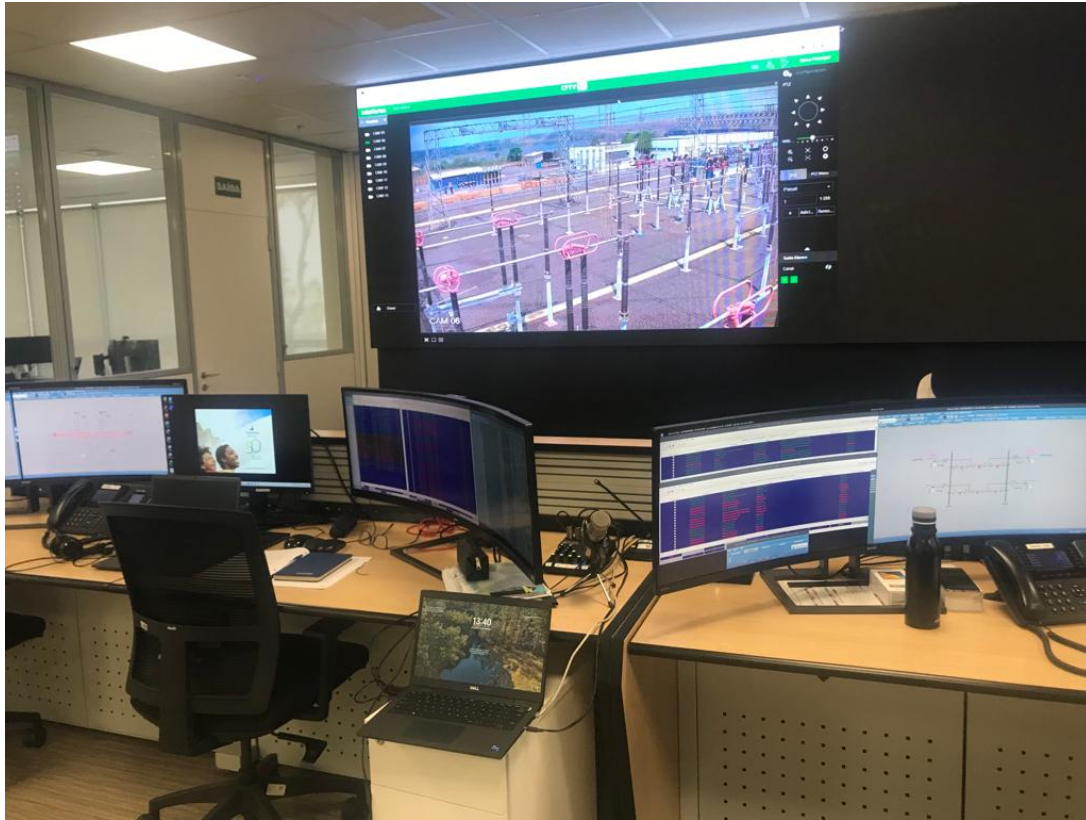


Figura 1 – COI-GT Eletronorte (Fonte: Autor)

Com aplicação dos instrumentos de IA, é possível analisar o grande volume de dados do SAGE produzido pelo conjunto de medições. Ainda assim, considerando os prazos e a complexidade dos dados, a análise será restrita à família de equipamentos reatores (Figura 2) pertencente à Rede Básica do Sistema Interligado Nacional de Transmissão de Energia Elétrica - SIN.

Reatores são dispositivos elétricos de impedância indutiva, utilizados para limitação de corrente e controle de tensão. São equipamentos indispensáveis para o setor de transmissão elétrica e, devido ao custo elevado, a aquisição é regulamentada pela ANEEL, a qual exige estudos e planejamento de médio a longo prazo (BELTRÃO, 2014).

### 1.3 Objetivo

Ao aplicar o método de mineração de dados chamado Regras de Associação, espera-se encontrar os alarmes que precedem o desligamento de equipamentos, conforme escopo definido em 1.2. De forma analítica, são estes os objetivos:

Objetivos gerais:

1. Compreender o sistema de alarmes e eventos do SAGE;



Figura 2 – Reator de 230 kV da Eletronorte (Fonte: Autor)

2. Criar um sistema de pesquisa capaz de relacionar desligamentos (caso principal de indisponibilidade) aos seus alarmes;
3. Criar um banco de dados com desligamentos e seus respectivos alarmes;
4. Aplicar os conhecimentos sobre IA no estudo desse banco de dados.

Objetivos específicos:

1. Encontrar relações entre alarmes e desligamentos com grau satisfatório de confiança;
2. Propor a manutenção mais eficiente de reatores, consoante às relações identificadas
3. Diminuir o número de eventos não programados;
4. Identificar melhorias no sistema de monitoramento da Eletrobras Eletronorte;
5. Validar o método para aplicação em outras famílias de equipamentos,



## 2 FUNDAMENTAÇÃO TEÓRICA E TRABALHOS RELACIONADOS

### 2.1 Sistema SAGE

Conforme apresentado na seção 1.1, a Eletrobras Eletronorte utiliza o software SAGE para aquisição, tratamento e distribuição de dados do sistema elétrico. O sistema utiliza um banco de dados relacional, modelado por meio de entidades e atributos, no qual as entidades se relacionam entre si.

O Serviço de Aquisição e Controle (SCD), subsistema do SAGE, é modelado por meio de ligações SCADA entre entidades. A telemedição do SCADA abrange o sistema elétrico da Eletrobras Eletronorte, interligações com demais agentes do sistema elétrico de potência e outros sistemas relevantes desses agentes. O conjunto desses sistemas é a Rede de Análise do SCD.

A Figura 3 ilustra o banco de dados relacional do SAGE referente à aquisição e controle. Sobre as medições realizadas nos equipamentos, destacam-se as entidades:

- PAS (Entidade Ponto de Medição Analógica): configura os pontos analógicos e calculados analógicos da Rede de Análise;
- PTS (Entidade Ponto Totalizado Lógico): configura os pontos totalizadores e calculados analógicos da Rede de Análise;

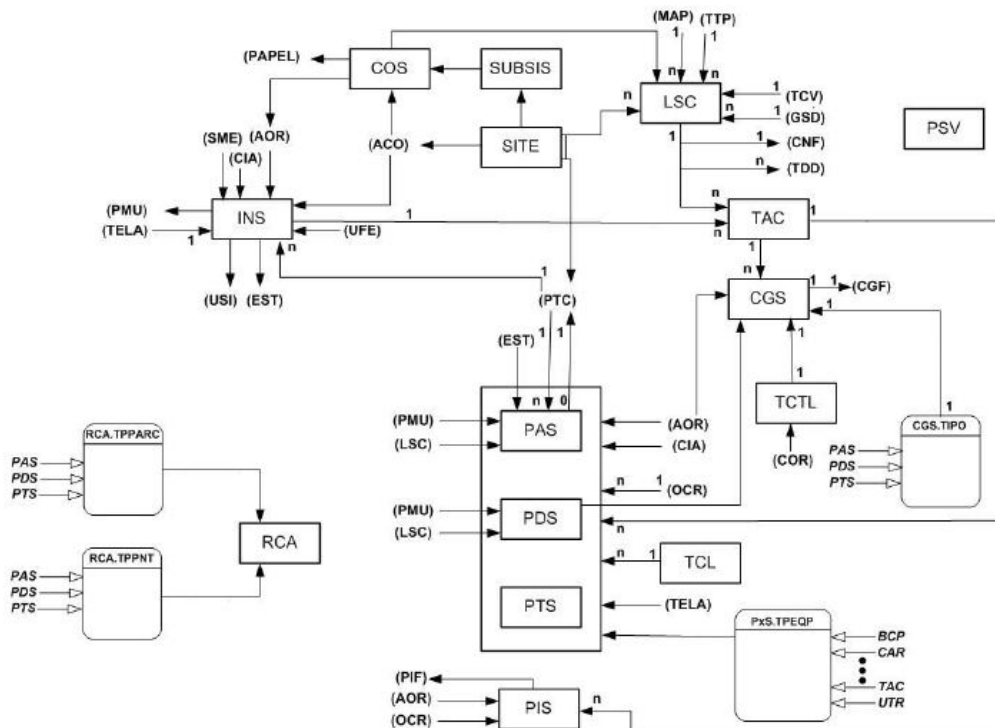


Figura 3 – Modelo de Dados da Configuração do SCD (MANUAL SAGE, 2020)

