

CluAlarm: Agrupamento inteligente de fluxos contínuos de alarmes em um Centro de Operação de Energia Elétrica.

Felipe Castelar Torres Silva

Trabalho de Conclusão de Curso
MBA em Inteligência Artificial e Big Data

UNIVERSIDADE DE SÃO PAULO

Instituto de Ciências Matemáticas e de Computação

CluAlarm

Felipe Castelar Torres Silva

USP - São Carlos

2023

Felipe Castelar Torres Silva

CluAlarm: Agrupamento inteligente de fluxos contínuos de alarmes em um Centro de Operação de Energia Elétrica.

Trabalho de conclusão de curso apresentado 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: Dados em Larga Escala (Big Data)

Orientador: Prof. Dr. Ricardo Cerri

USP - 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)

C348c CASTELAR TORRES SILVA, FELIPE
CluAlarm: Agrupamento inteligente de fluxos
contínuos de alarmes em um Centro de Operação de
Energia Elétrica. / FELIPE CASTELAR TORRES SILVA;
orientador Ricardo Cerri. -- São Carlos, 2023.
58 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. Inteligência Artificial . 2. Data Stream. 3.
Clustering . 4. SAGE EMS. 5. Sistema Interligado
Nacional (Energia Elétrica). I. Cerri, Ricardo,
orient. II. Título.

Bibliotecários responsáveis pela estrutura de catalogação da publicação de acordo com a AACR2:
Gláucia Maria Saia Cristianini - CRB - 8/4938
Juliana de Souza Moraes - CRB - 8/6176

DEDICATÓRIA

A toda a minha família.

AGRADECIMENTOS

Ao corpo docente do ICMC/USP, em especial ao meu professor orientador Dr. Ricardo Cerri, que com excelente competência e didática contribuiu para o meu crescimento científico. Também agradeço aos gestores da Eletronorte por confiarem a mim a missão de me qualificar cada vez mais como profissional. Finalmente, agradeço a minha família pelo suporte, compreensão e carinho durante mais esse empreendimento.

EPÍGRAFE

Os dados não são apenas o novo petróleo,
mas também a nova eletricidade, uma fonte
essencial de energia para o futuro" -
Masayoshi Son (CEO do SoftBank Group,
Conferência de Tecnologia Móvel,
Barcelona, 2022).

RESUMO

CASTELAR, F. **CluAlarm**: Agrupamento inteligente de fluxos contínuos de alarmes em um Centro de Operação de Energia Elétrica. 2023. 59 f. Trabalho de conclusão de curso (MBA em Inteligência Artificial e Big Data) – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos, 2023.

A operação teleassistida de subestações e usinas do sistema elétrico brasileiro tem um grande potencial na redução de custos e aumento da eficiência operacional. Especificamente, destaca-se que uma equipe concentrada e mais capacitada pode substituir uma operação local distribuída geograficamente nas diversas plantas, reduzindo o quadro de funcionários necessários em cerca de 90%. No entanto, em cenários esporádicos de perturbações relevantes no sistema elétrico, como um apagão regional ou até mesmo nacional, pode haver risco de sobrecarga dos operadores do centro de operação, comprometendo a recomposição do sistema elétrico. Um dos principais desafios nesse contexto é a avaliação rápida e assertiva da condição do sistema elétrico após uma perturbação, com base nos alarmes apresentados em tempo real pelo Energy Management System – EMS. Para abordar esse problema, esse trabalho apresenta uma ferramenta para processamento automático de alarmes, visando agrupar os fluxos contínuos de alarmes e diminuir a carga de trabalho analítico dos operadores, especialmente em cenários de perturbações relevantes no sistema elétrico.

Palavras-chave: Inteligência Artificial 1; Data Stream 2; Clustering 3; SAGE – EMS 4; Sistema Interligado Nacional 5.

ABSTRACT

CASTELAR, F. **CluAlarm**: Intelligent clustering of continuous streams of alarms in an Electric Power Operation Center – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos, 2020.

The teleassisted operation of substations and power plants in the Brazilian electrical system has great potential for reducing costs and increasing operational efficiency. Specifically, it highlights that a concentrated and more skilled team can replace a geographically distributed local operation in various plants, reducing the required staff by about 90%. However, in sporadic scenarios of relevant disturbances in the electrical system, such as a regional or even national blackout, there may be a risk of overload for the operators in the operation center, compromising the electrical system's reestablishment. One of the main challenges in this context is the quick and assertive evaluation of the electrical system's condition after a disturbance, based on alarms presented in real-time by the Energy Management System (EMS). To address this problem, this work presents a tool for automatic processing of alarms, aiming to group continuous flows of alarms and reduce the analytical workload of the operators, especially in scenarios of relevant disturbances in the electrical system.

Keywords: Artificial Intelligence 1; Data Stream 2; Clustering 3; SAGE - EMS (Energy Management System) 4; Brazilian Interconnected Electric Power Grid 5.

LISTA DE ILUSTRAÇÕES

Figura 1 – Capacidade de Geração Instalada do SIN.....	23
Figura 2 – Organização das principais entidades do SEB.....	24
Figura 3 – Algoritmos de Agrupamento em Fluxos Contínuos.....	28
Figura 4 – Pipeline de Dados do CluAlarm.....	34
Figura 5 – Resultado da Consulta ao Banco de Dados Histórico do SAGE EMS.....	35
Figura 6 – Janelamento de Eventos de Alarmes em Tempo Real.....	36
Figura 7 – Recorte dos Diagramas Unifilares Operacionais.....	40
Figura 8 – Recorte Instalações 500 kV PA, MA e TO.....	43
Figura 9 – Apresentação e Compartilhamento.....	46
Figura 10 – Perturbação de Grande Porte.....	49
Figura 11 – Perturbação Consecutivas Disjuntas.....	50
Figura 12 – Distribuição Tempo de Resposta Inicial.....	52
Figura 13 – Distribuição até a última atualização.....	53

LISTA DE TABELAS

Tabela 1 – Agrupamento Tradicional e em Fluxos de Dados Contínuos.....	28
Tabela 2 – Agrupamento Tradicional e em Fluxos de Dados Contínuos	39
Tabela 3 – Ligações SAGE EMS.....	42
Tabela 4 – Estatísticas Tempo Resposta Inicial	52

LISTA DE ABREVIATURAS E SIGLAS

ANEEL	–	Agência Nacional de Energia Elétrica
CCEE	–	Câmara de Comercialização de Energia Elétrica
CEPEL	–	Centro de Pesquisas de Energia Elétrica da Eletrobras
EMS	–	Energy Management System (Sistema de Gerenciamento de Energia)
MOA	–	Massive Online Analysis (Análise Online Massiva)
ONS	–	Operador Nacional do Sistema Elétrico
MQTT	–	Message Queuing Telemetry Transport
OPC-UA	–	Open Platform Communications – Unified Architecture
SAGE	–	Sistema Aberto de Gerenciamento de Energia (CEPEL)
SEB	–	Sistema Elétrico Brasileiro
SIN	–	Sistema Interligado Nacional
SQL	–	Structured Query Language

SUMÁRIO

1 INTRODUÇÃO.....	23
1.1 O ONS e os Procedimentos de Rede.....	24
1.2 Centros de Operação de Sistema.....	25
1.3 Motivação e Objetivos.....	26
2 REVISÃO BIBLIOGRÁFICA.....	27
2.1 Algoritmos de Agrupamento em Fluxos Contínuo de Dados.....	29
2.1.1 Incremental k-means.....	29
2.1.2 CluStream.....	29
2.1.3 DenStream.....	30
2.2 Agrupamento de Fluxos Contínuos de Rede de Sensores Distribuídos.....	31
2.3 Energy Management System – EMS.....	31
3 CLUALARM.....	33
3.1 Proposta de Desenvolvimento.....	33
3.2 Preparação para o desenvolvimento.....	34
3.3 Metodologia de Ingestão de Dados.....	35
3.4 Metodologia de Processamento em Tempo Real.....	37
3.5 Metodologia de Apresentação e Compartilhamento.....	38
4 AVALIAÇÃO DOS RESULTADOS.....	39
4.1 Descrição dos conjuntos de dados e do processamento.....	39
4.1.1 Identificação de perturbações.....	39
4.1.2 Proximidade Elétrica entre Instalações.....	40
4.1.3 Proximidade Elétrica entre Equipamentos.....	42
4.1.4 Janelamento e agrupamento inteligente.....	44
4.1.5 Atualização dos agrupamentos em tempo real.....	45
4.1.6 Exportação dos dados e Apresentação.....	47
4.2 Configuração experimental.....	47
4.3 Resultados e Discussão.....	48
4.3.1 Resultados experimentais preliminares.....	49
4.3.1.1 Perturbação de Grande Porte.....	48
4.3.1.2 Perturbações consecutivas em área elétricas disjuntas.....	50
4.3.2 Comparação com os relatórios de perturbação do sistema de gestão.....	51

4.3.3 Número de alarmes.....	51
4.3.4 Tempos para apresentação preliminar do agrupamento de perturbações.....	51
4.3.5 Análise Crítica e Limitações.....	53
5 CONCLUSÃO.....	55
REFERÊNCIAS.....	56

1. INTRODUÇÃO

O Sistema Elétrico Brasileiro – SEB é um dos maiores do mundo em termos de capacidade instalada, com uma capacidade total de geração de energia elétrica de mais de 180 *GW*, dos quais mais de 85% são gerados a partir de fontes renováveis, como hidrelétricas, eólicas e solares, como apresentado na Figura 1. Ademais, o Sistema Integrado Nacional – SIN, que conecta todos os estados da federação, exceto o estado de Roraima, possui mais de 270.000 *km* de linhas de transmissão apenas em sua rede básica – tensão superior a 230 *kV* (ONS, 2023).

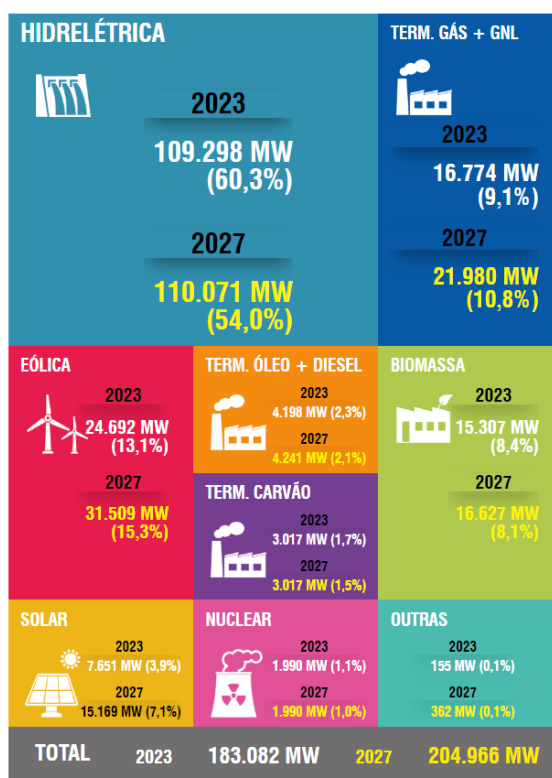


Figura 1 – Capacidade de Geração Instalada do SIN. Fonte: ONS (2023).

O desenvolvimento do chamado novo modelo do SEB foi impulsionado por diversos marcos regulatórios, que permitiram a criação de uma infraestrutura robusta e moderna, e que vem se aprimorando ao longo dos anos.

Dentre esses marcos, destacam-se: a abertura do mercado para a livre concorrência e a criação de um órgão regulador, a Agência Nacional de Energia Elétrica – ANEEL (Lei 9.074/1995 e Lei 9.427/1996); a criação da Câmara de Comercialização de Energia Elétrica – CCEE, que é responsável pela contabilização e liquidação financeira das transações realizadas no mercado de energia elétrica (Lei 9.648/1998); a criação do Operador Nacional do Sistema

Elétrico – ONS, que tem como função coordenar e controlar a operação do sistema elétrico brasileiro, garantindo a segurança, a qualidade e a confiabilidade do fornecimento de energia elétrica em todo o território nacional (Lei 9.648/1998); a regulamentação da contratação de energia pelas distribuidoras, que passaram a ser obrigadas a realizar leilões para adquirir a energia necessária para atender a sua demanda (Lei 10.848/2004); condições para a renovação das concessões de geração, transmissão e distribuição de energia elétrica, que venceram entre 2015 e 2017 para redução do preço das tarifas (Lei 12.783/2013); e a desestatização da Eletrobras, que é a maior empresa de geração e transmissão de energia elétrica do país Lei nº 14.182/2021. A Figura 2 mostra um organograma das principais entidades do SEB.

