

# Agrupamento Inteligente de Ocorrências para Priorização de Serviços em Sistemas de Distribuição

Saulo G. Rodrigues\* Rachel de O. C. dos Santos\* Bruno R. Sodré\*  
Carolina C. Milhomem\* Joe L. T. Morra\* Bruna M. Mulinari\*\*  
Vinicius T. Parede\*\* Clayton H. da Costa\*\* Flavio L. Grando\*\*  
Fabio A. Guerra\*\* Milton P. Ramos\*\*

\* *Equatorial Energia, Maceió, AL (e-mail: {saulo.rodrigues;  
rachel.santos; bruno.sodre; carolina.milhomem;  
joe.morra}@equatorialenergia.com.br).*

\*\* *Tech2Think, Curitiba, PR (e-mail: {bruna.mulinari; vinicius.parede;  
clayton.costa; flavio.grando; fabio.guerra;  
milton.ramos}@tech2think.com.br)*

---

**Abstract:** This research presents a machine learning-based application developed to improve the energy restoration process in distribution systems. Based on energy outage events generated by consumer complaints, the proposed method automatically analyzes data patterns to identify collective events, i.e., outages that affect several consumers simultaneously. In this way, operators can prioritize service in occurrences with a higher amount of affected customers and as a consequence, improve the quality of service. The method has been applied in an energy utility bringing practical experiences in the dispatch process of maintenance teams. The results contemplate improvements of up to 20% in the classification of occurrences, a reduction in the identification time of collective occurrences from a few hours to a few minutes and significant improvement in the reliability indicators.

**Resumo:** Esta pesquisa apresenta uma aplicação baseada em aprendizado de máquina desenvolvida para melhorar o processo de restabelecimento de energia em sistemas de distribuição. Com base em eventos de falta de energia gerados por reclamações de consumidores, o método proposto analisa automaticamente os padrões de dados para identificar eventos coletivos, ou seja, faltas de energia que afetam vários consumidores simultaneamente. Dessa forma, os operadores podem priorizar o atendimento em ocorrências com maior número de clientes afetados e, conseqüentemente, melhorar a qualidade do serviço. O método foi aplicado em uma concessionária de energia trazendo experiências práticas no processo de despacho das equipes de manutenção. Os resultados contemplam melhorias de até 20% na classificação de ocorrências, uma redução no tempo de identificação de ocorrências coletivas de algumas horas para alguns minutos e uma melhoria significativa nos indicadores de confiabilidade.

*Keywords:* machine learning; clustering; service prioritization; distribution systems.

*Palavras-chaves:* aprendizado de máquina; agrupamento; priorização de serviços; sistemas de distribuição.

---

## 1. INTRODUÇÃO

O setor de distribuição de energia elétrica é voltado para o fornecimento de equipamentos e serviços que maximizam a qualidade da energia fornecida ao consumidor. Entretanto, há fatores que não podem ser controlados, como condições climáticas adversas que levam a falhas na rede de distribuição e interrupções no fornecimento de energia. Em situações como essa, a maior preocupação é restaurar o fornecimento de energia o mais rápido possível.

Para melhorar a qualidade do serviço, os órgãos reguladores do setor elétrico em todo o mundo utilizam métricas

de confiabilidade, como o Índice de Duração Média de Interrupção do Sistema (SAIDI), que indica a duração das interrupções no fornecimento de energia (IEEE, 2012). No cenário brasileiro, a ANEEL (Agência Nacional de Energia Elétrica) apresenta a resolução 956/2021 denominada Procedimentos de Distribuição (PRODIST) onde estabelece alguns indicadores de qualidade de serviços para as concessionárias. Dentre eles o CHI (Consumidor Hora Interrompido), que indica o tempo (em horas) que um conjunto de consumidores ficou sem fornecimento de energia elétrica (ANEEL - Agência Nacional de Energia Elétrica, 2021). Esse indicador também está relacionado a outros indicadores de confiabilidade, de modo que um CHI elevado implica penalidades e multas para a empresa de distribuição de energia.

---

\* Projeto financiado pelo Programa P&D da ANEEL e pela Equatorial Energia (PD-00037-0044/2022).