Timestamp	Narrativa	Objeto	Nome do Ponto	Sobreposições
24/04/2023 11:30:15	Operacional	ENPDC0102LIAPM01	ENPDC-01/02 Ligação Aquisição SIMRDF1	
24/04/2023 11:30:13	Operacional	ENPDC0102LIAPV01	ENPDC-01/02 Ligação Aquisição PVRDF1	
24/04/2023 11:27:31	Desbloqueado	CHBP6018/BQ	CHBP6-01 878 Bloqueio Prot Dif Barra	
24/04/2023 11:23:10	Operacional	ENWAD0102LIANG	ENWAD-01/02 Ligação Aquisição SEM	
24/04/2023 11:21:38	Transição Normal para Desligamento	DDTFC004R/FHH	DDTFC-06 204 Corrente Fuga Alta Falha	
24/04/2023 10:01:48	Aberto	PBDJ204POS	PBDJ2-04 Posição Disjuntor	
24/04/2023 10:01:33	Aberto	PBDJ203POS	PBDJ2-03 Posição Disjuntor	
24/04/2023 08:54:10	Aberto	RPD015POS	RPD08-15 Posição Disjuntor	
24/04/2023 08:36:43	Aberto	UDJ414POS	UDJ4-14 Posição Disjuntor	
24/04/2023 01:10:03	Aberto	CP_DJ1084-3_DJDXE	CPV DJ1084-3 Estado	
24/04/2023 00:53:15	Aberto	CP_DJ1124-2_DJDXE	CPV DJ1124-2 Estado	
23/04/2023 14:24:32	Aberto	CP_DJ802_DJDXE	CPV DJ802 Estado	
23/04/2023 14:21:28	Fechado	CP_DJ802_DJDXE	CPV DJ802 Estado	
20/04/2023 11:39:36	Desligamento	RBDJ404SF62	RBDJ4-04 63 Baixa Pressão SF6 2oEst	
20/04/2023 11:39:36	Bloqueado	LEQJ201	27 Bloq Prot SubTensão 13kV	
20/04/2023 11:39:35	Desligimped	IZUCD1CS1C88M	IZCS2-01 96M Bloq por Defeito Mecânico	
20/04/2023 11:39:35	Desligimped	IZUCD1CS1C88E	IZCS2-02 96E Bloq por Defeito Elétrico	
20/04/2023 11:39:35	Desligimped	IZUCD1CS1C88E	IZCS2-01 96E Bloq por Defeito Elétrico	
20/04/2023 11:39:35	Desligamento	CTACQD248PTMSA81	CT-ACQD2 48 Prot Term DiaSaída Barra81	
20/04/2023 11:39:35	Desligamento	CTACQD148PTMSA83	CT-ACQD1 48 Prot Term DiaSaída Barra83	
20/04/2023 11:39:35	Desligamento	CTACCU8P248PTM81	CT-ACCU8-P2 48 Prot Term DiaSaída Barra81	
20/04/2023 11:39:35	Desligamento	CTACCU8P148PTM83	CT-ACCU8-P1 48 Prot Term DiaSaída Barra83	
20/04/2023 11:39:35	Desligamento	CTACCU8P248PTM83	CT-ACCU8-P2 48 Prot Term DiaSaída Barra83	
20/04/2023 11:39:35	Desligamento	CTACCU8P148PTM82	CT-ACCU8-P1 48 Prot Term DiaSaída Barra82	
20/04/2023 11:39:35	Desligamento	QBC60159	QBC6-01 50 Prot Sobretensão	
20/04/2023 11:39:35	Desligimped	QBC6015051V	QBC6-01 50/61 Prot Sobrecor PV	
20/04/2023 11:39:35	Desligamento	AQSM178016SARTDO	SAHQ-LT8-01 SSA Recuperação TDO	
20/04/2023 11:39:35	Desligamento	AQSM178016SARTDO	AQDM-01 50/62BF Prot Falha Disjuntor	
20/04/2023 11:39:35	Desligamento	ANDJ6035F62	ANDJ6-03 63 Baixa Pressão SF6 2oEst	
20/04/2023 11:39:35	Normal	ANDJ6035F62	ANDJ6-03 48 Discordância de Poles	
20/04/2023 11:39:35	Normal	ANDJ6035F62	ANDJ6-03 48 Bloqueio Fechamento Disj	
20/04/2023 11:39:35	Desligamento	ADACQD348PTMSA83	AD-ACQD3 48 Prot Term DiaSaída Barra83	
20/04/2023 11:39:35	Desligamento	ADACQD248PTMSA81	AD-ACQD2 48 Prot Term DiaSaída Barra81	
20/04/2023 11:39:35	Desligamento	ADACQD448PTMSA81	AD-ACQD4 48 Prot Term DiaSaída Barra81	
20/04/2023 11:39:35	Desligamento	ADACQD148PTMSA83	AD-ACQD1 48 Prot Term DiaSaída Barra83	

Figura 4 – Tela de Alarmes/Eventos do SAGE (Fonte: autor)

lados totalizadores lógicos da Rede de Análise;

- PDS (Entidade Ponto Digital): especifica os pontos digitais e calculados digitais lógicos da Rede de Análise;
- RCA (Relação de Cálculos do SCD): especifica os relacionamentos entre as entidades PAS, PTS e PDS (pontos lógicos simples) e as entidades PAS, PTS e PDS (pontos lógicos calculados);
- OCR (Entidade Ocorrência): configura todas as ocorrências (eventos) da Rede de Análise, conforme o tipo de ponto especificado nas entidades PAS, PTS e PDS.

Para uso dos operadores do sistema elétrico de potência, outro subsistema do SAGE efetiva a organização dos dados para consulta ou exibição em uma interface gráfica (Figura 4). O Serviço de Alarmes e Eventos (SSC) é configurável e o analista de sistemas pode organizar os resultados conforme necessário, por exemplo, por ponto de medição, por ocorrência, por tipo de equipamento ou por instalação.

Na configuração da Eletrobras Eletronorte, o banco de dados possui as entidades EVE\_H (Eventos), PAS\_R e PDS\_R, as quais serão utilizadas na extração de dados.



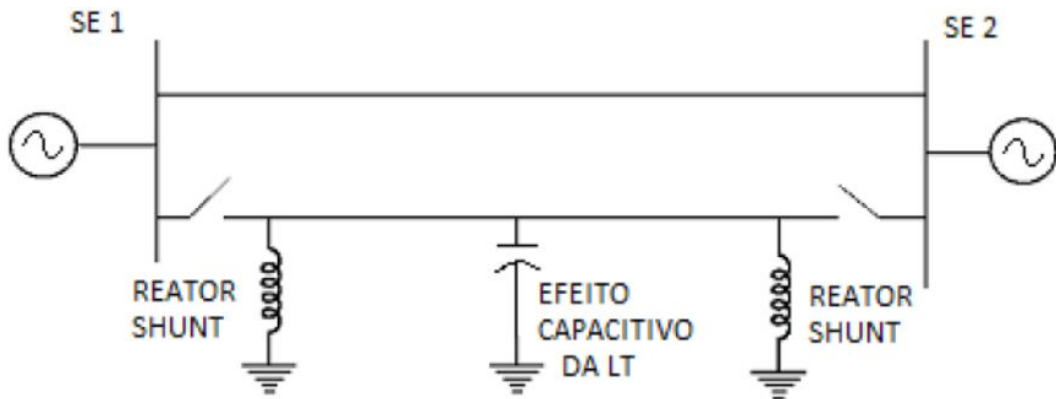


Figura 5 – Diagrama elétrico de linhas de transmissão com uso de reatores para compensação (CIGRÉ, 2023)

## 2.2 Equipamento Reator

Em um sistema elétrico de potência, a eficiente transmissão de energia depende do controle do nível de tensão, o qual varia em resposta às dinâmicas de geração e consumo de energia. Esse controle é efetuado por meio da gestão da potência reativa no sistema elétrico (Figura 5).

As linhas de transmissão atuam como condutores de campos elétricos e magnéticos, interagindo com o sistema, gerando ou absorvendo energia reativa. Um dos principais dispositivos empregados para controlar a elevação da tensão de operação da linha (conhecida como Efeito Ferranti) é o reator. Ele absorve o excesso de potência reativa capacitiva do sistema, reduzindo os níveis de tensão para valores adequados aos demais componentes do sistema elétrico de potência, além de atenuar sobretensões durante manobras (BELTRÃO, 2014).

A confiabilidade do reator, no entanto, depende de sua manutenção adequada ao longo da vida. Ainda que o tipo de manutenção seja uma decisão estratégica empresarial, a manutenção preventiva é mais benéfica para o sistema elétrico de potência, pois evita indisponibilidades em decorrência de falhas e defeitos:

- Falha: diminuição parcial da capacidade de um equipamento, suscetível a evoluir para defeito;
- Defeito: evento súbito, cujo resultado é a indisponibilidade do equipamento.

Segundo (SILVA, 2012), a manutenção preditiva preconiza que o monitoramento regular das condições reais das máquinas irá maximizar o tempo médio entre reparos, além de minimizar o número e o custo das paradas não programadas, resultantes de defeitos (Figura 6). Esse é o propósito ao minerar regras de associação.

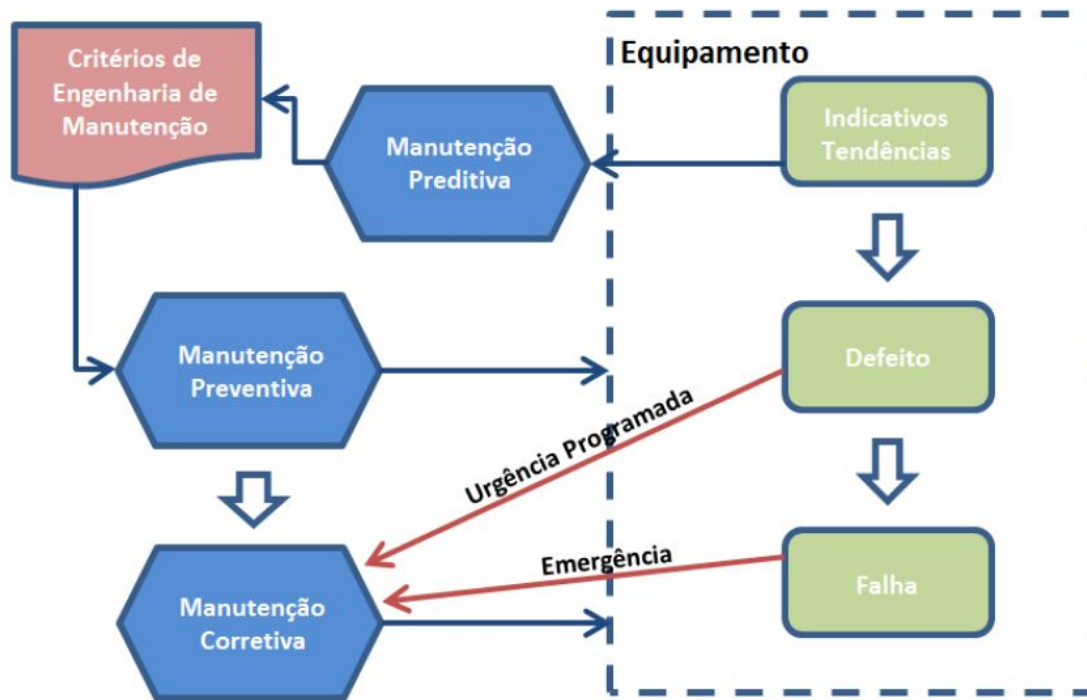


Figura 6 – Tipos de Manutenção conforme Criticidade (MARQUES, 1996)

### 2.3 Regras de Associação

Associação é uma das técnicas centrais da mineração de dados, cujo objetivo é identificar relações úteis entre os elementos de um conjunto de dados.