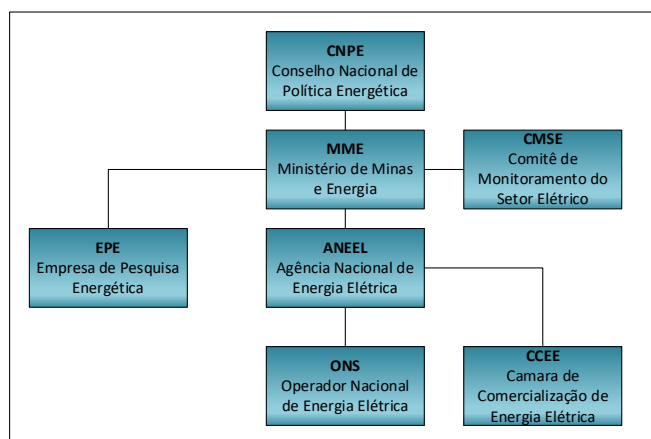


Figura 2 – Organização das principais entidades do SEB. Fonte: Adaptado de CCEE (2023).

1.1. O ONS e os Procedimentos de Rede

O ONS é responsável pela coordenação e controle da operação das instalações de geração e transmissão de energia elétrica no SIN e pelo planejamento da operação dos sistemas isolados do país, sob a fiscalização e regulação da ANEEL. Esse órgão como uma pessoa jurídica de direito privado, sob a forma de associação civil sem fins lucrativos, é composto por membros associados e membros participantes, que são as empresas de geração, transmissão, distribuição, consumidores livres, importadores e exportadores de energia. Também participam o Ministério de Minas e Energia (MME) e representantes dos Conselhos de Consumidores.

Para o exercício de suas atribuições legais e o cumprimento de sua missão institucional, o ONS desenvolve uma série de estudos e ações exercidas sobre o sistema e seus agentes proprietários para gerenciar as diferentes fontes de energia e a rede de transmissão, de forma a garantir a segurança do suprimento contínuo em todo o país, com os objetivos de promover a

otimização da operação do sistema eletroenergético, visando ao menor custo para o sistema, observados os padrões técnicos e os critérios de confiabilidade; garantir que todos os agentes do setor elétrico tenham acesso à rede de transmissão de forma não discriminatória; e contribuir, de acordo com a natureza de suas atividades, para que a expansão do SIN se faça ao menor custo e vise às melhores condições operacionais futuras.

Para tanto, são definidos nos Procedimentos de Rede do ONS e aprovados pela ANEEL, um conjunto de normas e diretrizes técnicas que regulamentam a operação do SEB. Esses procedimentos estabelecem as regras e os procedimentos que devem ser seguidos pelos agentes do setor elétrico, como geradores, distribuidoras e consumidores, para garantir a segurança, a qualidade e a confiabilidade do fornecimento de energia elétrica no país.

1.2. Centros de Operação de Sistema

No submódulo 5.1 dos procedimentos de rede (ONS, 2023) é definida uma hierarquia operacional, que rege o relacionamento entre a operação e o controle de instalações, como usinas geradoras e subestações de energia elétrica dos agentes, e a operação e o controle de sistema, executada pelos próprios centros de operação do ONS. Nesse mesmo submódulo, é facultado ao agente realizar a operação de suas instalações localmente, diretamente a partir de suas subestações e usinas, ou por meio de um ou mais centros de operação do agente.

Entretanto, para que haja a assistência remota de suas instalações a partir de um centro de operações, o agente do sistema elétrico deve cumprir uma série de requisitos que visam garantir a operação segura e eficiente, como: possuir uma infraestrutura física e tecnológica, que permita a supervisão e controle remoto do sistema elétrico em tempo real, a comunicação por voz sem discagem (*hotline*), e sistemas de monitoramento por imagem de suas instalações; possuir equipe qualificada e treinada para desempenhar suas funções, que incluem o monitoramento do sistema elétrico, a identificação e solução de problemas e a tomada de decisões em situações críticas em regime ininterrupto (24/7); seguir procedimentos operacionais padronizados, que estabeleçam as diretrizes e os protocolos para o monitoramento e controle do sistema elétrico, bem como para a comunicação com os demais agentes do setor elétrico; contar com um sistema de gestão de emergências que permita a rápida identificação e solução de problemas no sistema elétrico, bem como a comunicação com as autoridades e demais agentes envolvidos em situações de emergência.

Com a evolução da tecnologia a digital para operação remota das plantas de transmissão e geração, é patente o movimento estratégico das empresas agentes do SEB em concentrar seus

processos de operação, inclusive o atendimento ininterrupto em tempo real, em centros de operação. Esse conceito de assistência remota das plantas ficou conhecido como teleassistência e está cada vez mais em prática nas principais empresas do setor.

1.3. Motivação e Objetivos

A teleassistência tem um potencial relevante na redução de custos e no aumento da eficiência operacional dos agentes do setor elétrico, tendo em vista que para realizar as atividades de operação, apenas uma equipe concentrada e eventualmente mais capacitada será responsável por lidar com os desafios da operação ininterrupta. Essa equipe toma o lugar de uma operação local distribuída geograficamente nas diversas plantas.

A redução do quadro de funcionários necessários ao atendimento ininterrupto das plantas pode cair num fator de cerca de 90%, considerando um cenário com 20 instalações por um centro de operação com 04 postos operando em regime 24/7. Ademais, por se tratar de uma equipe mais enxuta e concentrada em uma única localidade geográfica, os desafios de capacitação constante e melhoria dos processos são bastante facilitados.

Contudo, em cenários esporádicos de perturbações relevantes no sistema elétrico, como um apagão regional ou até mesmo nacional, há um risco significativo de sobrecarga dos operadores do centro de operação. Essa sobrecarga pode comprometer a recomposição do sistema elétrico acarretando grande prejuízo à população e às atividades econômicas do país, bem como penalidades impostas pela agência reguladora, e até mesmo riscos aos ativos e pessoas podem decorrer de falhas operacionais durante uma recomposição do sistema em um cenário complexo como esse.

Um dos principais desafios em relação a essa sobrecarga está em o operador avaliar rapidamente e assertivamente a condição do sistema elétrico após uma perturbação, com base nos alarmes apresentados pelo sistema de supervisão em tempo real das diversas plantas subordinadas ao centro de operação. Esse esforço analítico pode ser bastante relevante tendo em vista uma avalanche de alarmes em uma perturbação de grande porte.

Assim, esse trabalho tem o propósito de desenvolver uma ferramenta para processamento de alarmes apresentados ao operador nos centros de operação. O objetivo central dessa ferramenta é realizar o agrupamento dos fluxos contínuos de alarmes, utilizando técnicas de processamento inteligentes, diminuindo a carga de trabalho analítico dos operadores, especialmente em cenários de perturbações relevantes no sistema elétrico.

2. REVISÃO BIBLIOGRÁFICA

Um dos desafios no processamento de um número grande alarmes em um centro de operações, decorrentes de perturbações relevantes no sistema elétrico, é lidar com a sua natureza de tempo real. É importante, nesse caso, que os esquemas de processamento e a análise desses dados tenham respostas próximas do imediato e que sejam mutáveis com a chegada de novos dados em tempo real.

Fluxos contínuos de dados ou *Data Streams* são abstrações que dão suporte a análise de dados em tempo real (BIFET, 2018). Esses fluxos podem ser modelados como uma sequência, por vezes infinita, em que cada item além de seus atributos característicos, possui uma estampa de tempo, que os ordena. De acordo com BIFET, 2010, formalmente um fluxo contínuo de dados (*data stream*) é um conjunto de dados multidimensionais $X_1, X_2, \dots, X_k, \dots$, que chegam com selos de tempo $T_1, T_2, \dots, T_k, \dots$, onde cada ponto desse conjunto de dados X_i tem dimensão d e, portanto, $X_i = \{xi1, xi2, \dots, xid\}$.

Nota-se que no processamento de fluxos contínuos de dados o tempo de resposta e uma acomodação razoável em memória são chaves para que o processo de análise seja eficaz quando se compara com mineração dos dados em conjuntos de dados convencionais (BIFET, 2018). Acumular em memória todos os dados que chegam em um fluxo contínuo de dados pode se tornar impraticável e ao mesmo tempo, persistidos para um processamento em lotes pode acarretar uma demora que não contemple os requisitos desejados.

Assim, esse tipo de processamento tem como requisitos: o processamento de cada item do fluxo contínuo em sua chegada; que, preferencialmente, não haja reiterações sobre o processamento de cada item; que há um tempo limite de processamento de cada item; que existe um limite de capacidade de memória; que existe um limite de tempo também para que as respostas sejam úteis; e que existe a necessidade de adaptação de sua resposta ao longo do tempo, com a chegada de novos dados (BIFET, 2018).

Também é comum nesse tipo de processamento que não se tenha conhecimento prévio sobre a rotulação dos dados em classes, tampouco o número de classes. Com isso, a mineração de dados geralmente é feita com métodos de agrupamento não supervisionados, que buscam a partir de similaridades entre os atributos de cada item do fluxo contínuo, agrupá-los em classes distintas, minimizando as dissimilaridades dentro de uma mesma classe. Entretanto, é relevante notar que a maioria dos algoritmos clássicos para esse agrupamento não são preparados para o processamento de fluxos contínuos (BIFET, 2010).

A Tabela 1 mostra lado a lado as principais diferenças entre os mecanismos de agrupamento não supervisionada tradicional e aquelas que são propostos para o processamento de fluxos contínuos de dados.

Tabela 1 – Agrupamento Tradicional e em Fluxos de Dados Contínuos

Agrupamento em fluxos de dados contínuos	Agrupamento Tradicional
Processamento em Tempo Real	Processamento em lotes
Novos dados chegam a cada instante	Dados estáticos
Dados são processados em memória	Dados intermediários podem ser persistidos
Apenas dados sumarizados são persistidos	Dados resultantes são facilmente persistidos

Fonte: Adaptado LIN et. al (2003)

Para resolver essas questões, diversos métodos e algoritmos de agrupamento não supervisionado de fluxos contínuos de dados foram propostos nos últimos anos, como ilustra a Figura 3. Tipicamente, esses métodos consistem na subdivisão da tarefa em duas partes, uma inicial baseada no cenário típico, estático, realizada de modo *offline* e outra incremental com o fluxo de dados, realizada de modo *online* (BIFET, 2018). É também comum em um fluxo contínuo, incessante, que essa fase *online* se repita periodicamente, em um mecanismo de janelamento. Existem outras tantas propostas que consideram, por exemplo, janelas amortecidas, ou deslizantes. É comum o uso de janelamento para o tratamento da obsolescência de agrupamentos que perderam relevância com o tempo – *concept drift* (ZUBAROGLU e ATALAY, 2020).

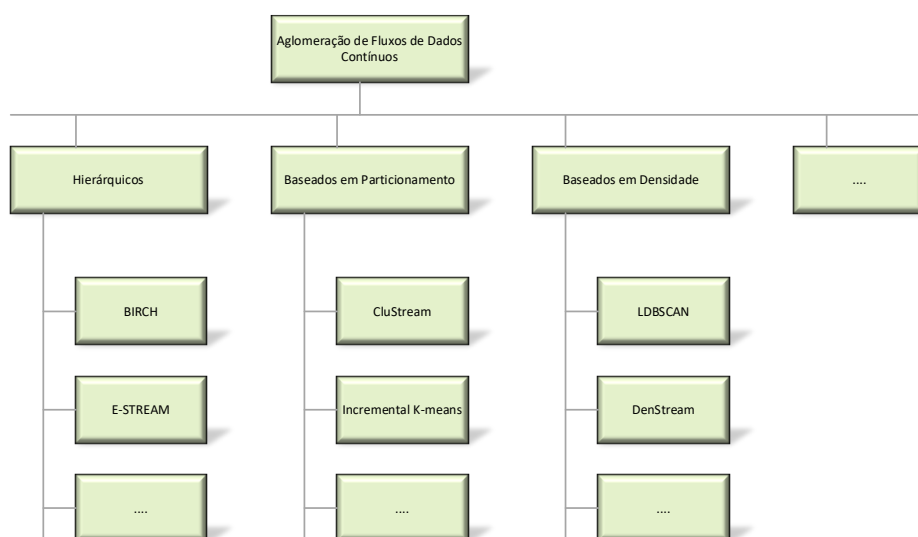


Figura 3 – Algoritmos de Agrupamento em Fluxos Contínuos de Dados. Fonte: Adaptado de ZUBAROGLU e ATALAY (2020).