Diferentes métricas e métodos de cálculo são adotados em todo o mundo, o que motiva diferentes estratégias de gerenciamento. Alguns autores propõem a integração das tarefas de gerenciamento de ativos e interrupções, com o objetivo de melhorar a confiabilidade e o retorno sobre o investimento (Dong et al., 2009). Outros sugerem novas métricas ou formas de cálculo com o objetivo de melhorar a regulamentação do desempenho das empresas (Jooshaki et al., 2018). Independentemente das diferenças, o consenso é buscar a minimização das interrupções de energia. Com uma grande quantidade de informações disponíveis nas empresas de energia, as abordagens orientadas por dados trazem novas oportunidades.

### 1.1 Trabalhos Relacionados

Em Chen et al. (2014), discute-se o uso de big data para o gerenciamento de interrupções, apontando várias fontes de dados úteis para essa finalidade. Em Dunn et al. (2019), a análise de dados revela heterogeneidade espaço-temporal, sugerindo que as estratégias adotadas devem ser adaptáveis ao longo do tempo e abordar as particularidades regionais.

Com relação aos métodos, uma comparação das estruturas de interrupção de energia é apresentada por Bahmanyar et al. (2017). Os métodos avançados têm altos custos para implantar e operar a infraestrutura de medição. O método tradicional é baseado em chamadas de consumidores, cujas principais limitações são o consumo de tempo e um número suficiente de chamadas para identificar interrupções de energia. Esse é exatamente o cenário vivenciado na presente pesquisa.

A literatura carece de trabalhos que envolvam métodos baseados em chamadas dos consumidores. No entanto, trabalhos recentes propõem sistemas de regras para identificar e priorizar a manutenção, conforme apresentado por Guang (2019). Porém, esse modelo baseia-se apenas na percepção do consumidor para evitar mais reclamações. Outros trabalhos sobre o processamento de reclamações podem ser encontrados em Uzunoğlu (2017) e Al-Kanj et al. (2017). Esses trabalhos se baseiam na probabilidade Bayesiana para estimar o local das interrupções, mas não há estratégias de priorização. Por outro lado, o estudo apresentado por Rosales-Asensio et al. (2022) propõe o uso de um sistema de ponderação para priorizar consumidores críticos. Entretanto, o método baseia-se em uma pré-classificação de pesos para consumidores especiais e não no maior número de consumidores afetados.

A pesquisa proposta aqui se baseia no processo de chamadas por falta de energia e tem como objetivo agrupar reclamações referentes ao mesmo problema, identificar o maior número de consumidores afetados e usar isso como uma estratégia de priorização para o atendimento. Esse problema específico não foi encontrado na literatura, cujos detalhes são apresentados na próxima seção.

### 1.2 Descrição do Problema

Alguns alimentadores de distribuição são caracterizados por seu longo comprimento, atendimento ao cliente disperso em locais remotos e falta de monitoramento em tempo real. Isso significa que, para identificar e localizar

uma queda de energia, a empresa depende do contato com o cliente. As empresas têm vários canais de comunicação, que coletam reclamações de falta de energia e abastecem o centro de operações da empresa de distribuição. A tarefa dos operadores é analisar as reclamações e enviar equipes de manutenção para restaurar a energia para o maior número possível de clientes no menor tempo possível. Esse processo é ilustrado na Figura 1.

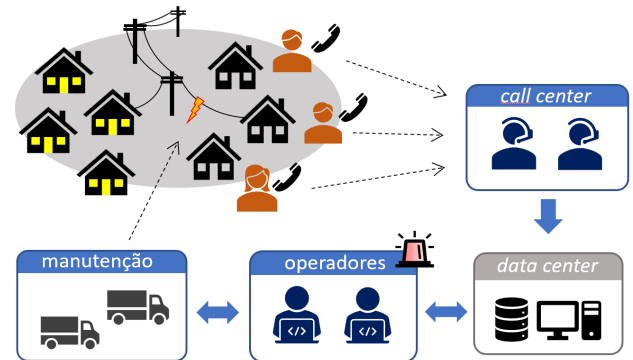


Figura 1. Consumidores manifestam falta de energia e suas reclamações entram na fila de análise. Os operadores do sistema analisam as solicitações e tentam identificar o problema, passando a tarefa de restabelecimento para a equipe de manutenção.