Regras de associação expõem correlações de difícil percepção, ampliando as possibilidades para tomada de decisões (AL-ZAWAIDAH; JBARA; ABU-ZANONA, 2019).

O projeto utiliza o modelo matemático proposto por (AGRAWAL; IMIELINSKI; SWAMI, 1993) o qual tratou de vendas em um mercado.

Seja  $Z = I_1, I_2, \dots, I_n$  um conjunto de produtos, atributos ou alarmes. Para o modelo,  $Z$  é um conjunto de itens.

Seja  $T$  um conjunto de observações ou registros dos relacionamentos entre os itens de  $Z$  que ocorrem juntos num determinado momento ou contexto. Cada registro no conjunto  $T$  é chamado de transação.

Sejam  $X$  e  $Y$  dois subconjuntos de itens contidos em  $Z$ , porém excludentes. Ou seja,  $X \subseteq Z$ ,  $Y \subseteq Z$  e  $X \cap Y = \emptyset$ . Além disso, o conjunto  $T$  inclui transações envolvendo  $X$  e  $Y$ , logo  $X \subseteq T$  e  $Y \subseteq T$ .

Dadas essas condições, uma regra de associação é a implicação na forma  $X \Rightarrow Y$  (seja o conjunto de transações  $T$ , a ocorrência de  $X$  implica na ocorrência de  $Y$  com determinado percentual de confiança).

Nem todas as regras de associação, porém, possuem utilidade para compreensão dos



dados. Dessa forma, o modelo de (AGRAWAL; IMIELINSKI; SWAMI, 1993) estabeleceu outras condições na avaliação das regras de associação.

1. Condições sintáticas: o antecedente  $X$  deve conter determinado item  $I_k$ , ou o conseqüente  $Y$  deve conter o item  $I_k$ , ou simultaneamente  $X$  deve conter o item  $I_k$  enquanto  $Y$  deve conter o item  $I_j$ ;
2. Condição de suporte: a regra deve ter uma frequência mínima de recorrências no conjunto de transações  $T$ . Ou seja, a regra  $X \Rightarrow Y$  é aceita se um número suficiente de registros em  $T$  contém  $X \cup Y$ .

Estabelecidas as premissas, o modelo então calcula as regras de associação em duas etapas.

1. identificar todos os subconjuntos de  $Z$  que atendem à condição de suporte. Esses subconjuntos são chamados de *itemsets frequentes*;
2. identificar todas as regras de associação envolvendo *itemsets frequentes* e com percentual mínimo de confiança.

Há limitações nos algoritmos que encontram regras de associação. Uma delas é a escalabilidade, pois a geração de todas as combinações possíveis de *itemsets frequentes* pode ser demorada e exigir recursos computacionais avançados. Outra desvantagem é o grande número de regras obtidas, dificultando a compreensão dos resultados e, por consequência, diminuindo sua utilidade. Por último, não é simples estabelecer os parâmetros adequados para o algoritmo, como suporte mínimo, sendo necessárias tentativas de ajuste e, em cada uma delas, as desvantagens anteriores se repetem (AL-ZAWAIDAH; JBARA; ABU-ZANONA, 2019).

## 2.4 Algoritmo Apriori

O algoritmo Apriori, pioneiro na área de mineração de conjuntos de itens e regras de associação, foi introduzido por (AGRAWAL; SRIKANT, 1994). Com a evolução dos recursos computacionais, e aprimoramentos do próprio algoritmo, tem-se mostrado eficiente para lidar com a escalabilidade.

Conforme estabelecido na seção 2.3, a primeira etapa identifica os *itemsets frequentes* ((AGRAWAL; IMIELINSKI; SWAMI, 1993)).

Algoritmos para identificação dos *itemsets frequentes* realizam múltiplas varreduras nos dados. Na primeira varredura, identificam-se os *itemsets frequentes* formados por conjuntos individuais. Em cada varredura posterior, os conjuntos encontrados com

sucesso na etapa anterior são tratados como *seeds*, ou seja, são usados na composição de novos conjuntos de itens potencialmente frequentes, os quais serão avaliados segundo a condição de suporte mínimo. O processo se repete até nenhum novo *itemset frequente* ser encontrado.

O algoritmo Apriori também realiza essa sequência de varreduras, porém apresenta um novo princípio ao processo: qualquer subconjunto de um *itemset frequente* também deve ser um *itemset frequente*. Dessa forma, cada varredura avalia somente candidatos formados exclusivamente pela combinação de *seeds*, descartando candidatos com subconjuntos que não sejam frequentes. Esse procedimento resulta na geração de um número muito menor de conjuntos candidatos (AGRAWAL; SRIKANT, 1994).

O algoritmo Apriori foi aplicado por meio da biblioteca MLxtend. Para avaliar a eficácia de uma regra identificada, na forma  $X \Rightarrow Y$  definida na seção 2.3, a biblioteca oferece as seguintes métricas (RASCHKA, 2018):

1. Suporte (Support): a medida avalia a frequência relativa de uma regra  $X \Rightarrow Y$  em um determinado conjunto de dados.

$$\text{suporte}(X \Rightarrow Y) = \frac{|X \cup Y|}{|T|}, \quad \text{no intervalo } [0, 1]$$

2. Confiança (Confidence): a medida avalia a probabilidade de  $Y$  em uma transação que contenha  $X$ .

$$\text{confiança}(X \Rightarrow Y) = \frac{\text{suporte}(X \Rightarrow Y)}{\text{suporte}(X)}, \quad \text{no intervalo } [0, 1]$$

Nota-se que a medida não é simétrica, pois  $\text{confiança}(X \Rightarrow Y)$  é diferente de  $\text{confiança}(Y \Rightarrow X)$ .

3. Interesse (Lift): utilizada para encontrar dependências, a medida avalia o quão mais frequente torna-se  $Y$  quando  $X$  ocorre (caso  $Y$  e  $X$  sejam independentes, o Lift é igual a 1).

$$\text{interesse}(X \Rightarrow Y) = \frac{\text{confiança}(X \Rightarrow Y)}{\text{suporte}(Y)}, \quad \text{no intervalo } [0, \infty]$$

Uma limitação da medida de Confiança reside na tendência de subestimar a importância de uma associação. Por exemplo, em uma associação  $X \Rightarrow Y$ , a pontuação de Confiança leva em consideração predominantemente  $X$ , deixando de considerar a relevância de  $Y$ . Se  $Y$  for igualmente crucial, isso pode aumentar a probabilidade de que um padrão que contenha  $X$  também contenha  $Y$ , o que pode inflar a medida de Confiança. A métrica

Interesse é uma solução para esse problema, uma vez que avalia a força da associação entre  $X$  e  $Y$  de maneira mais equilibrada (RODRIGUES *et al.*, 2019).

A biblioteca MLxtend (RASCHKA, 2018) oferece outras três medidas, não mencionadas nos trabalhos observados.

5. Zhangs Metric (YAN; ZHANG; ZHANG, 2009): a medida combina Suporte, Confiança e Interesse para avaliar a qualidade de uma regra de associação.

$$\text{zhangs metric}(X \Rightarrow Y) = \frac{\text{confiança}(X \Rightarrow Y) - \text{confiança}(X' \Rightarrow Y)}{\text{Máx}[\text{confiança}(X \Rightarrow Y), \text{confiança}(X' \Rightarrow Y)]}$$

sendo  $\text{zhangs metric}(X \Rightarrow Y)$  no intervalo  $[-1, 1]$

6. Leverage: a medida avalia a divergência entre a ocorrência da regra de associação e a ocorrência aleatória de  $Y$  e  $X$ .

$$\text{leverage}(X \Rightarrow Y) = \text{suporte}(X \Rightarrow Y) - \text{suporte}(X) \times \text{suporte}(Y)$$

sendo  $\text{leverage}(X \Rightarrow Y)$  no intervalo  $[-1, 1]$

7. Conviction: a medida avalia a independência entre  $X$  e  $Y$ . No caso ideal, quando há total independência, o resultado é igual a 1, assim como no Interesse. Por outro lado, se a regra  $X \Rightarrow Y$  for perfeita (Confiança igual a 1), a medida Conviction tende ao infinito.

$$\text{conviction}(X \Rightarrow Y) = \frac{1 - \text{suporte}(Y)}{1 - \text{confiança}(X \Rightarrow Y)}, \quad \text{no intervalo } [0, \infty]$$

## 2.5 Processo KDD

Segundo (FAYYAD; PIATETSKY-SHAPIRO; SMYTH, 1996), a mineração de dados é somente uma etapa de um processo maior chamado KDD (extração de conhecimento em bancos de dados, ou *knowledge-discovery in databases*). Dessa forma, a extração do conhecimento útil envolve etapas adicionais de seleção, limpeza e preparação dos dados, incorporação de conhecimento prévio e interpretação adequada dos resultados (Figura 7).

Se aplicado isoladamente, a mineração de dados possivelmente resultará em relações sem significado e inválidas (FAYYAD; PIATETSKY-SHAPIRO; SMYTH, 1996).

## 2.6 Trabalhos Relacionados

O método não supervisionado de minerar regras de associação foi proposto por (AGRAWAL; IMIELINSKI; SWAMI, 1993), inicialmente aplicado ao banco de dados de um supermercado para embasar estratégias de vendas.