As próximas subseções sintetizarão alguns dos principais métodos de agrupamento de *data streaming* considerados relevantes para o desenvolvimento desse trabalho. Para uma revisão mais detalhada a respeito do tema sugere-se a leitura de (ZUBAROGLU e ATALAY, 2020). Ademais, alguns problemas postos na literatura a respeito da utilização de métodos de agrupamento de fluxos de tempo real provenientes de redes de sensores distribuídos são apresentados. Por fim será apresentado um *Energy Management System* – EMS e seu papel com principal ferramenta de suporte a operação em tempo real em um centro de operação.

2.1. Algoritmos de Agrupamento em Fluxos Contínuo de Dados

2.1.1. *Incremental k-means*:

Esse algoritmo é uma variação do *k-means* tradicional que permite a atualização incremental dos grupos, ao invés de processar todo o agrupamento novamente. Isso torna o algoritmo útil para grandes conjuntos de dados, onde o reprocessamento pode ser caro em termos de tempo e recursos.

O método começa com a seleção aleatória de k centróides iniciais. Em seguida, para cada novo ponto de dados, o algoritmo encontra o centróide mais próximo do ponto de dados e o adiciona ao cluster correspondente. Depois, os centróides são recalculados com a média dos pontos de dados de cada grupo (ORDONEZ, 2003).

Uma vez que é baseado no *k-means*, esse algoritmo não é adequado para agrupar dados com formato não globular.

Uma implementação em Python desse algoritmo está disponível no pacote Cardinal (DAITAKU, 2023).

2.1.2. CluStream

O CluStream é um framework de mineração de fluxo de dados que pode ser utilizado para identificar grupos ou clusters em fluxos contínuos de dados proposto por VISWARUPAN, 2017. O objetivo desse framework é prover o agrupamento baseado em distância de fluxos de dados em larga escala e em evolução constante. Ao contrário de outras técnicas convencionais que tentam agrupar todo o fluxo em uma única passagem, o CluStream visualiza o fluxo como um processo em constante mudança ao longo do tempo.

Nesse método, há a divisão das tarefas em dois processos: a coleta e sumarização de dados estatísticos (online) e a análise dos componentes analíticos (offline). O processo online

é projetado para capturar as estatísticas essenciais dos fluxos de dados em evolução sem sacrificar a eficiência de espaço e de tempo subjacente do processo de agrupamento de fluxo. (AGGARWAL, 2003) e (AGGARWAL, 2014).

O CluStream usa uma janela de tempo piramidal e micro agrupamentos para garantir a coleta de dados escalável e manter a suficiência de dados necessária para o agrupamento efetivo. Em AGGARWAL, 2003 sugere-se que a estrutura desenvolvida para separar a coleta de dados estatísticos periódicos por meio de uma janela de tempo piramidal fornece um ambiente único para examinar outros paradigmas de agrupamento para fluxos de dados no futuro.

Uma implementação tradicional desse framework foi proposta em BIFET, 2010. O Massive Online Analysis – MOA é um ambiente de software para implementação e experimentação de aprendizado de máquina online para fluxos contínuos de dados. Esse ambiente foi escrito em linguagem Java, mas também possui interface para a linguagem Python. A implementação do CluStream presente no MOA é altamente eficiente e escalável para grandes conjuntos de dados.

Além dessa implementação, outra que se tornou bastante popular é a presente na biblioteca scikit-multiflow (MONTIEL et. al., 2018). Essa biblioteca é uma extensão do pacote scikit-learn e oferece uma implementação eficiente de CluStream que permite trabalhar com fluxos de dados, além de outras técnicas de mineração de fluxo de dados. Outra opção, também em linguagem Python, é a implementação presente no projeto ClusOpt Core (MACEDO et. al., 2020)

2.1.3. DenStream

O Density-based clustering over an evolving data stream with noise – DenStream (CAO et. al., 2006) é também uma proposta para mineração de dados em fluxos de dados em tempo real, entretanto, é baseado em um método de densidade que tende a lidar melhor com ruídos e *outliers*. Além disso, o método também é um algoritmo incremental, que processa os dados um a um e mantém uma representação compacta dos dados, o que permite otimizar o desempenho do algoritmo, sendo capaz de lidar com diferentes formatos de clusters e detectar a sua evolução no tempo.

A representação compacta do DenStream é baseada em microclusters, que são estruturas compactas que representam subconjuntos de dados do fluxo. Essa representação permite detectar mudanças na distribuição dos dados, sem a necessidade de manter todos os dados no

fluxo. O DenStream usa uma estratégia de atualização adaptativa dos parâmetros do algoritmo para lidar com a mudança na distribuição dos dados e minimizar o impacto dos ruídos.

Há uma variedade razoável de implementações em linguagem Python para esse algoritmo, como por exemplo biblioteca PyDenStream (PATEL, 2022), DenStream Clustering (WAYLON, 2020) e DenStream (MEMARI, 2020).

2.2. Agrupamento de Fluxos Contínuos Proveniente de Rede de Sensores Distribuídos

Em situações em que se pretende criar uma estrutura de agrupamento de fluxos de dados distribuídos, provenientes de sensores espalhados em uma larga área geográfica, é comum que se tenha problemas com o custo da comunicação de dados. Isso acontece, pois a abordagem mais convencional propõe a concentração dos dados em um sistema central para então processar.

Alguns métodos de processamento se propõem a reduzir esses custos de comunicação. Geralmente esses métodos propõem que os sensores distribuídos iniciem o processamento de agrupamento localmente e, em seguida, transmitam seu resultado para um servidor centralizado que por sua vez irá processar os agrupamentos de modo global. Em GAMA et. al., 2011 é proposto o algoritmo DGClust, que reduz a dimensionalidade e o custo de comunicação, permitindo que cada sensor local mantenha seu fluxo de dados online.

Cada novo sensor de dados dispara para o servidor central uma célula com dimensão reduzida, refletindo o estado atual do fluxo de dados no local. Assim, a cada momento, o servidor central tem o estado multivariado global de toda a rede.

Para evitar monitorar todos os estados possíveis, o servidor central mantém uma pequena lista de contadores dos estados globais mais frequentes. Finalmente, um algoritmo simples de agrupamento (k-means adaptativo) é aplicado aos pontos centrais dos estados frequentes para fornecer uma definição dos centros dos grupos.

2.3. Energy Management System – EMS

O EMS é um acrônimo em inglês de um Sistema de Gerenciamento de Energia. Esse tipo de sistema permite monitorar e determinar a condição operativa do sistema elétrico em tempo real, bem como controlar e otimizar o despacho de energia elétrica em uma rede de subestações e usinas (CEPEL, 2023). O objetivo do EMS é melhorar a eficiência energética e reduzir os custos associados à energia.

É comum que as ferramentas do EMS estejam disponíveis aos usuários também fora do ambiente de tempo real, permitindo a simulação e análise de uma condição operativa passada ou futura. Assim, tanto na sala de tempo real, quanto nas atividades de pré-operação e pós operação, o EMS dispõe de ferramentas para a análise da segurança operativa do sistema, bem como de ferramentas de otimização destinadas a auxiliar o usuário na tomada de decisões, sugerindo medidas corretivas e/ou preventivas (CEPEL, 2023).

No SEB o uso de EMS é bastante disseminado. Alguns exemplos de implementações desse tipo de sistema no SEB são o SIEMENS SPECTRUM POWER, o ELIPSE POWER EMS e o CEPEL SAGE EMS. Todos esses sistemas contam com uma série de ferramentas que tornam possível a operação do sistema elétrico proporcionando ferramentas gráficas de monitoração e controle, sinalizadores de alarmes, historiadores entre outros.

Está cada vez mais comum nesses sistemas a utilização de tecnologias que permitem se acoplar com ferramentas de análise de dados e inteligência de negócios. Por vezes essas ferramentas são próprias da solução e em outros casos são integradas a soluções externas como as ferramentas em nuvem. Em todos os casos, a integração dessas ferramentas com o sistema EMS se dá geralmente por um gerenciador de banco de dados tradicional (SQL) ou por servidores de fluxos contínuos em protocolos como o OPC-UA e o MQTT.

3. CLUALARM

3.1. Proposta de Desenvolvimento

O modelo de agrupamento inteligente dos alarmes em um centro de operação pressupõe o processamento contínuo e em tempo real, de modo que suas respostas sejam compatíveis com a janela de menos de 2 (dois) minutos para identificação de uma perturbação e de tomada de decisão de disponibilização ONS. Nesse caso, a equipe de operação deve ser capaz de reconhecer rapidamente, em questão de segundos, quais são as subestações e usinas envolvidas, quais são os equipamentos e funções de transmissão que foram desligadas e quais possuem bloqueios ou estão disponíveis para serem novamente energizadas, com o menor nível de esforço cognitivo e de sujeição a falhas humanas. Além do tempo limite para o tempo de resposta, o modelo deve levar em conta as restrições de processamento e memória e a necessidade de adaptação de sua resposta ao longo do tempo, com a chegada de novos dados.

Para tanto, propõe-se a integração com um sistema de gerenciamento de energia (EMS) largamente utilizado no SEB, Sistema Aberto de Gerenciamento de Energia – SAGE desenvolvido e mantido pelo Centro de Pesquisas de Energia Elétrica – CEPEL. O SAGE SEM possui um módulo historiador que registra em quase tempo real uma sequência de eventos de alarmes em um banco histórico PostgreSQL. A opção para realizar essa integração com o histórico e não utilizando-se um protocolo de tempo real, como o MQTT, deve-se ao fato de ser possível utilizar dados ocorridos no passado para o desenvolvimento e depuração do modelo. A construção do modelo, entretanto, continua com o foco em compatibilidade com o processamento em tempo real, sendo necessários ajustes apenas na parte da ingestão dos dados.

Com essa compatibilidade em mente foi proposto um *pipeline* de dados com ingestão assíncrona de dados do banco de dados relacional e o armazenamento em DATAFRAMES da biblioteca PANDAS, fazendo o uso da técnica de janelamento, limitando um estreito horizonte temporal tendo em vista a necessidade de processamento em um curto espaço de tempo. Ademais, entende-se que eventos relevantes a uma mesma perturbação no SEB geralmente acontecem sequencialmente na ordem de poucas dezenas de segundos.

Após a ingestão em memória dos eventos historiados pelo SAGE EMS, há necessidade de se identificar, a partir dos pontos de monitoramento, uma perturbação inicial do sistema de potência e, conseqüentemente, as subestações e os equipamentos. Com base nessa identificação, é que são calculadas a partir de grafos de conexões entre subestações e equipamentos, as

distâncias elétricas. Parte-se do pressuposto que eventos com proximidade elétrica em uma janela temporal estreita têm altíssima probabilidade de compor uma mesma perturbação.

É com base nessa assunção que se propõe a utilização de um algoritmo de agrupamento, não supervisionado, para identificação inteligente dos eventos de alarmes correlacionados eletricamente. A escolha da família de algoritmos para fase de testes inicial será a de agrupamento por densidade, uma vez que, a análise preliminar dos dados demonstrou que o formato dos grupos não é globular e o processamento hierárquico dependeria de metadados ora indisponíveis para serem profícuos.

Para visualização inicial dos dados na sala de operação os dados processados online serão processados em um banco de dados relacional para visualização em uma página web dinâmica ou em algum visualizador de painéis de mercado como o MS Power BI. A Figura 4 sintetiza o *pipeline* de dados proposto para o CluAlarm.