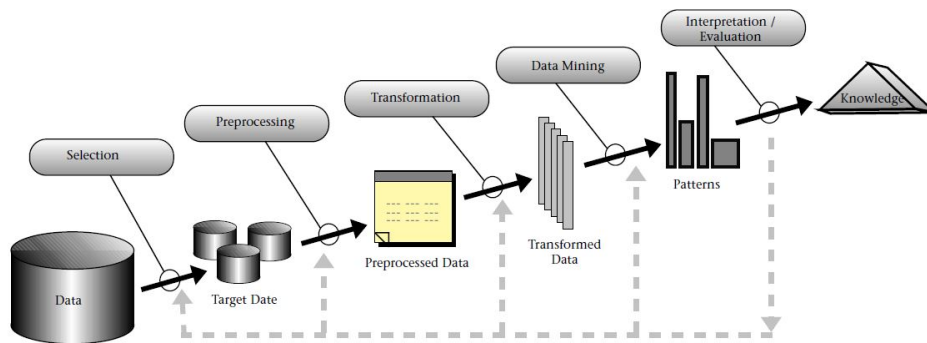


Figura 7 – Visão geral das etapas do processo KDD (FAYYAD; PIATETSKY-SHAPIRO; SMYTH, 1996)

Sobre a predição de desligamentos, (VASCONCELOS, 2020) propôs um modelo híbrido de predição, a partir de padrões extraídos das sequências de alarmes geradas por motores de uma usina termoeletrica. Destaca-se que o trabalho analisa as consequências do disparo excessivo de alarmes, ou *inundação de alarmes*, seja por falha de configuração ou de manutenção apropriada.

Outro estudo sobre manutenção preventiva foi realizado por (SILVA, 2012), o qual utiliza outra técnica de mineração de dados, SVM (Support Vector Machine), para detecção e diagnóstico de falhas mecânicas e elétricas (desalinhamento, desbalanceamento, folga mecânica, curto-circuito entre espiras do estator, desequilíbrio de fase e barras quebradas) em motores de indução trifásicos. Também apresenta um comparativo entre redes neurais artificiais, lógica fuzzy e SVM na classificação correta de falhas e defeitos.

(RODRIGUES *et al.*, 2019) discute o custo de analisar as regras de associação após a mineração de dados e apresenta um novo algoritmo para solucionar o problema. Nesse trabalho as medidas usadas na avaliação de regras são Suporte, Confiança e Interesse.

Por fim, observou-se a metodologia apresentada em (ROMÃO *et al.*, 1999) para organização e apresentação dos dados.

### 3 APLICAÇÃO DE REGRAS DE ASSOCIAÇÃO NA PREDIÇÃO DE DESLIGAMENTOS DE REATORES DA ELETROBRAS ELETRONORTE

Conforme abordagem KDD (seção 2.5) o desenvolvimento do trabalho pode ser compreendido em etapas. A primeira, extração de dados, envolve compreender o banco de dados e selecionar uma amostra adequada. A segunda etapa, pré-processamento, trata da limpeza e preparação dos dados. A terceira etapa, mineração de dados, é a aplicação da técnica de regras de associação. Por fim, a quarta etapa é a interpretação dos resultados.

#### 3.1 Extração de dados

O projeto avaliou registros do banco de dados SAGE, no intervalo de 1º de janeiro de 2022 a 30 de junho de 2023, organizados em três entidades:

- PAS\_R: medidas analógicas dos pontos supervisionados (Figuras 8 e 9);
- PDS\_R: medidas digitais (alarmes) dos pontos supervisionados (Figuras 10 e 11);
- EVE\_H: conjunto de medidas analógicas e digitais nos pontos supervisionados, mais as ocorrências do sistema elétrico associadas aos pontos (Figuras 12 e 13).

A entidade EVE\_H registra as medições nos pontos, classificadas por severidade, enquanto as entidades PAS\_R e PDS\_R descrevem os pontos. Dessa maneira, as três entidades são complementares.

O resultado é um conjunto de transações envolvendo reatores, com suas medidas analógicas e digitais, balanceadas entre transações com desligamento e sem desligamento.

A sequência de etapas adotada e comandos SQL empregados na extração dos dados são descritos no Apêndice A.

#### 3.2 Pré-Processamento

Os dados extraídos do SAGE foram organizados em uma só tabela, ou dataframe, na qual os pontos de medida foram convertidos em atributos binários de cada transação. Também foi acrescentado um atributo *Target* para identificar as transações em que houve desligamento de reatores.

O dataframe resultante, adequado para a mineração de dados, possui as seguintes características:

- Transações: 259, sendo 130 com desligamento;

column_name name	data_type character varying	column_name name	data_type character varying
bh_dthr_ini	timestamp with time zo...	histlin	smallint
bh_dthr_fim	timestamp with time zo...	histslc	smallint
bh_chave	integer	id	character
bh_indtr	integer	ins	character
bh_mrid	character	pmu	character
compdev	real	tac	character
descr	character varying	tpeq	smallint
eqp	character	unidade	character
excdev	real	ocr	character

Figura 8 – Atributos da Entidade PAS\_R (Fonte: Autor)

bh_dthr_ini timestamp with time zone	bh_dthr_fim timestamp with time zone	id character (32)	descr character varying (63)
2021-12-10 12:31:59-03	2023-04-20 11:36:48-03	SBRTSBRE1	POT. REATIVA SBRTSBRE1
2021-12-10 12:31:59-03	2023-04-20 11:36:48-03	RNRE7-RT04RN_Q	POT. REATIVA RNRE7-RT04RN
2021-12-10 12:31:59-03	2023-04-20 11:36:48-03	RNRE7-RTRBRN_Q	POT. REATIVA RNRE7-RTRBRN
2021-12-10 12:31:59-03	2022-09-14 16:45:12-03	BPRE601TMBMBITSFB6	BPRE6-01 BITS Alarmes FB 230kV
2021-12-10 12:31:59-03	2022-09-14 16:45:12-03	BPRE601TMBMFB1003	BPRE6-01 BM FA 230kV Tend Tan Delta
2021-12-10 12:31:59-03	2022-09-14 16:45:12-03	BPRE601TMBMFB1006	BPRE6-01 BM FA 230kV Tempo TDelta Alta
2021-12-10 12:31:59-03	2022-09-14 16:45:12-03	BPRE601TMBMFB1025	BPRE6-01 BM FV 230kV Tend Tan Delta
2021-12-10 12:31:59-03	2022-09-14 16:45:12-03	BPRE601TMBMFB1028	BPRE6-01 BM FV 230kV Tempo TDelta Alta
2021-12-10 12:31:59-03	2022-09-14 16:45:12-03	BPRE601TMBMFB1034	BPRE6-01 BM 230kV Ângulo da Somatória
2021-12-10 12:31:59-03	2022-09-14 16:45:12-03	BPRE601TMBMMCFA6	BPRE6-01 Med Capacitância FA 230kV

Figura 9 – Amostra da Entidade PAS\_R (Fonte: Autor)

column_name name	data_type character varying	column_name name	data_type character varying
bh_dthr_ini	timestamp with time zo...	histslc	smallint
bh_dthr_fim	timestamp with time zo...	histslc_soe	smallint
bh_chave	integer	id	character
bh_indtr	integer	ins	character
bh_mrid	character	pmu	character
descr	character varying	tac	character
eqp	character	tpeq	smallint
histform	smallint	ocr	character

Figura 10 – Atributos da Entidade PDS\_R (Fonte: Autor)

bh_dthr_ini timestamp with time zone	bh_dthr_fim timestamp with time zone	id character (32)	descr character varying (63)
2021-12-10 12:32:30-03	[null]	PDPDRE7117TOV	PDRE7-11 26 SobreTemp Óleo 1oEst FV
2021-12-10 12:32:30-03	[null]	IZSR715BPOS	IZSR7-15 Posição Seccionadora Fase B
2021-12-10 12:32:30-03	[null]	ATDJ609TM	ATDJ6-09 Térmico/Def Motor
2022-10-05 16:17:37-03	[null]	MB3SR705SELC	MBSR7-05 Seleção p/ Comando
2021-12-10 12:32:30-03	[null]	PDRE703E0492	PDRE7-03 49 SobreTemp Enrol 2oEst
2021-12-10 12:32:30-03	[null]	PDSR704E00FT	PDSR7-04 Falta Tensão/Térmico
2021-12-10 12:32:30-03	[null]	IZDJ712CBQSC	IZDJ7-12 Desbloq Abertura Seccs p/50BF
2021-12-10 12:32:30-03	[null]	MBRE70726V	MBRE7-07 26 SobreTemp Óleo FV 1oEst
2021-12-10 12:32:30-03	[null]	ATDJ60974B1A	ATDJ6-09 74 Superv Bob 1 Abertura
2021-12-10 12:32:30-03	[null]	ATDJ60974B2A	ATDJ6-09 74 Superv Bob 2 Abertura

Figura 11 – Amostra da Entidade PDS\_R (Fonte: Autor)

column_name name	data_type character varying	column_name name	data_type character varying
bh_dthr	timestamp with time zone	ocr	character
bh_variacao	integer	severidade	character
mrid	character varying	texto	character varying
obj	character	tpobj	character

Figura 12 – Atributos da Entidade EVE\_H (Fonte: Autor)

bh_dthr timestamp with time zone	bh_variacao integer	mrid character varying (55)	obj character (50)	ocr character (30)	severidade character (30)	texto character varying (152)	tpobj character (15)
2023-05-22 20:59:59-03	2	_81de79d4_98cd_486...	JRQSACCRET1	OcrAnaFaixa01	K_SEV_NORML	Retornou a regiao normal (-2.00)	pas
2023-05-22 20:59:59-03	2	_81de79d4_98cd_486...	JRQSACCRET1	OcrAnaFaixa01	K_SEV_NORML	Retornou a regiao normal (-2.00)	pas
2023-05-22 20:59:57-03	2	_4807bc45_bd6b_f91...	CX6AT4VDF2FB	OcrUrg1Alarm...	K_SEV_NORML	Normal	pds
2023-05-22 20:59:57-03	2	_4a6c0f20_0a5a_651...	SSTF603MBCHAFI	OcrUrg0Alarm...	K_SEV_NORML	Normal	pds
2023-05-22 20:59:57-03	2	_4807bc45_bd6b_f91...	CX6AT4VDF2FB	OcrUrg1Alarm...	K_SEV_NORML	Normal	pds
2023-05-22 20:59:57-03	2	_4a6c0f20_0a5a_651...	SSTF603MBCHAFI	OcrUrg0Alarm...	K_SEV_NORML	Normal	pds

Figura 13 – Amostra da Entidade EVE\_H (Fonte: Autor)

- Atributos: 62 pontos de medição analógica ou digital, mais *Target*;
- Valores: 0 ou 1, sendo 1 o registro de uma medida, analógica ou digital, de Severidade igual a Fatal ou, para o atributo *Target*, o registro de que naquela transação houve desligamento;
- Frequência: a Figura 14 apresenta as medidas mais frequentes no intervalo de estudo.

O algoritmo Python utilizado no pré-processamento (o qual inclui correção de erros identificados na base) está disponível no Apêndice B

### 3.3 Mineração de dados

Conforme seção 2.4, a mineração utilizou a biblioteca MLxtend, especificamente duas de suas funções: *apriori* (Figura 15), para identificar os *itemsets frequentes*, e



	Nº de Medidas		Nº de Medidas
Corrente Fase V	30	Temperatura Enrol Fase A	15
Corrente Fase B	29	Temperatura Enrol Fase B	14
Corrente Fase A	27	Temperatura de oleo Fase A	13
Potência Reativa	26	Temperatura Enrol Fase R	13
Tensão entre Fases BV	21	Temperatura Enrol	12
Tensão entre Fases VA	21	Temperatura de oleo Fase V	12
Frequência	20	Temperatura de oleo Fase R	11
Tensão entre Fases AB	19	Tensão Fase A	10
Temperatura de oleo Fase B	17	Potência Ativa	10
Temperatura Enrol Fase V	16	Tensão Fase V	10

Figura 14 – Contagem dos Atributos mais Frequentes (Fonte: Autor)

*association\_rules* (Figura 16), para identificar as regras de associação.

```
itemfreq = apriori(transacoes, min_support = 0.2, use_colnames = True)
itemfreq.sort_values(by='support', ascending=False)
```

	support	itemsets
0	0.521236	( Corrente Fase A)
3	0.513514	( Frequência)
14	0.501931	(target)
12	0.416988	( Tensão entre Fases BV)
13	0.378378	( Tensão entre Fases VA)
...	...	...
62	0.200772	( Temperatura de oleo Fase B, Temperatura...
10	0.200772	( Temperatura de oleo Fase V)
64	0.200772	( Temperatura de oleo Fase B, Temperatura...
40	0.200772	( Tensão entre Fases AB, Frequência)
66	0.200772	( Temperatura de oleo Fase B, Corrente Fa...

67 rows × 2 columns

Figura 15 – Exemplo de uso da função *apriori*: Suporte mínimo ajustado em 20% (Fonte: Autor)

Ainda que a biblioteca MLxtend possua diversas métricas para avaliação de eficácia, a função *association\_rules* permite somente uma métrica, como parâmetro de corte, por execução. Dessa forma, duas outras condições foram adicionadas após as funções MLxtend.

Algoritmo de mineração de dados:



```
rules = association_rules(itemfreq, metric = "confidence", min_threshold = 0.5)
rules = rules[['antecedents', 'consequents', 'support', 'confidence', 'lift']]
rules
```

	antecedents	consequents	support	confidence	lift
0	( Corrente Fase B)	( Corrente Fase A)	0.250965	0.764706	1.467102
1	( Corrente Fase V)	( Corrente Fase A)	0.247104	0.752941	1.444532
2	( Corrente Fase A)	( Frequência)	0.347490	0.666667	1.298246
3	( Frequência)	( Corrente Fase A)	0.347490	0.676692	1.298246
4	( Potência Reativa)	( Corrente Fase A)	0.235521	0.782051	1.500380
...	...	...	...	...	...
136	( Temperatura Enrol Fase R, Temperatura d...	( Temperatura de óleo Fase B, Corrente Fa...	0.200772	0.852459	4.014307
137	( Temperatura de óleo Fase B)	( Corrente Fase B, Temperatura Enrol Fase...	0.200772	0.838710	3.949560
138	( Corrente Fase B)	( Temperatura de óleo Fase B, Temperatura...	0.200772	0.611765	2.731846
139	( Temperatura Enrol Fase R)	( Temperatura de óleo Fase B, Corrente Fa...	0.200772	0.812500	3.970519
140	( Temperatura de óleo Fase R)	( Temperatura de óleo Fase B, Corrente Fa...	0.200772	0.838710	4.098600

141 rows x 5 columns

Figura 16 – Exemplo de uso da função *association\_rules*: Confiança mínima ajustada em 50% para *itemsets frequentes* com Suporte mínimo ajustado em 20% (Fonte: Autor)

1. Aplicação da *apriori* (com aplicação inerente da métrica Suporte);
2. Aplicação da *association\_rules* com adoção da métrica Confiança;
3. Seleção exclusiva de regras com o Atributo *Target* contido no Consequente;
4. Aplicação adicional da métrica Interesse (Lift) como parâmetro de corte.

Uma vez estabelecido, o algoritmo foi submetido a iterações para ajuste dos parâmetros, conforme Tabela 1

Suporte mínimo	Nº <i>itemsets frequentes</i>	Confiança mínima	Interesse mínimo	Nº regras de associação
0,2	67	0,5	2,0	0
0,2	67	0,5	1,0	1
0,1	6049	0,5	1,0	127
0,1	6049	0,5	2,0	56
0,1	6049	0,55	2,0	6

Tabela 1 – Parâmetros e resultados a cada aplicação do algoritmo (Fonte: autor)

O resultado do método de mineração de dados é apresentado na Tabela 2.

### 3.4 Resultados

As regras de associação encontradas, ainda que obtidas pelo melhor uso das técnicas disponíveis, não apresentam força suficiente para explicar os defeitos dos reatores.

Conforme seção 3.3, o algoritmo encontrou uma regra com Suporte acima de 20%, porém foi descartada por sua medida Interesse (Tabela 3) próxima do valor 1, indicando

<b>Antecedente</b>	<b>Consequente</b>	<b>Suporte</b>	<b>Confiança</b>	<b>Interesse</b>
Temperatura de óleo Fase A, Corrente Fase V	Target, Corrente Fase B	0,1004	0,5532	3,0484
Temperatura de óleo Fase A, Corrente Fase B	Target, Corrente Fase V	0,1004	0,5532	3,2563
Temperatura de óleo Fase A, Corrente Fase B	Target, Temperatura Enrol Fase R	0,1042	0,5745	4,3761
Temperatura de óleo Fase A, Corrente Fase B	Temperatura de óleo Fase R, Target	0,1004	0,5532	4,3417
Temperatura de óleo Fase A, Corrente Fase V	Target, Temperatura Enrol Fase R	0,1042	0,5745	4,3761
Temperatura de óleo Fase A, Corrente Fase V	Temperatura de óleo Fase R, Target	0,1004	0,5532	4,3417

Tabela 2 – Regras de Associação (Fonte: autor)

que os *itemsets frequentes* classificados de Antecedente e Consequente são, na verdade, independentes.

<b>Antecedente</b>	<b>Consequente</b>	<b>Suporte</b>	<b>Confiança</b>	<b>Interesse</b>
Tensão entre Fases VA	Target	0,2008	0,5306	1,0571

Tabela 3 – Regra de Associação descartada (Fonte: autor)

Quanto ao resultado do projeto, Tabela 2, as regras de associação apresentam pouco Suporte (os relacionamentos aparecem em apenas 10% das 259 transações observadas) e Confiança pouco acima de 50% (na melhor delas, a Confiança é 57,45%).

No entanto, as regras apuradas mostram indícios para a Manutenção Preventiva. As medições analógicas de temperatura no óleo, especificamente na Fase A, precederam todos os desligamentos nas regras apuradas. É preciso investigar, no entanto, se é verdadeira a maior influência da Fase A nesses equipamentos (a base pesquisada inclui números semelhantes de medições de temperatura em todas as Fases).

Há medições de temperatura, de óleo e de enrolamento, desassociadas de fases. Podem ser uma metodologia diferente para monitorar reatores específicos ou podem ser erros na base de dados. Considerando o peso das medidas de temperatura nas regras encontradas, o esclarecimento deve alterar os resultados do estudo, o que levaria a uma possível revisão em breve.

## 4 CONCLUSÃO

O objetivo do projeto foi analisar o uso de regras de associação na identificação prévia de falhas em reatores. Em seu desenvolvimento, foram cumpridos os objetivos gerais de aprendizado a respeito da base de dados e ferramentas de IA:

- Compreender o sistema de alarmes e eventos do SAGE: seção 2.1;
- Criar um sistema de pesquisa capaz de relacionar desligamentos (caso principal de indisponibilidade) aos seus alarmes: seção 3.1;
- Criar um banco de dados com desligamentos e seus respectivos alarmes: seção 3.2;
- Aplicar os conhecimentos sobre IA no estudo desse banco de dados: capítulo 3, em especial a seção 3.3 (Mineração de dados).

O resultado do projeto atende parcialmente os objetivos específicos:

- Encontrar relações entre alarmes e desligamentos com grau satisfatório de confiança;
- Propor a manutenção mais eficiente de reatores, consoante às relações identificadas;
- Diminuir o número de eventos não programados.

Conforme apresentado na seção 3.4, há indícios de um fator predominante que, se confirmado, irá aprimorar a Manutenção Preditiva, reduzindo as indisponibilidades do sistema e custos associados.

- Identificar melhorias no sistema de monitoramento da Eletrobras Eletronorte.

As etapas de Extração (seção 3.1) e Pré-Processamento (seção 3.2) produziram um sistema prático de consulta de eventos e seus respectivos alarmes, assim como identificaram erros na base de aquisição e controle do SAGE.

- Validar o método para aplicação em outras famílias de equipamentos.

Comprovou-se a validade do método. O uso de regras de associação, dentre as técnicas de mineração de dados disponíveis, foi essencialmente a opção por um método versátil, de fácil aplicação para conjuntos diferentes de equipamentos e medições. O segmento de energia elétrica está em contínua expansão e utiliza um vasto número de famílias de equipamentos, portanto é imprescindível um sistema dinâmico de gestão.