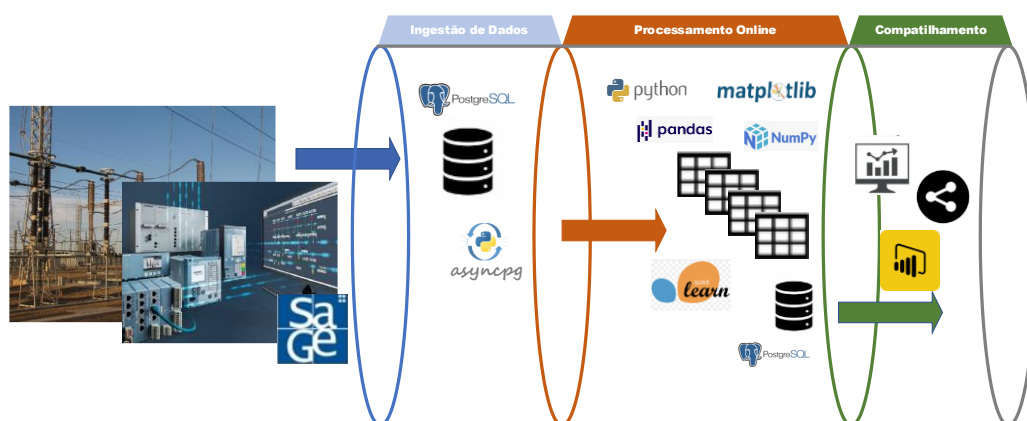


Figura 4 – Pipeline de Dados do CluAlarm. Fonte: Própria (2023).

3.2. Preparação para o desenvolvimento

Para o desenvolvimento dessa solução de software optou-se pela abordagem interativa proporcionada pelos notebooks Jupyter. A instalação foi baseada em um container Docker para ciência de dados. Esse container foi gerado pelo próprio projeto Jupyter e está disponível no Docker Hub (jupyter/datascience-notebook). Além da instalação ser bastante simples a conexão para leitura do banco de dados do SAGE EMS pode ser facilmente estabelecida por meio de um cliente VPN, que acessa o ambiente do centro de operações com segurança. Ademais, essa abordagem permitirá o lançamento em produção de modo bastante facilitado, uma vez que o

ambiente computacional do centro de operações em questão possui a capacidade de virtualização necessária.

3.3. Metodologia de Ingestão de Dados

A consulta ao banco de dados histórico do SAGE EMS foi realizada utilizando-se uma biblioteca assíncrona (asyncpg) para manutenção de uma janela de dados de um minuto. Essa biblioteca é uma interface ao banco de dados projetada especificamente para o PostgreSQL em conjunto com a biblioteca de processamento assíncrona do Python (asyncio). Sua instalação pode ser realizada de modo simples pelo comando de gerenciamento de pacotes Python (PIP) diretamente de um notebook Jupyter.

A consulta projetada para o CluAlarm traz as informações essenciais para a identificação da perturbação no SEB. A Figura 5, mostra um exemplo do resultado de uma consulta durante um evento de médio porte, o desligamento completo da subestação de 500 kV na cidade de Marabá, que envolveu outras três subestações adjacentes. O DATAFRAME descreve os eventos de alarmes registrados pelo sistema SAGE EMS em seu banco histórico durante a janela de um minuto. Os campos de cada registro representam, respectivamente, a data e hora do evento, o identificador do objeto de monitoramento, a descrição do que aconteceu no evento, a descrição do objeto de monitoramento o equipamento e a instalação associadas ao objeto de monitoramento.

	datetime	obj	ocr	descr	eqp	ins
1	2023-06-10 02:34:00+00:00	MBCA603FT ...	OcrUrg1Alarme03	MBCA6-03 Falta Tensão/Térmico/RFF Secc	MBCA6-03	MB
3	2023-06-10 02:34:04+00:00	MBCA603FT ...	OcrUrg1Alarme03	MBCA6-03 Falta Tensão/Térmico/RFF Secc	MBCA6-03	MB
5	2023-06-10 02:34:08+00:00	MBCA603FT ...	OcrUrg1Alarme03	MBCA6-03 Falta Tensão/Térmico/RFF Secc	MBCA6-03	MB
8	2023-06-10 02:34:12+00:00	BVCE201CTBALAR ...	OcrUrg1Alarme04	BVCE2-01 Ctrl Interno 1oEst SistB	BVCE2-01	BV
9	2023-06-10 02:34:12+00:00	MBCA603FT ...	OcrUrg1Alarme04	MBCA6-03 Falta Tensão/Térmico/RFF Secc	MBCA6-03	MB
...
1440	2023-06-10 02:35:24+00:00	MBQSACA220CMDAUT ...	OcrAdv1Atuado06	MBESA-QCA2 Comando Automático 220	MBESA-QCA2	MB
1441	2023-06-10 02:35:24+00:00	MBSD601FT ...	OcrUrg1Alarme05	MBSD6-01 Falta Tensão/Térmico/RFF Secc	MBSD6-01	MB
1442	2023-06-10 02:35:24+00:00	MBSD602FT ...	OcrUrg1Alarme05	MBSD6-02 Falta Tensão/Térmico/RFF Secc	MBSD6-02	MB
1443	2023-06-10 02:35:24+00:00	MBSY601FT ...	OcrUrg1Alarme05	MBSY6-01 Falta Tensão/Térmico/RFF Secc	MBSY6-01	MB
1444	2023-06-10 02:35:24+00:00	MBSY606FT ...	OcrUrg1Alarme05	MBSY6-06 Falta Tensão/Térmico/RFF Secc	MBSY6-06	MB

606 rows × 6 columns

Figura 5 – Resultado da Consulta ao Banco de Dados Histórico do SAGE EMS. Fonte: Própria (2023).

Nota-se que em uma janela de apenas um minuto o resultado da consulta possui cerca de 600 eventos de alarme. Nota-se, ainda, que dentro dessa lista existem diversos registros que guardam relação com a perturbação em questão e outros diversos registros que não devem ser levados ao foco do tomador de decisão em uma situação tão estressante como a de um desligamento dessa magnitude.

No contexto desse trabalho, a utilização da técnica de janelamento possui dois propósitos. O primeiro deles diz respeito a uma abstração da camada de ingestão de dados que é entregue para camada seguinte de processamento online. Em outras palavras, a entrada do processamento online sempre será um DATAFRAME da biblioteca PANDAS, com os dados de uma janela deslizante de um minuto que avança em tempo real. Assim em futuras implementações e melhorias, caso os dados não sejam obtidos do banco de dados históricos e sim diretamente de um fluxo contínuo de dados, por exemplo utilizando-se o protocolo MQTT, o resultado da saída da camada de ingestão sempre será o mesmo.

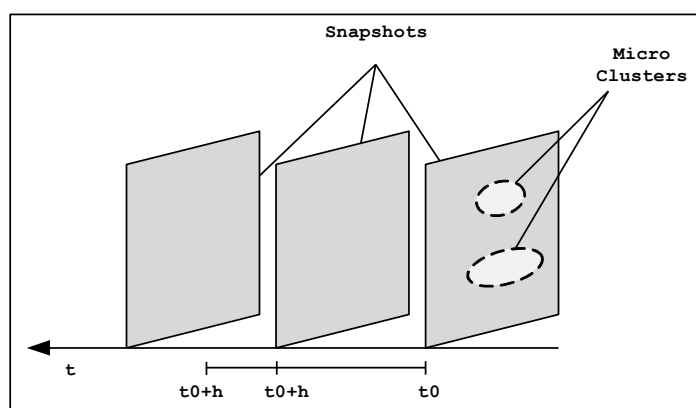


Figura 6 – Janelamento de Eventos de Alarmes em Tempo Real. Fonte: Adaptado de VISWARUPAN (2027).

O segundo propósito diz respeito a estreitar o universo de eventos que terão de ser processados na camada seguinte. É importante perceber que os eventos de uma perturbação no SEB têm uma probabilidade alta de ocorrerem em janelas de tempo estreita. Assim, uma vez identificada uma perturbação o processamento dos agrupamentos pode ocorrer de modo eficiente e com tempo de resposta adequado à tomada de decisão da equipe de operação. A Figura 6 ilustra o mecanismo de janelamento usado na ingestão de dados.

3.4. Metodologia de Processamento em Tempo Real

O primeiro passo na camada de processamento em tempo real é a identificação da perturbação. Nesse passo são caracterizadas como perturbações eventos onde há desligamento automático de um conjunto de funções de transmissão. Assim o algoritmo deve ser capaz de identificar e não considerar uma perturbação eventos de desligamento deliberado, por ação de controle do centro de operação. Os desligamentos automáticos são caracterizados pela atuação do sistema de proteção e pelos eventos de alarmes gerados por esses sistemas. Assim por meio de dessa lógica pode-se identificar os desligamentos automáticos e os equipamentos e as subestações envolvidas.

Cumprida essa etapa, é possível calcular com base em um evento de alarme com a atuação do sistema de proteção, quais são as distâncias elétricas entre cada um dos eventos, baseando-se em grafo pré-configurado da ligação entre equipamentos da rede elétrica. Esse grafo foi exportado em um arquivo tipo CSV do SAGE EMS e sua leitura e transformação em um dicionário chave valor é realizado somente na inicialização do software. A biblioteca QUEUE foi utilizada para cálculo da distância elétrica a partir do dicionário que representa a rede elétrica.

Outra dimensão utilizada para representar a distância elétrica se baseia em um outro grafo pré-configurado da ligação entre as instalações da rede elétrica. Apesar de correlacionada com a primeira dimensão, a distância entre as instalações traz uma noção de direcionalidade para a distância elétrica. Ademais, essa direcionalidade permite melhor visualização dos dados em um gráfico bidimensional.

Uma vez identificada a perturbação, as instalações envolvidas e as distâncias elétricas é realizado um agrupamento não supervisionado dos eventos de alarme. A biblioteca SCIKITLEARN foi utilizada e o modelo selecionado foi o DBSCAN sobre os eventos presentes no janelamento, que realiza o agrupamento baseando-se nas duas dimensões supracitadas. Os diversos parâmetros do DBSCAN foram ajustados com base em leituras históricas de perturbações ocorridas durante o ano de 2023. Assim como proposto para o DenStream, o DBSCAN foi considerado apropriado uma vez que a distribuição dos dados não é globular. Ao contrário do DenStream, o processamento mais simples, sem adaptação dos parâmetros, demonstrou-se suficiente, como mostrado na sessão de resultados.

É importante notar que o algoritmo de agrupamento só roda quando é identificada uma ou mais perturbações na janela de um minuto. Assim, boa parte do tempo a rotina principal de

processamento online demora poucos milissegundos para rodar. Entretanto, alguns segundos são gastos no cálculo das distâncias e no próprio agrupamento não supervisionado.

A evolução dos grupos de eventos de alarmes (clusters) identificados à medida que chegam novos dados pode ser realizada considerando-se apenas o cálculo da distância elétrica para o centro dos cluster obedecendo-se a um critério fixo, não necessariamente havendo a necessidade de rodar novamente o algoritmo de agrupamento. Também na sessão de resultados será avaliado o custo-benefício da utilização desse critério fixo ou de reiterar a execução do DBSCAN.

A fase de processamento em tempo real termina com uma lógica de identificação da imutabilidade dos clusters de alarmes identificados. Assim, é interrompida a persistência dos dados de uma perturbação no repositório de saída. Inicialmente, esse repositório de saída reside no sistema de arquivos do container Docker, em formato CSV. Entretanto, propõe-se utilizar um banco de dados relacional para armazenar os clusters identificados para facilitar a ingestão para camada subsequente de apresentação de dados.

3.5. Metodologia de Apresentação e Compartilhamento

É bastante relevante para o resultado desse trabalho que a saída do processamento em tempo real possa ser apresentada visualmente para equipe de operação a fim de auxiliar no processo de tomada de decisão dia a dia no acompanhamento da operação do sistema elétrico. Assim propõe-se inicialmente uma apresentação em painel baseado em interface via navegador de Internet, por meio de um servidor de páginas dinâmicas ou por meio de uma solução de painéis de apresentação como o Power BI.

Essa abordagem inicial tem o objetivo de que a solução prove seu valor dentro da sala de operações. Futuros desenvolvimentos poderão ser realizados com o intuito de mostrar na mesma interface do SAGE EMS os clusters de alarmes, de modo a facilitar ainda mais a utilização e o aproveitamento da solução proposta.