Como trabalho futuro, podem ser aprofundados os estudos e extrações de dados da base do SAGE. É possível que dados adicionais estejam armazenados em outro servidor e, dessa forma, com uma base ampliada de transações para estudo, seria possível fortalecer os resultados.

Outro aspecto para aprofundamento é a seleção de medidas. Na ausência de um critério abrangente, capaz de padronizar os diferentes reatores da Eletrobras Eletronorte, com suas características operacionais próprias, o estudo considerou as medidas analógicas apenas sob o aspecto de severidade máxima, aplicado pelo SAGE (uma medida classificada como Fatal recebeu o valor 1, enquanto severidades menores foram descartadas). É necessário analisar as medições descartadas, pois é possível inferir que a identificação de defeitos antes de uma medição Fatal irá ampliar as possibilidades de Manutenção.

A pesquisa de trabalhos relacionados encontrou projetos mais abrangentes de predição de falhas, porém utilizando técnicas diferentes, como deep learning. É possível que um trabalho futuro exija essas abordagens.

## REFERÊNCIAS

- AGÊNCIA NACIONAL DE ENERGIA ELÉTRICA – ANEEL. Regras dos serviços de transmissão de energia elétrica. Módulo 4: Prestação dos Serviços, 2022. Available at: [https://www2.aneel.gov.br/cedoc/aren2020905\\_2\\_3.pdf](https://www2.aneel.gov.br/cedoc/aren2020905_2_3.pdf). Access at: 21 out. 2023.
- AGRAWAL, R.; IMIELINSKI, T.; SWAMI, A. N. Mining association rules between sets of items in large databases. *In: SIGMOD/PODS93*, Washington D.C. 1993. p. 207–216. Available at: <https://dl.acm.org/doi/10.1145/170035.170072>. Access at: 16 mar. 2023.
- AGRAWAL, R.; SRIKANT, R. Fast algorithms for mining association rules. *In: PROC. 20TH INT. CONF. VERY LARGE DATA BASES, VLDB*. 1994. v. 1215, p. 487–499. Available at: [https://www.it.uu.se/edu/course/homepage/infoutv/ht08/vldb94\\_rj.pdf](https://www.it.uu.se/edu/course/homepage/infoutv/ht08/vldb94_rj.pdf). Access at: 17 out. 2023.
- AL-ZAWAIDAH, F. H.; JBARA, Y. H.; ABU-ZANONA, M. A. An improved algorithm for mining association rules in large databases. **World of Computer Science and Information Technology Journal (WCSIT)**, v. 11, n. 3, p. 158–177, 2019. ISSN 2175-7275. Available at: <http://www.ojs.unirg.edu.br/index.php/1/article/view/2788/1541>. Access at: 17 set. 2023.
- BASE Fonte SAGE - Manual de Configuração. [S.l.], 2020.
- BELTRÃO, V. **Metodologia de Gestão do Ciclo de Vida de Reatores**. 2014. 197 p. Dissertação (Mestrado em Engenharia Elétrica) — Instituto de Tecnologia, Universidade Federal do Pará, Belém, 2014. Available at: <https://repositorio.ufpa.br/jspui/handle/2011/6286>. Access at: 17 set. 2023.
- CIGRÉ. **Brochura 16 - Guia de Manutenção para Transformadores de Potência**. [S.l.], 2023. Available at: <https://cigre.org.br/brochuras/>. Access at: 2023.09.17.
- ELETRONORTE Eletronorte: História. 2023. Available at: <https://www.eletronorte.com.br/historia>. Access at: 16 jul. 2023.
- FAYYAD, U.; PIATETSKY-SHAPIO, G.; SMYTH, P. From data mining to knowledge discovery in databases. **AI Magazine**, v. 17, n. 3, 1996. Available at: <https://doi.org/10.1609/aimag.v17i3.1230>. Access at: 25 nov. 2023.
- MARQUES, A. Apostila. **Tópicos de Subestações**. Goiânia: [S.l.: s.n.], 1996.
- PEREIRA, L. A. C. *et al.* Sage - um sistema aberto para a evolução. 2018. Available at: <https://sage.cepel.br/index.php/pt/documentacao/artigos/download/23-artigos/157-sage-um-sistema-aberto-para-a-evolucao>. Access at: 05 mar. 2023.
- RASCHKA, S. Mlxtend: Providing machine learning and data science utilities and extensions to python’s scientific computing stack. **The Journal of Open Source Software**, The Open Journal, v. 3, n. 24, Apr 2018. Available at: <https://joss.theoj.org/papers/10.21105/joss.00638>. Access at: 20 out. 2023.

RODRIGUES, D. C. *et al.* Proposta de método para redução do conjunto de regras de associação resultantes do algoritmo apriori. **Revista Cereus**, v. 11, n. 3, p. 158–177, 2019. ISSN 2175-7275. Available at: <http://www.ojs.unirg.edu.br/index.php/1/article/view/2788/1541>. Access at: 17 set. 2023.

ROMÃO, W. *et al.* Extração de regras de associação em c&t: O algoritmo apriori. *In: XIX ENCONTRO NACIONAL EM ENGENHARIA DE PRODUÇÃO*. 1999. Available at: [https://www.researchgate.net/publication/268338920\\_EXTRACAO\\_DE\\_REGRAS\\_DE\\_ASSOCIACAO\\_EM\\_CT\\_O\\_ALGORITMO\\_APRIORI](https://www.researchgate.net/publication/268338920_EXTRACAO_DE_REGRAS_DE_ASSOCIACAO_EM_CT_O_ALGORITMO_APRIORI). Access at: 17 set. 2023.

SILVA, V. A. D. **Detecção de falhas em motores elétricos através das máquinas de vetores de suporte**. 2012. 117 p. Dissertação (Mestrado em Engenharia Mecânica) — Universidade Estadual de Campinas, SP, 2012. Available at: <https://doi.org/10.47749/T/UNICAMP.2012.847459>. Access at: 16 set. 2023.

VASCONCELOS, B. R. A. **Predição de desligamentos de motores de uma usina termoeletrica baseada no histórico de eventos**. 2020. 127 p. Dissertação (Mestrado em Ciência da Computação) — Universidade Federal de Campina Grande, Paraíba, 2020. Available at: <http://dspace.sti.ufcg.edu.br:8080/jspui/handle/riufcg/18429>. Access at: 15 out. 2023.

YAN, X.; ZHANG, C.; ZHANG, S. Confidence metrics for association rule mining. **Applied Artificial Intelligence**, v. 23, n. 8, p. 713–737, Apr. 05 2009. Available at: <https://doi.org/10.1080/08839510903208062>. Access at: 20 out. 2023.

## APÊNDICE A – EXTRAÇÃO DE DADOS DO SAGE

A ferramenta de consulta e extração de dados do SAGE foi o SQL. Em razão do custo computacional e da necessidade de evitar erros, o código foi fracionado para execuções parciais.

1. Selecionar, na entidade EVE\_H, todos os desligamentos de reatores no intervalo de janeiro de 2022 até junho de 2023. Essa pesquisa define os eventos do estudo.

```
CREATE TEMPORARY TABLE eventos AS (
    SELECT date_trunc('minute', bh_dthr) AS minuto, obj
    FROM eve_h
    WHERE bh_dthr BETWEEN '2022-01-01' AND '2023-06-30'
    AND tpobj = 'rea'
    AND texto LIKE '%DESLIGOU%')
```

2. Selecionar, na entidade EVE\_H, todas as medições analógicas e digitais para os reatores, classificadas com Severidade igual a Fatal, ocorridas no mesmo minuto ou anterior a cada um dos eventos. Essas medidas são as transações.

```
SELECT date_trunc('minute', bh_dthr) AS transacao,
    bh_variacao, mrid, obj, ocr,
    severidade, texto, tpobj
FROM eve_h
WHERE severidade = 'K_SEV_FATAL'
    AND (obj LIKE '%RE6%' OR obj LIKE '%RE7%')
    AND (((date_trunc('minute', bh_dthr) -
        INTERVAL '1 minute')
        IN (SELECT minuto FROM eventos)) OR
        (date_trunc('minute', bh_dthr)
        IN (SELECT minuto FROM eventos)))
```

3. Cruzar os dados das entidades EVE\_H e PAS\_R para identificar os tipos de medições analógicas selecionadas como transações

```
SELECT DISTINCT T1.obj, T2.id, T2.descr
FROM eve_h AS T1
JOIN pas_r AS T2 ON T1.obj = T2.id
```

```
WHERE T1.severidade = 'K_SEV_FATAL'
AND T1.tpobj = 'pas'
AND (T1.obj LIKE '%RE6%' OR T1.obj LIKE '%RE7%')
AND (((date_trunc('minute', T1.bh_dthr) -
      INTERVAL '1 minute')
      IN (SELECT minuto FROM eventos)) OR
      (date_trunc('minute', T1.bh_dthr)
      IN (SELECT minuto FROM eventos)))
```

O resultado foi armazenado no arquivo 'de\_para\_colunas\_pas\_22\_23.csv'

4. Cruzar os dados das entidades EVE\_H e PDS\_R para identificar os tipos de medições digitais selecionadas como transações

```
SELECT DISTINCT T1.obj, T2.id, T2.descr
FROM eve_h AS T1
JOIN pds_r AS T2 ON T1.obj = T2.id
WHERE T1.severidade = 'K_SEV_FATAL'
AND T1.tpobj = 'pds'
AND (T1.obj LIKE '%RE6%' OR T1.obj LIKE '%RE7%')
AND (((date_trunc('minute', T1.bh_dthr) -
      INTERVAL '1 minute')
      IN (SELECT minuto FROM eventos)) OR
      (date_trunc('minute', T1.bh_dthr)
      IN (SELECT minuto FROM eventos)))
```

O resultado foi armazenado no arquivo 'de\_para\_colunas\_pds\_22\_23.csv'

5. Selecionar, na entidade EVE\_H, uma amostra aleatória, limitada a 7000 eventos de reatores, sem desligamento, no intervalo de janeiro de 2022 até junho de 2023

```
CREATE TEMPORARY TABLE amostra AS (
  SELECT date_trunc('minute', bh_dthr) AS minuto, obj
FROM eve_h
WHERE bh_dthr BETWEEN '2022-01-01' AND '2023-06-30'
AND tpobj = 'rea'
AND texto NOT LIKE '%DESLIGOU%'
ORDER BY RANDOM()
LIMIT 7000)
```



6. Selecionar, na entidade EVE\_H, todas as transações (data/hora/minuto) de reatores, classificadas com Severidade igual a Fatal, ocorridas no mesmo minuto ou anterior à amostra

```
SELECT date_trunc('minute', bh_dthr) AS transacao,
       bh_variacao, mrid, obj, ocr,
       severidade, texto, tpobj
FROM eve_h
WHERE severidade = 'K_SEV_FATAL'
      AND (obj LIKE '%RE6%' OR obj LIKE '%RE7%')
      AND (((date_trunc('minute', bh_dthr) -
                INTERVAL '1 minute')
            IN (SELECT minuto FROM amostra)) OR
            (date_trunc('minute', bh_dthr)
            IN (SELECT minuto FROM amostra)))
```

7. Cruzar os dados das entidades EVE\_H e PAS\_R para identificar os tipos de medições analógicas selecionadas como transações

```
SELECT DISTINCT T1.obj, T2.id, T2.descr
FROM eve_h AS T1
JOIN pas_r AS T2 ON T1.obj = T2.id
WHERE T1.severidade = 'K_SEV_FATAL'
      AND T1.tpobj = 'pas'
      AND (T1.obj LIKE '%RE6%' OR
            T1.obj LIKE '%RE7%')
      AND (((date_trunc('minute', T1.bh_dthr) -
                INTERVAL '1 minute')
            IN (SELECT minuto FROM amostra)) OR
            (date_trunc('minute', T1.bh_dthr)
            IN (SELECT minuto FROM amostra)))
```

O resultado foi armazenado no arquivo 'de\_para\_colunas\_pas\_22\_23\_sem2.csv'

8. Cruzar os dados das entidades EVE\_H e PDS\_R para identificar os tipos de medições digitais selecionadas como transações

```
SELECT DISTINCT T1.obj, T2.id, T2.descr
```

```
FROM eve_h AS T1
JOIN pds_r AS T2 ON T1.obj = T2.id
WHERE T1.severidade = 'K_SEV_FATAL'
      AND T1.tpobj = 'pds'
      AND (T1.obj LIKE '%RE6%' OR
           T1.obj LIKE '%RE7%')
      AND (((date_trunc('minute', T1.bh_dthr) -
              INTERVAL '1 minute')
            IN (SELECT minuto FROM amostra)) OR
           (date_trunc('minute', T1.bh_dthr)
            IN (SELECT minuto FROM amostra)))
```

O resultado foi guardado no arquivo 'de\_para\_colunas\_pds\_22\_23\_sem2.csv'

## APÊNDICE B – PRÉ-PROCESSAMENTO DOS DADOS

### 1. Importar os dados extraídos

```
import pandas as pd

df = pd.read_csv('de_para_colunas_pas_22_23.csv', sep=',')
df1 = pd.read_csv('de_para_colunas_pas_22_23_sem2.csv', sep=',')
df2 = pd.read_csv('de_para_colunas_pds_22_23.csv', sep=',')
df3 = pd.read_csv('de_para_colunas_pds_22_23_sem2.csv', sep=',')
df4 = pd.read_csv('eve_2022_2023_deslig.csv', sep=',')
df5 = pd.read_csv('eve_2022_2023_sem_deslig2.csv', sep=',')
```

### 2. Concatenar os quatro arquivos do subconjunto “de\_para\_colunas”, criando um único dataframe ‘df’ com registro das medições analógicas e digitais, de eventos com e sem desligamento de reatores (Figura 17)

```
df = pd.concat([df, df1, df2, df3])
```

[21] df

	obj	id	descr
0	AN_ANRE6-02_167 ...	AN_ANRE6-02_167	ANRE6-02 Tensão entre Fases AB
1	AQRE601MIA ...	AQRE601MIA	AQRE6-01 Corrente Fase A
2	AQRE601MIB ...	AQRE601MIB	AQRE6-01 Corrente Fase B
3	AQRE601MIV ...	AQRE601MIV	AQRE6-01 Corrente Fase V
4	AQRE601MQ ...	AQRE601MQ	AQRE6-01 Potência Reativa
...	...	...	...
9	RB_RBRE6-01_16383 ...	RB_RBRE6-01_16383	RBRE6-01 63 Relé de Gás 2oEst
0	IZRE704087RV ...	IZRE704087RV	IZRE7-04 87R Prot Diferencial Reator FV
1	IZRE712ETPA ...	IZRE712ETPA	IZRE7-12 Proteções Reator FA
2	IZRE712ETPB ...	IZRE712ETPB	IZRE7-12 Proteções Reator FB
3	IZRE712ETPV ...	IZRE712ETPV	IZRE7-12 Proteções Reator FV

406 rows × 3 columns

Figura 17 – Concatenação dos arquivos em um dataframe (Fonte: Autor)

3. Inserir no dataframe a coluna 'medida', resultado da cópia da coluna 'descr' sem a especificação do reator e mais um processo de padronização dos nomes das medições (Figura 18). A padronização é necessária, pois a mesma classe aparece eventualmente com nomes diferentes no SAGE

```
df['medida'] = df['descr'].str.strip().str[9:]

df['medida'] = df['medida'].replace('    Tensão entre Fases BV 230kV',
'    Tensão entre Fases BV')
df['medida'] = df['medida'].replace('Tensão Fases BV',
'    Tensão entre Fases BV')

df['medida'] = df['medida'].replace('    Tensão entre Fases VA 230kV',
'    Tensão entre Fases VA')
df['medida'] = df['medida'].replace('    Tensão entre Fases AB 230kV',
'    Tensão entre Fases AB')

df['medida'] = df['medida'].replace('    Frequência', '    Frequência')
df['medida'] = df['medida'].replace('    Frecuencia', '    Frequência')

df['medida'] = df['medida'].replace('    Corrente Fase A 230kV',
'    Corrente Fase A')
df['medida'] = df['medida'].replace('Corrente Fase B',
'    Corrente Fase B')

df['medida'] = df['medida'].replace('    Temp Óleo Fase A',
'    Temperatura de oleo Fase A')
df['medida'] = df['medida'].replace('Temp Óleo Fase B',
'    Temperatura de oleo Fase B')
df['medida'] = df['medida'].replace('    Temp Óleo Fase V',
'    Temperatura de oleo Fase V')
df['medida'] = df['medida'].replace('Temp Óleo',
'    Temperatura de Óleo')
df['medida'] = df['medida'].replace('    Temp Enrol Fase A',
'    Temperatura Enrol Fase A')
df['medida'] = df['medida'].replace('Temp Enrolamento Fase B',
'    Temperatura Enrol Fase B')
df['medida'] = df['medida'].replace('Temp Enrolamento Fase V',
'    Temperatura Enrol Fase V')
df['medida'] = df['medida'].replace('    Temp Enrol Fase V',
```

df

	obj	id	descr	medida
0	AN_ANRE6-02_167 ...	AN_ANRE6-02_167	ANRE6-02 Tensão entre Fases AB	Tensão entre Fases AB
1	AQRE601MIA ...	AQRE601MIA	AQRE6-01 Corrente Fase A	Corrente Fase A
2	AQRE601MIB ...	AQRE601MIB	AQRE6-01 Corrente Fase B	Corrente Fase B
3	AQRE601MIV ...	AQRE601MIV	AQRE6-01 Corrente Fase V	Corrente Fase V
4	AQRE601MQ ...	AQRE601MQ	AQRE6-01 Potência Reativa	Potência Reativa
...	...	...	...	...
9	RB_RBRE6-01_16383 ...	RB_RBRE6-01_16383	RBRE6-01 63 Relé de Gás 2oEst	63 Relé de Gás 2oEst
0	IZRE704087RV ...	IZRE704087RV	IZRE7-04 87R Prot Diferencial Reator FV	87R Prot Diferencial Reator FV
1	IZRE712ETPA ...	IZRE712ETPA	IZRE7-12 Proteções Reator FA	Proteções Reator FA
2	IZRE712ETPB ...	IZRE712ETPB	IZRE7-12 Proteções Reator FB	Proteções Reator FB
3	IZRE712ETPV ...	IZRE712ETPV	IZRE7-12 Proteções Reator FV	Proteções Reator FV

406 rows x 4 columns

Figura 18 – Dataframe modificado (Fonte: Autor)

```

    Temperatura Enrol Fase V')
df['medida'] = df['medida'].replace(
    '    Temperatura Enrolamento 230kV', '    Temperatura Enrol')

df['medida'] = df['medida'].replace('    63 Relé de Gás 2oEst FV',
    '63 Relé de Gás Fase V 2oEst')

```

4. Copiar as colunas 'obj' e 'medida' para criar o dicionário de medidas 'map'

```
map = dict(zip(df['obj'], df['medida']))
```

5. Extrair as colunas 'transacao', 'bh\_variacao' e 'obj' do subconjunto de arquivos "eve\_2022\_2023", copiadas nos dataframes 'df4' e 'df5'. O resultado é copiado primeiro nos dataframes modificados, depois padronizados pelo dicionário 'map'

```

df4_modif = df4[['transacao', 'bh_variacao', 'obj']]
df5_modif = df5[['transacao', 'bh_variacao', 'obj']]

df4_modif['obj'] = df4_modif['obj'].replace(map)
df5_modif['obj'] = df5_modif['obj'].replace(map)