4. AVALIAÇÃO DOS RESULTADOS

4.1. Descrição dos conjuntos de dados e do processamento

4.1.1. Identificação de perturbações

Conforme mencionado, no sistema elétrico é comum que por conveniência ou necessidade operacional equipamentos sejam ligados e desligados de modo deliberado, por ação de controle de um operador. As perturbações são então normalmente caracterizadas por desligamentos automáticos, que ocorrem quando da ação de sistemas de proteção e automação que visam preservar o funcionamento do sistema elétrico como um todo, fazendo com que desligamentos pontuais não se alastrem por uma grande área e se tornem um blecaute, por exemplo.

Assim, a principal entrada de dados é o fluxo contínuo de alarmes provenientes do sistema SAGE EMS. Nesse fluxo, o conjunto de dados consiste basicamente dos seguintes itens:

- Estampa de tempo de quando ocorreu o alarme (Data/Hora);
- O identificador (Objeto) e a descrição de um ponto de monitoramento (Descrição);
- As transições de estado desses pontos de monitoramento que geraram o alarme (Ocr);
- O equipamento ao qual o ponto de monitoramento está associado (Eqp);
- A instalação (subestação ou a usina) ao qual o equipamento pertence (Ins).

A Tabela 2 mostra um exemplo com os alarmes referentes ao desligamento automático do reator de linha VCRE7-01 na subestação Vila do Conde (VC), no estado do Pará. A abertura dos disjuntores associados VCDJ7-01 e VCDJ7-02 ocorreu pela atuação do relé de gás, proteção impeditiva, intrínseca ao equipamento reator. O recorte abaixo mostra ainda os bloqueios de fechamento dos disjuntores, que denota que o operador não pode disponibilizar imediatamente o equipamento para reenergização.

Tabela 2 – Agrupamento Tradicional e em Fluxos de Dados Contínuos

Data/Hora	Objeto	Descrição	Ocr	Eqp	Ins
2023-06-02 03:24:16	VCDJ701	Posição Disjuntor	Abriu	VCDJ7-01	VC
2023-06-02 03:24:16	TCDJ726POS	Posição Disjuntor	Abriu	TCDJ7-26	TC
2023-06-02 03:24:16	VCRE701632	63 Relé de Gás 2 Est	Desligou Imp	VCRE7-01	VC
2023-06-02 03:24:23	TCDJ727POS	Posição Disjuntor	Abriu	TCDJ7-26	TC
2023-06-02 03:24:24	VCDJ70186	86 Bloqueio Fechamento	Bloqueou	VCDJ7-01	VC
2023-06-02 03:24:24	VCDJ702	Posição Disjuntor	Abriu	VCDJ7-02	VC
2023-06-02 03:24:24	VCDJ70286	86 Bloqueio Fechamento	Bloqueou	VCDJ7-02	VC

Os 07 (sete) alarmes acima destacados foram registrados em um período de cerca de 08 (oito) segundos, em meio a outros 32 (trinta e dois) alarmes de menor relevância ocorridos em outras 08 (oito) subestações diferentes.

Assim um dos desafios iniciais do processamento é reconhecer quando ocorreu uma perturbação, quais são as instalações e os equipamentos associados. Para em seguida, por meio de clusterização, saber se há relação de proximidade elétrica, portanto, pertencem ao mesmo grupo de perturbações.

4.1.2. Proximidade Elétrica entre Instalações:

A rede elétrica do SIN pode ser representada por um grande grafo onde todos os equipamentos potência dentro das instalações são interconectados e onde as linhas de transmissão fazem a conexão entre essas instalações. É comum no setor, especialmente para operação do sistema elétrico, que essas conexões sejam representadas em diagramas unifilares.

A Figura 7 apresenta o recorte de dois diagramas unifilares de duas instalações no estado do Pará: a subestação Vila do Conde e a subestação de Tucuruí. Ambos os recortes são referentes ao nível de tensão de 500 kV. Nota-se, pelos recortes que há ao menos uma linha de transmissão que conecta essas instalações, que é o circuito número 1, codificado nos diagramas como TCVC-LT7-01. Essa conexão entre as instalações permite dizer que são instalações adjacentes.

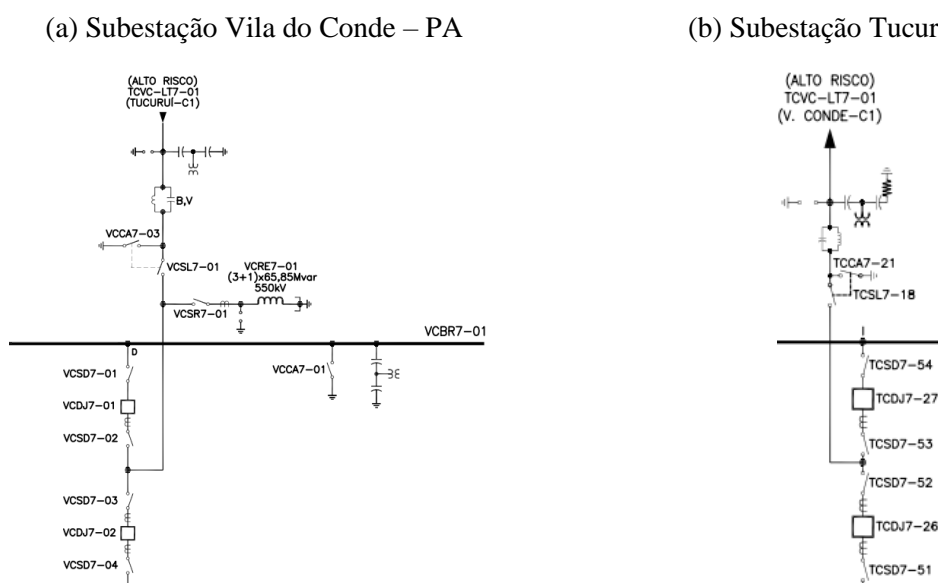


Figura 7 – Recorte dos Diagramas Unifilares Operacionais. Fonte: Eletrobras Eletronorte (2023).

Para se obter uma medida de distância elétrica entre subestações, inicialmente, sintetizou-se, a partir dos diagramas unifilares, um grafo simplificado que descreve a adjacência entre as instalações. A Figura 8 mostra um recorte desse grafo das instalações de 500 kV operadas pela Eletrobras Eletronorte nos estados do Pará, Maranhão e Tocantis.

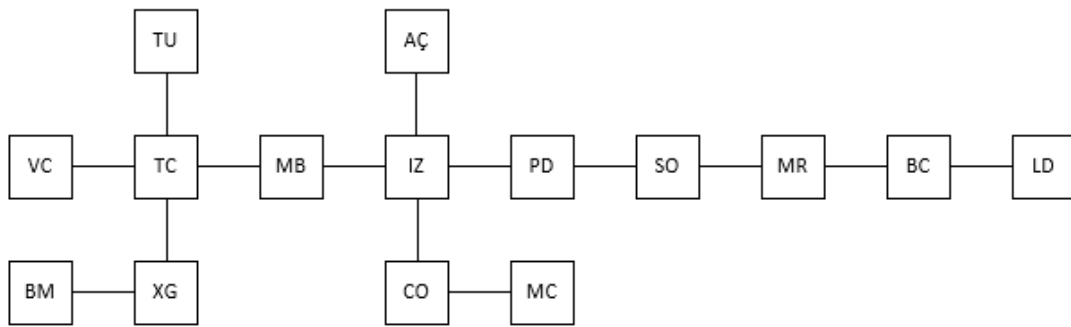


Figura 8 – Recorte Instalações 500 kV PA, MA e TO. Fonte: Própria (2023).

No diagrama acima, tomando-se como referência a subestação Vila do Conde (VC), intuitivamente há uma distância de 5 (cinco) saltos até subestação Santo Antônio dos Lopes (SO). Entretanto, a contagem de saltos não se mostrou adequada como medida de distância nas experiências com a clusterização, uma vez que não há uma relação biunívoca entre instalações e os valores de saltos contados no diagrama. Por exemplo, a subestação Miracema (MC) também está a uma distância de 5 (cinco) saltos da mesma referência.

Para desfazer a ambiguidade e manter a simplicidade, propôs-se representar o grafo com distâncias unitárias para o tronco tido como principal e, em caso de bifurcação, distâncias unitárias para cada ramo em uma segunda dimensão. Considerando as instalações mencionadas, a tupla cartesiana (0,0) poderia representar a subestação VC, (5,0) a subestação SO e (3,2) a subestação MC. Nesses casos as distâncias são calculadas pela distância euclidiana entre os pares de tuplas, ou seja, igual a 5 para distância relativa entre as subestações VC e SO e igual a $\sqrt{(3)^2 + (2)^2} = 3,6$.

Com essa modelagem, cada nova bifurcação, dentro de uma bifurcação prévia é tratada como uma nova dimensão. Para o grafo de todas as 70 instalações foi então gerada uma matriz com tuplas com 7 dimensões. O cálculo das distâncias euclidiana D_I entre as instalações as i e j nesse hiperplano é dada pela equação (1):

$$D_I(i, j) = \sqrt{\sum_{d=0}^7 (x_{0,j} - x_{0,i})^2} \quad (1)$$

4.1.3. Proximidade Elétrica entre Equipamentos

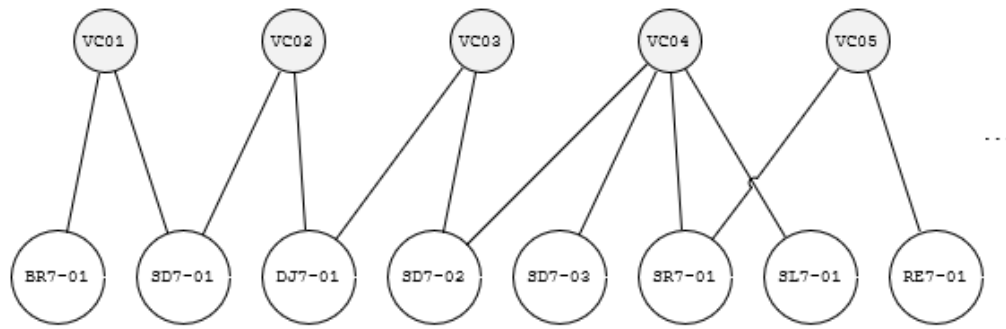
Além da vista simplificada da distância elétrica entre subestações, também é possível a partir dos diagramas unifilares do sistema elétrico, a construção de um grafo mais detalhado onde são representadas todas as conexões entre os equipamentos. O sistema SAGE EMS possui em sua modelagem uma tabela com essas conexões na forma de grafo bipartido.

Retomando os recortes da Figura 7, no SAGE EMS essas ligações entre equipamentos são descritas na forma de tabela, onde são representados os nós elétricos (lig.id) e os equipamentos os quais estão conectados nesses nós. A Tabela 3 mostra um exemplo de configuração para o recorte da subestação Vila do Conde mencionado acima. O conteúdo dessa tabela pode também ser visualizado na forma do grafo bipartido da Figura 8(a). Ademais, a título de exemplo, na Figura 8(b) pode ser visualizado em destaque vermelho a representação do nó elétrico com identificador VC04.

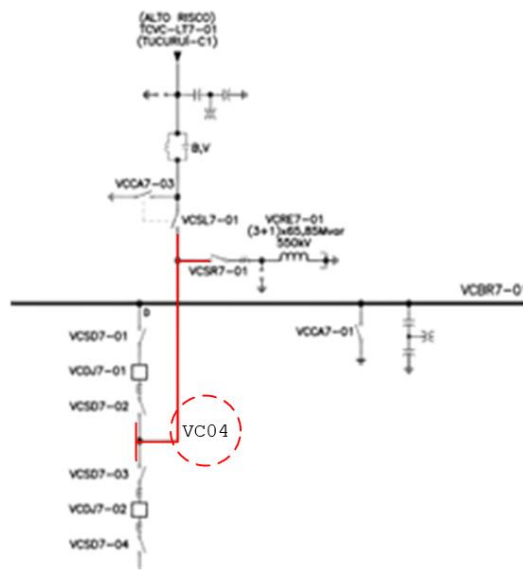
Tabela 3 – Ligações SAGE EMS

Identificador da Ligação (lig.id)	Equipamento
VC01	VCBR7-01
VC01	VCSD7-01
VC02	VCSD7-01
VC02	VCDJ7-01
VC03	VCDJ7-01
VC04	VCSD7-02
VC04	VCSD7-03
VC04	VCSR7-01
VC04	VCSL7-01
VC05	VCSR7-01
VC05	VCRE7-01

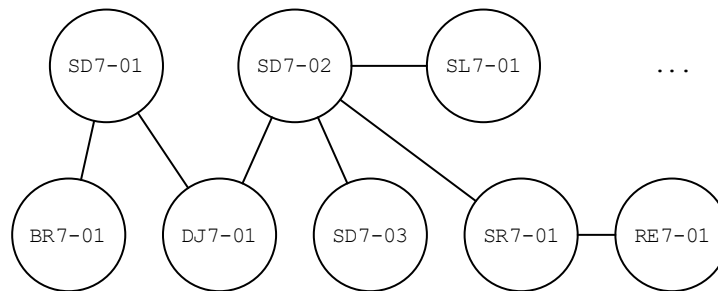
(...)



(a) Grafo bipartido.



(b) Destaque do nó VC04 no diagrama unifilar.



(c) Grafo Direto.

Figura 8 – Equipamentos e nós elétricos. Fonte: Própria (2023).

Para realizar o cálculo da distância elétrica entre equipamentos optou-se pela utilização de um algoritmo inspirado na busca em largura (*Breadth-First Search* – BFS). Entretanto, para utilizar esse tipo de algoritmo é necessário transformar a representação bipartida proveniente do SAGE EMS em um grafo direto. Para tanto, a Tabela 3 foi transformada em um dicionário

(Python) que representa o grafo direto mostrado na Figura 8(c). Cada chave desse dicionário representa um equipamento e os valores representam os equipamentos conectados.

O cálculo da distância entre todos os equipamentos do grafo (`input_graph`) a partir de um equipamento origem (`source`) é dados pelo Algoritmo 1. O retorno desse algoritmo é também um dicionário onde as chaves são os equipamentos e os valores são as distâncias elétricas entre os equipamentos.

Algoritmo 1 – Distância entre equipamentos.

```
def calculate_distance(input_graph, source):
    Q = Queue()
    distance_dict = {k: 999999999 for k in input_graph.keys()}
    visited_vertices = list()
    Q.put(source)
    visited_vertices.append(source)
    while not Q.empty():
        vertex = Q.get()
        if vertex == source:
            distance_dict[vertex] = 0
        for u in input_graph[vertex]:
            if u not in visited_vertices:
                # update the distance
                if distance_dict[u] > distance_dict[vertex] + 1:
                    distance_dict[u] = distance_dict[vertex] + 1
                Q.put(u)
                visited_vertices.append(u)
    return distance_dict
```

4.1.4. Janelamento e agrupamento inteligente

Uma vez identificadas as perturbações, as instalações e equipamentos envolvidos, o próximo passo do processamento inteligente é identificar se as perturbações compõem um único grupo ou se caracterizam perturbações independentes. Para tanto, parte-se do pressuposto de que perturbações distintas podem ser correlacionadas pelas dimensões temporal e distal, essa última representada pela proximidade elétrica.

Em outras palavras, perturbações que acontecem em instantes de tempo próximos tendem a compor grupos. De mesmo modo, perturbações em que há uma curta distância elétrica também tendem a fazer parte de um mesmo grupo.

Para abordar a dimensão temporal optou-se pelo uso da técnica de janelamento. Essa abordagem em tese possui duas vantagens nesse contexto. A primeira é que em uma mesma

janela estarão presentes os alarmes dos conjuntos de perturbações de interesse. O segundo é a diminuição substancial do volume de dados que serão processados de modo on-line, fazendo como que o tempo de resposta do algoritmo seja condizente com a operação de tempo real do sistema elétrico.

O janelamento utilizado tem como referência o relógio do sistema operacional que roda o algoritmo, sincronizado com a fonte de dados do SAGE EMS. A janela deslizante consiste portanto do conjunto de alarmes que aconteceram no último minuto. É em cima dessa janela que são realizadas as operações de identificação das perturbações, equipamentos e instalações envolvidas.

A abordagem da dimensão distal foi realizada com duas dimensões, a distância entre instalações e a distância entre equipamentos. Apesar de correlacionadas a utilização das duas e não somente da última trouxe ganhos para o processamento mais assertivo do algoritmo de clusterização.

O agrupamento inicialmente foi realizado com as 3 dimensões, a temporal e as duas distais, entretanto, também descartou-se a dimensão temporal pois a mesma estava agregando ambiguidades no processamento. Foi experimentado o algoritmo KNN, entretanto, como o padrão dos grupos não é globular, o algoritmo que melhor se apresentou foi o DBSCAN da biblioteca SKLEARN.

A parametrização que melhor funcionou para o DBSCAN foi a utilização das distâncias absolutas sem a padronização dos valores de distância (STANDARD SCALER). Isso se deu porque o ajuste dos parâmetros *EPS*, que controla a distância máxima entre dois pontos de elementos de um mesmo grupo, e *MIN_SAMPLES*, que define o número mínimo de pontos para que seja caracterizado um mesmo grupo, são dificultados pela utilização dos valores padronizados. Um resultado efetivo foi obtido, portanto, com os parâmetros $EPS = 3$ e $MIN_SAMPLES = 2$.

4.1.5. Atualização dos agrupamentos em tempo real

Uma vez identificadas perturbações e realizado o primeiro agrupamento, é necessário continuar de modo on-line a processar os dados que continuam chegando. Assim, se dentro da mesma janela em que aconteceu a primeira perturbação, acontecerem uma sequência de outras perturbações, há que se atualizar os agrupamentos efetuados na primeira passagem. Desse modo, em questão de poucos segundos há a possibilidade de apresentar o resultado inicial e atualizá-lo em tempos compatíveis com o requisito operacional, ou seja, em menos de 2 minutos.

A complexidade computacional do DBSCAN é em média linear em relação ao volume de dados em processamento. No caso em pauta, mesmo sendo necessário uma nova execução para atualização dos agrupamentos, foi eficaz a utilização desse algoritmo uma vez que o mesmo está rodando em um volume de dados reduzido, dentro de uma janela de apenas 1 minuto, e depois de ter passado pelos filtros de identificação das perturbações, equipamentos e instalações envolvidas.

A atualização dos clusters continua acontecendo até um limite de tempo. Esse parâmetro foi ajustado inicialmente para 300 segundos do tempo corrente. Entretanto, houve atualizações desnecessárias, uma vez que o algoritmo passou a observar as manobras de recomposição. O valor final desse parâmetro foi ajustado para 180 segundos após a primeira clusterização com o DBSCAN.

4.1.6. Exportação dos dados e Apresentação

Uma vez identificada estagnação dos agrupamentos, após o processamento os dados são exportados em forma tabela novamente para o banco de dados relacional (PostgreSQL). Os dados exportados são semelhantes àqueles da Tabela 2. Duas colunas apenas são acrescentadas, uma com o identificados do conjunto de perturbações e outra com os agrupamentos. Esses dados são então consumidos e apresentados em um painel dinâmico (dashboard).



(a) Painel Dinâmico CluAlarm

00:24:16	Alarme	UGSB601FT	UGSB6-01	Falta Tensão/Térmico/RFF Secc
00:24:16	Alarme	UGSB620FTTRFF	UGSB6-20	Falta Tensão/Térmico/RFF Secc
00:24:16	Aberto	TCDJ701	TCDJ7-01	Posição Disjuntor
00:24:16	Desligamento	VCRE701632	VCRE7-01	63 Relé de Gás 2 Est
00:24:16	Alarme	VCCT701ETDMA	TCVCLT7-01	21A Transmissão TDD Mantido
00:24:16	Alarme	VCCT701ETDMP	TCVCLT7-01	21F Transmissão TDD Mantido
00:24:17	Retornou a regiao normal (-2.00)	JQBACCRC21	JQB8-CC	Corrente Corrente Ret 1
00:24:15	Retornou a regiao normal (-0.04)	COBR702MFDFFA	COBR7-02	PMU Taxa Variação Frequência FB
00:24:15	Retornou a regiao normal (-0.01)	JUBR602MFDFFB	JUBR6-02	PMU Taxa Variação Frequência FB
00:24:15	Retornou a regiao normal (0.00)	JUVDL1601MFDFFB	JUVD-L16-01	PMU Taxa Variação Frequência FB
00:24:15	Ultrapassou urgencia superior (0.03)	JUVDL1602MFDFFB	JUVD-L16-02	PMU Taxa Variação Frequência FB
00:24:15	Retornou a regiao normal (-0.04)	PDBR701MFDFFB	PDBR7-01	PMU Taxa Variação Frequência FB
00:24:15	Retornou a regiao normal (-0.08)	PDBR702MFDFFB	PDBR7-02	PMU Taxa Variação Frequência FB
00:24:15	Ultrapassou advertencia superior (60.46)	PDSOLT701MFDFFB	PDSO-LT7-01	PMU Frequência FB
00:24:15	Retornou a regiao normal (-0.03)	PDBR1L1701MFDFFB	PDBR-L17-01	PMU Taxa Variação Frequência FB
00:24:15	Retornou a regiao normal (-0.04)	PDTSL1701MFDFFB	TSPD-L17-01	PMU Taxa Variação Frequência FB
00:24:15	Retornou a regiao normal (-0.04)	PDTSL1702MFDFFB	TSPD-L17-02	PMU Taxa Variação Frequência FB
00:24:15	Ultrapassou urgencia superior (0.01)	SMBR601MFDFFB	SMBR6-01	PMU Taxa Variação Frequência FB
00:24:15	Ultrapassou urgencia superior (0.01)	SMBR602MFDFFB	SMBR6-02	PMU Taxa Variação Frequência FB
00:24:15	Ultrapassou urgencia superior (0.01)	SMPVL1601MFDFFB	SMPV-LT6-01	PMU Taxa Variação Frequência FB
00:24:15	Ultrapassou urgencia superior (0.03)	SMPVL1603MFDFFB	SMPV-LT6-03	PMU Taxa Variação Frequência FB
00:24:20	Normal	BPDJ201FTDM	BPDJ2-01	Falta Tensão/Térm/Def Motor
00:24:21	"NUL" Transição Normal para Alarme	CKAAT4VDF2FB	CKAT6-04	Vent Falta Tensão FB
00:24:21	"NUL" Transição Normal para Alarme	HBCA601FT	HBCA6-03	Falta Tensão/Térmico/RFF Secc
00:24:23	Alarme	SPDB50174FTM	SPDB5-01	Falta Tensão Motor
00:24:23	Alarme	SRCA601FT	SRCA6-01	Falta Tensão/Térmico/RFF Secc
00:24:23	Alarme	TCDJ72694B2	TCDJ7-26	94 Relé de Trip B2
00:24:23	Aberto	TCDJ72774B1A	TCDJ7-27	Posição Disjuntor
00:24:23	Normal	TCDJ72774B1A	TCDJ7-27	74 Supervisão Bobina 1 Abertura
00:24:23	Alarme	TCDJ72694B1	TCDJ7-26	94 Relé de Trip B1
00:24:23	Normal	TCDJ72774B2A	TCDJ7-27	74 Supervisão Bobina 2 Abertura

(b) Recorte da lista de Alarmes sem Filtros (SAGE EMS).

Figura 9 – Apresentação e Compartilhamento. Fontes: Própria (2023) e Eletrobras Eletronorte (2023).

O painel foi construído utilizando-se o Microsoft Power BI e contém as informações básicas sobre as perturbações. No primeiro quadro, tem-se uma caixa de seleção com as últimas perturbações. Uma vez selecionada a perturbação, nos demais quadros tem-se um contador de instalações envolvidas, se há desligamentos com impedimentos de energização, quantos equipamentos foram envolvidos e quantos agrupamentos foram identificados. Ademais, o painel mostra no mapa geográfico os estados da federação que foram afetados.