```

```

▶ print('Número de tuplas do dicionário "map": ', len(map))

Número de tuplas do dicionário "map": 244

▶ map

'PDRE702A0TOA'      ': ' Temperatura de oleo Fase A',
'PDRE702A0TOB'      ': ' Temperatura de oleo Fase B',
'PDRE702A0TOR'      ': ' Temperatura de oleo Fase R',
'PDRE702A0TOV'      ': ' Temperatura de oleo Fase V',
'PDRE702M000F'      ': ' Frequência',
'PDRE703A0TEB'      ': ' Temperatura Enrol Fase B',
'PDRE703A0TER'      ': ' Temperatura Enrol Fase R',
'PDRE703A0TOA'      ': ' Temperatura de oleo Fase A',
'PDRE703A0TOB'      ': ' Temperatura de oleo Fase B',
'PDRE703A0TOR'      ': ' Temperatura de oleo Fase R',
'PDRE703M000F'      ': ' Frequência',
'PDRE703M00IA'      ': ' Corrente Fase A',
'PDRE703M00IB'      ': ' Corrente Fase B',
'PDRE703M00IV'      ': ' Corrente Fase V',
'PDRE703M00QT'      ': ' Potência Reativa',

```

Figura 19 – Dicionário de medidas (Fonte: Autor)

[34] df4\_modif

	transacao	bh_variacao	obj
0	2022-01-27 16:51:00-03	1	Frequência
1	2022-01-27 16:51:00-03	1	Tensão entre Fases BV
2	2022-01-27 16:51:00-03	1	Frequência
3	2022-02-11 00:37:00-03	1	Tensão entre Fases BV
4	2022-02-11 00:37:00-03	1	Corrente Fase A
...	...	...	...
1290	2023-06-12 07:44:00-03	1	Proteções Reator FV
1291	2023-06-13 08:09:00-03	2	Corrente Fase A
1292	2023-06-13 08:09:00-03	2	Potência Reativa
1293	2023-06-13 08:09:00-03	2	Tensão entre Fases BV
1294	2023-06-13 08:09:00-03	2	Corrente Fase A

1295 rows × 3 columns

Figura 20 – Dataframe de medidas analógicas, padronizado pelo dicionário de medidas (Fonte: Autor)

```
[35] df5_modif
```

	transacao	bh_variacao	obj
0	2022-01-16 00:14:00-03	1	Proteções Reator FA
1	2022-01-18 14:39:00-03	1	Corrente Fase A
2	2022-01-18 14:39:00-03	1	Potência Reativa
3	2022-01-27 16:51:00-03	1	Frequência
4	2022-01-27 16:51:00-03	1	Tensão entre Fases BV
...	...	...	...
1272	2023-06-13 13:09:00-03	2	Frequência
1273	2023-06-13 13:09:00-03	2	Tensão entre Fases BV
1274	2023-06-13 13:09:00-03	2	Corrente Fase B
1275	2023-06-28 13:54:00-03	1	Proteções Reator FA
1276	2023-06-28 13:54:00-03	1	Proteções Reator FV

1277 rows × 3 columns

Figura 21 – Dataframe de medidas digitais, padronizado pelo dicionário de medidas (Fonte: Autor)

6. Transpor as colunas ‘obj’ nos dataframes anteriores, ‘df4\_modif’ e ‘df5\_modif’, de modo que seus valores se tornem colunas para dois novos dataframes, conforme as seguintes regras:

- as colunas ‘transacao’ se tornam index
- as colunas ‘bh\_variacao’ desaparecem, mas os valores são copiados para as novas colunas transpostas

```
df4_pivot = pd.pivot_table(df4_modif, index='transacao',
                           columns='obj', values='bh_variacao')
df5_pivot = pd.pivot_table(df5_modif, index='transacao',
                           columns='obj', values='bh_variacao')
```

7. Acrescentar a coluna ‘target’ a cada um dos dataframes pivot, conforme as seguintes regras:

- o dataframe ‘df4\_pivot’ lista transações com desligamento de reatores, portanto o Target é igual a 1

	transacao	bh_variacao	obj			
	0	2022-01-16 00:14:00-03	1	Proteções Reator FA		
	1	2022-01-18 14:39:00-03	1	Corrente Fase A		
	2	2022-01-18 14:39:00-03	1	Potência Reativa		
	3	2022-01-27 16:51:00-03	1	Frequência		
	4	2022-01-27 16:51:00-03	1	Tensão entre Fases BV		
	...	...	...	...		
	obj	Proteções Reator FA	Corrente Fase A	Potência Reativa	Frequência	Tensão entre Fases BV
	transacao					
	2022-01-16 00:14:00-03	1.0	NaN	NaN	NaN	NaN
	2022-01-18 14:39:00-03	NaN	1.0	1.0	NaN	NaN
	2022-01-27 16:51:00-03	NaN	NaN	NaN	1.0	1.0
	2022-02-01 04:51:00-03	NaN	NaN	NaN	1.0	1.0
	2022-02-10 00:37:00-03	NaN	NaN	NaN	1.0	1.0

Figura 22 – Exemplo de transposição com linhas do dataframe de medidas digitais (Fonte: Autor)

- o dataframe 'df5\_pivot' lista transações sem desligamento de reatores, portanto o Target é igual a 0

```
df4_pivot['target'] = 1.0
df5_pivot['target'] = 0.0
```

- Concatenar os dataframe pivot em um único dataframe final (Figuras 23 e 24). Os valores nulos são preenchidos com 0 (zero) e os valores acima de zero são substituídos por 1 (um)

```
dfrea = pd.concat([df4_pivot, df5_pivot])
dfrea = dfrea.fillna(0)
dfrea = dfrea.replace(2,1)

dfrea = dfrea.sort_index()

dfrea.reset_index(drop=True, inplace=True)
```

O resultado foi guardado no arquivo 'reatores.csv'



```
dfrea['target'].value_counts()
```

```
1.0    130
0.0    129
Name: target, dtype: int64
```

```
dfrea
```

obj	27 Proteção SubTensão	51N Prot SobreCor Neutro Temp	63 Relé de Gás 2oEst	BITS Alarmes FA 230kV	BM 230kV Ângulo BA	BM Capacitância Alta FV 500kV	BM FA 230kV Tempo TDelta Alta	BM FA 230kV Tend Capacitância	BM FV 230kV Tempo Cap M Alta	BM FV 230kV Tempo TDelta MAlta	...
0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...
1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...
2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...
3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...
4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...
...	...	...	...	...	...	...	...	...	...	...	...
254	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...
255	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...
256	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...
257	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...
258	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...

259 rows x 63 columns

Figura 23 – Dataframe final: todas as medidas (transações) analógicas e digitais de reatores (Fonte: Autor)

```
dfrea.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Index: 259 entries, 2022-01-16 00:14:00-03 to 2023-06-...
Data columns (total 63 columns):
#   Column                                     Non-Null Count
---  -
0   27 Proteção SubTensão                     259 non-null
1   51N Prot SobreCor Neutro Temp             259 non-null
2   63 Relé de Gás 2oEst                      259 non-null
3   BITS Alarmes FA 230kV                    259 non-null
4   BM 230kV Ângulo BA                       259 non-null
5   BM Capacitância Alta FV 500kV            259 non-null
6   BM FA 230kV Tempo TDelta Alta            259 non-null
7   BM FA 230kV Tend Capacitância            259 non-null
8   BM FV 230kV Tempo Cap M Alta             259 non-null
9   BM FV 230kV Tempo TDelta MAlta          259 non-null
10  BM MM2 Versão Firmware                   259 non-null
11  Corrente Fase A                          259 non-null
12  Corrente Fase B                          259 non-null
13  Corrente Fase V                          259 non-null
14  Frequência                              259 non-null
15  Med Corrente Fuga FA 230kV              259 non-null
16  Par Cap Inicial FV 230kV                259 non-null
17  Potência Ativa                           259 non-null
18  Potência Reativa                         259 non-null
19  Proteções Reator FA                     259 non-null
20  Proteções Reator FV                     259 non-null
21  TM Par NGR Número Grupos Resfr          259 non-null
22  Temperatura Enrol                        259 non-null
23  Temperatura Enrol Fase A                 259 non-null
24  Temperatura Enrol Fase B                 259 non-null
25  Temperatura Enrol Fase R                 259 non-null
26  Temperatura Enrol Fase V                 259 non-null
27  Temperatura de oleo Fase A              259 non-null
28  Temperatura de oleo Fase B              259 non-null
29  Temperatura de oleo Fase R              259 non-null
30  Temperatura de oleo Fase V              259 non-null
31  Temperatura de Óleo                     259 non-null
32  Tensão Fase A                           259 non-null
33  Tensão Fase B                           259 non-null
34  Tensão Fase V                           259 non-null
35  Tensão entre Fases AB                   259 non-null
36  Tensão entre Fases BV                   259 non-null
37  Tensão entre Fases VA                   259 non-null
38  49 SobreTemp Enrol 2oEst                 259 non-null
39  63 Relé de Gás Fase A 2oEst             259 non-null
40  63 Relé de Gás Fase V 2oEst             259 non-null
41  Potência Aparente Fase A               259 non-null
42  target                                  259 non-null
43  87R Prot Diferencial Reator FV          259 non-null
44  BITS Alarmes IRF                        259 non-null
45  BM Capacitância FB 230kV                259 non-null
46  BM FA 230kV Tend Tan Delta              259 non-null
47  BM FB 230kV Tend Capacitância           259 non-null
48  BM FV 230kV Tempo Cap Alta             259 non-null
49  BM HMI Número Série 2                  259 non-null
50  BM MM2 Número Série 1                  259 non-null
51  BM Memória Erro BM-MM End 1            259 non-null
52  BM Par Corr Fuga Alta FA 230kV          259 non-null
53  BM TangDelta FA 230kV                   259 non-null
54  BM Tend Capacitância FB 230kV          259 non-null
55  Corrente Neutro                         259 non-null
56  Med Tan Delta FB 230kV                  259 non-null
57  Par Tan Delta Inicial FA 230kV          259 non-null
58  Proteções Reator FB                     259 non-null
59  TM Corrente Sec TC Enrol 230kV          259 non-null
60  TM Par Histerese Resfriamento          259 non-null
61  Potência Aparente                       259 non-null
62  Potência Reativa Fase A                 259 non-null
dtypes: float64(63)
```

Figura 24 – Dataframe final: detalhamento (Fonte: Autor)