O design do painel, mostrado na Figura 9(a) tem a intenção de permitir ao operador do sistema elétrico a visualização rápida e concisa da informação necessária para tomada de decisão, sem a necessidade de um esforço cognitivo significativo. A Figura 9(b) mostra a lista de alarmes no sistema SAGE EMS no mesmo momento da ocorrência.

4.2. Configuração experimental

A proposta do CluAlarm é para o que mesmo opere em tempo real. Assim, como mencionado anteriormente, para modelagem e validação de sua eficácia, sua camada de ingestão de dados teve que ser adaptada para consumir dados históricos do SAGE EMS. Para realização da validação foram utilizados os dados históricos no período de 01/06/2023 até 31/07/2023 em que foram registradas 67 desligamentos automáticos de equipamentos da rede básica operados pela Eletronorte.

Além disso, foi utilizada nessa avaliação a mesma parametrização apresentada na seção anterior para os algoritmos para identificação das perturbações, equipamentos e instalações envolvidas; janelamento; cálculo das distâncias elétricas entre equipamentos e instalações; bem

como o ajuste dos parâmetros do algoritmo de agrupamento DBSCAN. O propósito é avaliar a eficácia dessa paremetrização e replicá-la para o processamento em tempo real.

A principal base para comparação da eficácia da ferramenta é o relatório de perturbações, produzido pelas equipes de tempo real e consolidados com as equipes de análise de desempenho da operação e manutenção. Um extrato dos meses em avaliação foi obtido do sistema de gestão da pós operação. Esse extrato traz quais os horários que foram registradas as perturbações, as instalações e equipamentos envolvidos.

Deste modo é possível uma comparação entre os grupos identificados pelas ferramentas e aquelas registradas no sistema de gestão da empresa. Para essa comparação foram filtrados apenas os desligamentos automáticos de ativos operados ou de propriedade da Eletrobras Eletronorte.

4.3. Resultados e Discussão

4.3.1. Resultados experimentais preliminares

Como mencionado, a gravidade ou porte de uma perturbação no sistema elétrico de potência pode ser avaliada em relação a diversos tipos de critérios, como por exemplo, a quantidade de equipamentos e instalações envolvidas, se há ou não impedimentos para o religamento, se houve ou não corte de carga. Assim a maior parte das perturbações não é de grande porte, como a exemplificada na sessão anterior, com o desligamento automático da linha TCVC-LT7-01 em função de uma proteção intrínseca do reator, em 02/06/2023.

Mesmos nesses casos mais simples o agrupamento inteligente e a apresentação em um painel intuitivo se mostram um caminho para reduzir o esforço cognitivo do operador de tempo real. Entretanto, entende-se que são nos casos de perturbações de maior porte que a ferramenta se justifica, especialmente com um número elevado de equipamento e instalações envolvidas.

Além das perturbações de grande porte, a ferramenta tem um grande potencial de facilitar a percepção e a tomada de decisão na operação de tempo real, quando acontecem perturbações subsequentes mas que não possuem correlação. Por exemplo, num espaço de poucos minutos há um desligamento de uma linha de transmissão no estado do Maranhão e outro de um banco de capacitores em Mato Grosso. As duas perturbações aconteceram em regiões elétricas distintas e devem ser facilmente isoladas e avaliadas para tomada de decisão.

4.3.1.1. Perturbação de Grande Porte

No período avaliado, separou-se uma perturbação de grande porte que ocorreu no dia 09/06/23 envolvendo as regionais do Pará e do Maranhão. O primeiro desligamento identificado

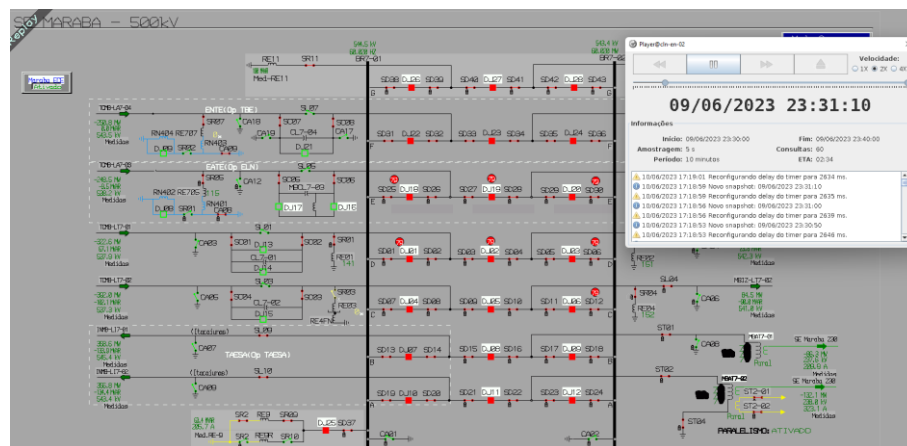
ocorreu às 23:34:19. Menos de 50 segundos depois já estava disponível a primeira versão do agrupamento, que seguiu sendo atualizado até 23:38:36 devido a seguidos desligamentos relacionados a mesma perturbação.

A Figura 10 (a) mostra a última versão com o resumo da perturbação com 61 equipamentos envolvidos em 4 instalações diferentes, sendo identificados 5 desligamentos impeditivos. Apesar de a última versão do agrupamento, às 23:38:36 possuir um conjunto razoável de 141 alarmes, o algoritmo foi capaz de agrupá-los em um mesmo grupo, correlacionado pela distância elétrica entre os sucessivos desligamentos.

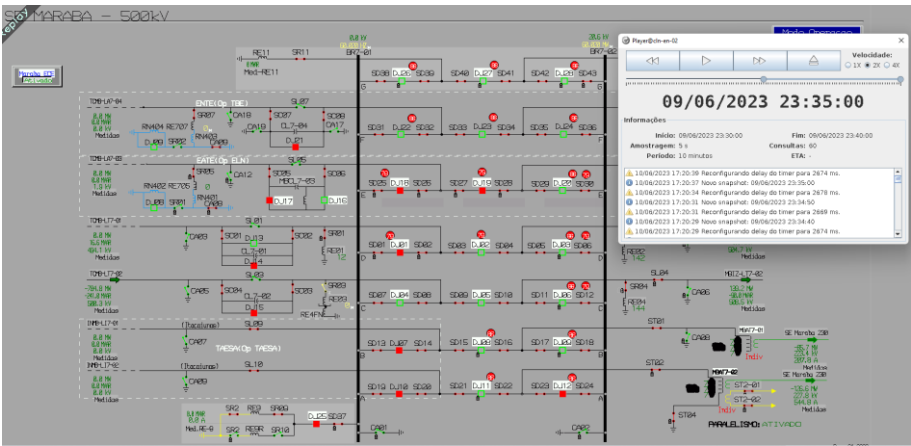
As Figuras 10(b) e 10(c) mostram recortes retirados da ferramenta de reprise (replay) do SAGE EMS. Na Figura 10(b) o estado do setor de 500kV na subestação Marabá antes da ocorrência e na Figura 10(c) estado com diversos disjuntores abertos.



(a) Painel Dinâmico CluAlarm



(b) Recorte anterior à perturbação (SAGE EMS).



(c) Recorte durante à perturbação (SAGE EMS).

Figura 10 – Perturbação de Grande Porte. Fontes: Própria (2023) e Eletrobras Eletronorte (2023).

4.3.1.2. Perturbações consecutivas em área elétricas disjuntas

Ainda no período avaliado, outro conjunto de perturbações teve destaque por apresentar dois conjuntos de desligamentos que ocorreram em áreas disjuntas, ou seja, as distâncias entre essas perturbações fizeram com que houvesse mais de um agrupamento identificado. Nesse caso, foi identificado o desligamento automático de uma linha de transmissão no estado do Tocantis e uma peturbação envolvendo a proteção de barras da subestação Tucuruí, que estava em manutenção.

A Figura 11 mostra mais uma vez o painel do CluAlarm sintetizando as perturbações concecutivas. Nota-se que foram identificados 2 grupos marcados na coluna CluAlr e classificados para facilitar a percepção do operador em tempo real.



Figura 11 – Perturbação Consecutivas Disjuntas. Fontes: Própria (2023) e Eletrobras Eletronorte (2023).

4.3.2. Comparação com os relatórios de perturbação do sistema de gestão

Como colocado anteriormente, de acordo com o sistema de gestão de perturbações da Eletrobras Eletronorte, foram registrados no período em avaliação, junho e julho de 2023, um total de 67 desligamentos automáticos, que geraram 26 relatórios de perturbações tendo em vista a data e hora das mesmas. Destaca-se que foram considerados para essa avaliação apenas os desligamentos automáticos de ativos próprios e operados pela empresa.

Desses 26 grupos de desligamento, a ferramenta foi capaz de identificar 24, ou seja, 92% dos grupos de perturbações foram reconhecidos de modo automático. A avaliação é tida como satisfatória, entretanto, há dois grupos que não foram identificados e que mostram duas limitações intrínsecas à ferramenta.

A identificação de uma perturbação causada por um ou mais desligamentos automáticos passa pela identificação da atuação dos sistemas de proteção e controle local das subestações e usina. Assim, a primeira limitação se refere a uma situação em que há falha de comunicação entre a instalação e o centro de operação, fazendo com que não chegue esse registro de atuação à ferramenta. A segunda limitação se refere a uma perturbação que teve origem na atuação do sistema de proteção e controle de um ativo que não é de propriedade da Eletronorte. Essa atuação externa também não é registrada e, portanto, não dispara a identificação automática da perturbação.

4.3.3. Número de alarmes

Um dos intúitos da ferramenta é a diminuição do esforço cognitivo do operador na identificação das perturbações, das instalações e dos equipamentos envolvidos. Os grupos de perturbações avaliados apresentaram em torno de 108 alarmes em média. A ferramenta foi capaz de reduzir significativamente o número de alarmes apresentados para o operador na ferramenta de visualização.

O número médio de alarmes apresentados é de 17, representando uma diminuição de aproximadamente 83%. Na maior perturbação avaliada, mostrada na Figura 10, a ferramenta ainda teve que apresentar 119 alarmes ao operador, entretanto, o total de alarmes nessa perturbação de grande porte, foi de 897.

4.3.4. Tempos para apresentação preliminar do agrupamento de perturbações

A avaliação dos tempos de resposta da ferramenta é fundamental, uma vez que há um requisito de tempo bastante estreito para tomada de decisão do operador em tempo real. Assim, uma vez identificada a ocorrência avaliou-se os tempos de resposta inicial do algoritmo, isto é,

a primeira atualização do DBSCAN que será disponibilizada ao operador por meio do painel dinâmico.

O histograma da Figura 12 mostra a distribuição dos tempos de resposta inicial. Nota-se que mais de 75% das vezes o algoritmo respondeu em um tempo inferior a 16 segundos e que, ademais, em nenhuma das avaliações o algoritmo ultrapassou 50 segundos para a apresentação inicial. A Tabela 4 mostra os resultados descritivos dessa distribuição de dados.

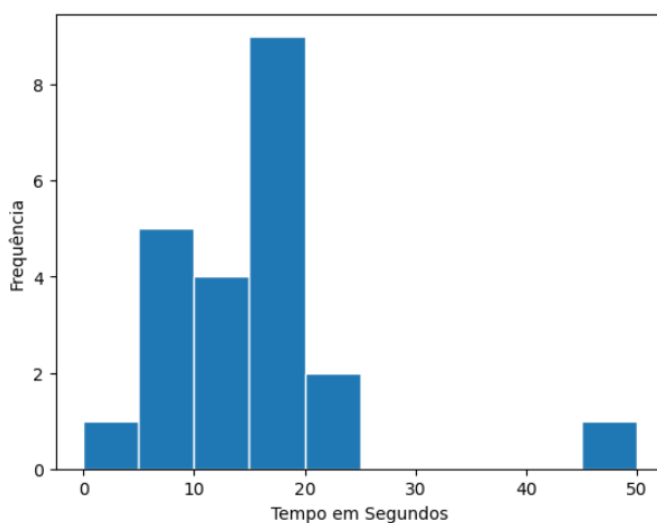


Figura 12 – Distribuição Tempo de Resposta Inicial. Fontes: Própria (2023).

Tabela 4 – Estatísticas Tempo Resposta Inicial

Média	14,9 s	25%	10,0 s
Desvio Padrão	8,32 s	50%	15,2 s
Mínimo	1,71 s	75%	16,7 s
Máximo	45,6 s	Amostras	23

A avaliação é de que os tempos apresentados são consistentes com o requisito para tomada de decisão em tempo real. Nota-se, entretanto, a presença de um único valor atípico (*outlier*). Analisando a causa desse tempo de resposta tão longo, verificou-se uma limitação do algoritmo, que dentro de uma mesma janela de um minuto identificou uma manobra deliberada, realizada pelo operador e um desligamento automático. Se for considerado somente o desligamento automático o tempo de resposta inicial seria computado em apenas 9 segundos, próximo à média da distribuição, demonstrada na Tabela 4.

A evolução das perturbações e a atualização dos clusters apresentados ao operador também são relevantes à tomada de decisão e uma visão global. A Figura 13 mostra os tempos

até a última atualização dos agrupamentos de perturbações, tendo como referência o tempo de resposta inicial. Tem-se como satisfatória a estratégia de atualização adotada, uma vez que conseguiu registrar a evolução das perturbações, mesmo em cenários mais complexos como os apresentados no item 4.3.1. Nota-se que os tempos maiores do que 150 segundos são exatamente os referentes aos dois casos destacados nesse item.

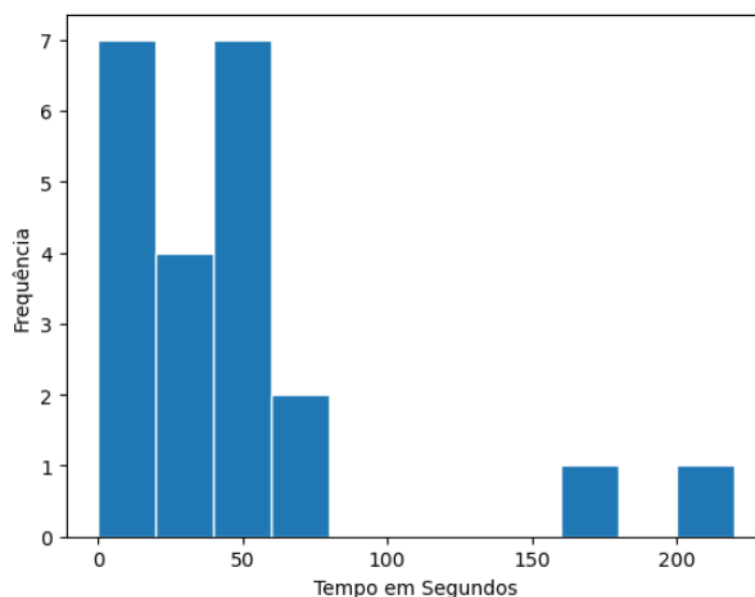


Figura 13 – Distribuição até a última atualização. Fontes: Própria (2023).

4.3.5. Análise Crítica e Limitações

De modo geral a ferramenta desenvolvida para o processamento automático dos alarmes, com a identificação e o agrupamento das perturbações é considerada satisfatória, tendo em vista as avaliações de resultados apresentadas. Destaca-se especialmente a efetividade do algoritmo da ferramenta na comparação com os dados do sistema de gestão da operação, em relação a redução do número de alarmes que são apresentados ao operador e ao tempo de resposta para tomada de decisão. Assim entende-se para os casos avaliados que a ferramenta é capaz de realizar o agrupamento dos fluxos contínuos de alarmes, diminuindo a carga de trabalho analítico dos operadores.

No entanto, destacou-se algumas limitações intrínsecas como a dependência de que a informação chegue ao centro de operação por meio do sistema SAGE EMS, não sendo tratadas falhas de comunicação ou perturbações causadas por sistemas externos, que não são de propriedade da Eletrobras Eletronorte. Ademais, destacou-se a ineficácia da ferramenta de

separar uma manobra espontânea e um desligamento automático que ocorreram em uma janela de tempo estreita.

Para superar as limitações do primeiro caso seria apropriado complementar as entradas de dados com outros sistemas além do SAGE EMS do centro de operação. Um exemplo seria uma integração com o sistema de oscilografias para complementar a identificação das perturbações, bem como sistemas de monitoramento dos ativos por câmeras e outros sensores inteligentes. Outro seria uma integração em paralelo com os registros (LOG) dos sistemas locais das instalações, de modo mitigar o número de perturbações não identificadas.

Com relação a separação de manobras e desligamentos automáticos em janelas de tempo estreitas, é possível implementar melhorias no próprio algoritmo de atualização dos clusters para eliminar do agrupamento final os desligamentos manuais.

Entende-se que algumas melhorias ainda poderiam ser propostas para acelerar ainda mais os tempos de processamento e a acurácia da ferramenta. Uma delas se refere a melhorias na padronização e classificação dos dados no SAGE EMS do centro de operação da Eletronorte. As lógicas para identificação das perturbações das instalações e dos equipamentos envolvidos poderia ser mais eficaz se a fonte estivesse mais bem padronizada.

Outra oportunidade nesse sentido seria rever a modelagem das dimensões distais, entre equipamentos e entre instalações. Acredita-se que a modelagem utilizando-se técnicas e bibliotecas do Python orientadas a grafos seria ainda mais apropriada dos que os modelos apresentados.

5. CONCLUSÃO

A operação teleassistida de subestações e usinas por centros de operação apresenta diversos desafios às empresas do Setor Elétrico. Apesar do potencial na redução de custos com a redução dos quadros de operadores, a eficiência operacional é também um objetivo central nesse cenário. Assim diminuir a sobrecarga cognitiva em cima das equipes de operadores é fundamental para desempenho da operação em tempo real.

Essa sobrecarga se torna bastante relevante em perturbações de grande porte, com o desligamento de diversos equipamentos em instalações espalhadas pelo sistema elétrico, abrangendo inclusive mais de uma área geo-elétrica.

A proposta desse trabalho foi apresentar uma ferramenta de avaliação rápida e assertiva da condição do sistema elétrico após perturbações, com base nos alarmes em tempo real apresentados pelo SAGE EMS. Como demonstrado, o CluAlarm é capaz de agrupar os fluxos contínuos de alarmes e diminuir/sumarizar os alarmes que são apresentados aos operadores de tempo real, mesmo em situações de perturbações de porte relevante.

A comparação com o sistema de gestão da operação da Eletrobras Eletronorte e a avaliação dos tempos de resposta da ferramenta mostraram a aplicabilidade da ferramenta bem como a compatibilidade com o requisito para tomada de decisão em tempo real, potencializando a eficiência do time de operação e diminuindo perdas financeiras.

Algumas limitações e oportunidades de melhoria foram apresentadas, como a possibilidade de se integrar a sistemas complementares ao SAGE EMS como os sistema de oscilografia e monitoramento por meio de câmeras e outros sensores inteligentes, bem como, melhorias na modelagem das dimensões distais com ferramentas mais avançadas para tratamento de grafos também podem ser exploradas em trabalhos futuros.

REFERÊNCIAS

- ABSOLAR, **Por que devemos apoiar o Projeto de Lei 5829/19**, Brasília: ABSOLAR, 2023. Disponível em: <https://www.absolar.org.br/noticia/por-que-devemos-apoiar-o-projeto-de-lei-5829-19/>. Acesso em: março de 2023.
- AGGARWAL, C.C. et al. **A Framework for Clustering Evolving Data Streams**. Proceedings of the 29th VLDB Conference, Berlin, Alemanha, 2003.
- AGGARWAL C. C. et al. **Data Streams: Models and Algorithms**. 2014. Springer ISBN: 978-1461497684
- BIFET, A.; et al. **Machine Learning for Data Streams: with Practical Examples in MOA**. MIT Press, 2018.
- BIFET, A.; et al. **MOA – Massive Online Analysis: A Framework for Stream Classification and Clustering**. Journal of Machine Learning Research, 2010, v.11, p.1601-1605. MIT Press, 2010.
- CAMARA DE COMERCIALIZAÇÃO DE ENERGIA ELÉTRICA – CCEE, **Governança**, Brasília: CCEE, 2023. Disponível em: <https://www.ccee.org.br/web/guest/governanca>. Acesso em: março de 2023.
- CAO, F.; ESTER, M; ZHOU, A. **Density-based clustering over an evolving data stream with noise**. In Proceedings of the 2006 SIAM international conference on data mining, p. 328-339. Society for Industrial and Applied Mathematics, 2006.
- CENTRO DE PESQUISAS DE ENERGIA ELÉTRICA – CEPEL, **Sistema Aberto de Gerenciamento de Energia – SAGE**. Disponível em: <https://sage.cepel.br/index.php/pt/>. Acesso em: março de 2023.
- DATAIKU, **Incremental KMeans**: Cardinal 0.0.9 Documentation Examples – Disponível em https://dataiku-research.github.io/cardinal/auto_examples/plot_incr_kmeans.html#sphx-glr-auto-examples-plot-incr-kmeans-py. Acesso em: março de 2023.
- DINIZ, F.F.M. **Análise do impacto regulatório quanto à aplicação da parcela variável no serviço público de transmissão de energia elétrica**. Dissertação (Mestrado) - Universidade Federal de Itajubá, Itajubá, 2021.

GAMA, J.; LOPES, L. M.; RODRIGUES, P.P **Clustering distributed sensor data streams using local processing and reduced communication**. 2011, Intelligent Data Analysis v. 15 n. 1, p. 3-28. IOS Press 2011.

LIN, J.; KEOGH, E.; LONARDI, S.; CHIU B. **A Symbolic Representation of Time Series, with Implications for Streaming Algorithms**. DMKD' 03, 2003, San Diego, Estados Unidos.

MACEDO, G. O. et al. **Um modelo de infraestrutura de microsserviços para algoritmos de agrupamento de fluxos contínuos de dados**. SODEBRAS journal, v.15, n. 178, p.48-52, Belo Horizonte, 2020.

MEMARI, I. **Python implementation of the data stream clustering algorithm "DenStream"**, 2020. Disponível em: <https://github.com/issamemari/DenStream>. Acesso em: março de 2023.

MONTIEL, J.; READ, J., BIFET, A.; ABDESSALEM, T. **Scikit-multiflow**: A multi-output streaming framework. The Journal of Machine Learning Research, v. 19, n. 72, p. 1–5. MIT Press, 2018.

OPERADOR NACIONAL DO SISTEMA ELÉTRICO – ONS, **O Sistema em números**, Brasília: ONS, 2023. Disponível em: <https://www.ons.org.br/paginas/sobre-o-sin/o-sistema-em-numeros>. Acesso em: março de 2023.

OPERADOR NACIONAL DO SISTEMA ELÉTRICO – ONS, **Procedimentos de Rede**, Brasília: ONS, 2023. Disponível em: <https://www.ons.org.br/paginas/sobre-o-ons/procedimentos-de-rede/o-que-sao>. Acesso em: março de 2023.

ORDONEZ, C. **Clustering binary data streams with K-means**. DMKD03: 8th ACM SIGMOD Workshop on Research Issues in Data Mining and Knowledge Discovery, p. 12–19, San Diego, Estados Unidos, 2003.

PATEL, P. **PyDenStream**: Implementation of the algorithm Density-Based Clustering over an Evolving Data Stream with Noise in Python, 2022. Disponível em: <https://github.com/MrParosk/pyDenStream>. Acesso em: março de 2023.

SEDGEWICK, R.; WAYNE, K. **Algorithms**, 2020, 4th Edition, p. 4.4 (Shortest Paths), disponível em <https://algs4.cs.princeton.edu/44sp/>. Acesso em março de 2023.

SIROIS C.B **MagikStack Asyncpg: A fast PostgreSQL Database Client Library for Python/asyncio**, 2022, Disponível em: <https://github.com/MagicStack/asyncpg>. Acesso em: junho de 2023.

VISWARUPAN, N. **CluStream — A Framework for Clustering Evolving Data Streams**, Medium, 2017, Disponível em: <https://niruhan.medium.com/clustream-a-framework-for-clustering-evolving-data-streams-b2f8b2d65ae>. Acesso em: março de 2023.

WAYLON, L. **DenStream clustering**, 2020, Disponível em: <https://github.com/waylongo/denstream>. Acesso em: março de 2023.

ZUBAROGLU A.; ATALAY V. Data Stream Clustering: A review. Ankara, Turquia, 2020.