A prioridade no atendimento é para a maior quantidade de consumidores afetados pelo mesmo problema elétrico, ou seja, uma *ocorrência coletiva*. Entretanto, inicialmente, todas as reclamações são individuais, cabendo aos operadores do sistema identificar quando uma ou mais reclamações se referem a uma falha coletiva. Esse processo de agrupamento de reclamações é chamado de *formação de ocorrência coletiva*.

O volume de informações que os operadores devem processar em tempo real pode aumentar exponencialmente com o aumento do número de reclamações durante períodos críticos. Isso sobrecarrega os setores da empresa, faz com que os consumidores passem mais tempo sem energia, a empresa paga multas por interrupções no fornecimento e aumenta o custo de manutenção com erros de prioridade.

Portanto, trata-se de uma tarefa que depende muito da heurística dos operadores para identificar e diferenciar as ocorrências individuais das coletivas. Para aprimorar esse processo, este documento apresenta um projeto de pesquisa e desenvolvimento focado no aprimoramento do processo de restauração de energia em uma distribuidora usando técnicas baseadas em aprendizado de máquina. A proposta está sendo testada e tem apresentado resultados promissores em 4 concessionárias de energia elétrica de diferentes regiões brasileiras: Alagoas (AL), Pará (PA), Maranhão (MA) e Piauí (PI).

### 1.3 Resumo das Contribuições

Com base nas considerações anteriores, as principais contribuições deste trabalho podem ser resumidas da seguinte forma:

- Taxas de acerto: melhoria significativa na identificação de ocorrências coletivas em comparação com a metodologia atualmente utilizada na empresa;

- Tempo de resposta: redução acentuada do tempo de formação de ocorrências coletivas;
- Confiabilidade: melhoria nos indicadores de confiabilidade proporcionada pela maior agilidade e precisão do processo;
- Experiências práticas: implantação na distribuidora, com análise de desempenho e considerações práticas das equipes de operação;

Desta forma, a seção seguinte 2 descreve o projeto de uma solução para classificação de ocorrências na rede de distribuição. A seção 3 apresenta alguns resultados obtidos, discussão e avaliação do impacto desses resultados para a empresa. Finalmente, a seção 4 traz conclusões e desenvolvimentos futuros.

## 2. SOLUÇÃO PROPOSTA

Um diagrama resumido da solução proposta é apresentado na Figura 2. A proposta usa a plataforma Python para o estágio de *back-end* e o PowerBI para o estágio de *front-end*. A partir do banco de dados, o processo tem dois fluxos de dados distintos. O primeiro deles é destinado ao treinamento do modelo no estágio de aprendizado supervisionado, portanto, é *off-line* e executado apenas uma vez. O segundo e principal fluxo de informações é *on-line*, no qual há solicitações contínuas no banco de dados que permitem a aplicação do modelo criado anteriormente no estágio de aprendizado.

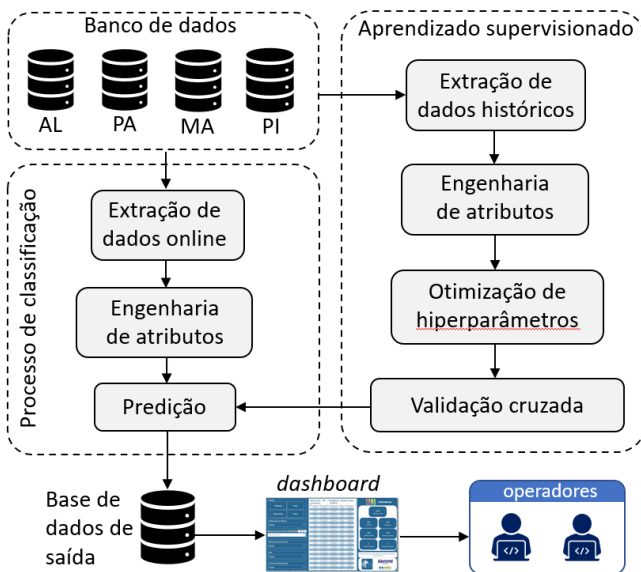


Figura 2. Visão geral da solução proposta. Um modelo de aprendizado de máquina é criado para operar *on-line* conectado ao banco de dados da empresa. O resultado da previsão é enviado a um painel de visualização para dar suporte à equipe de operação.

Portanto, o fluxo principal consiste basicamente em um estágio de extração de dados, passando por um estágio de processamento representado pelo bloco de engenharia de recursos e terminando com a aplicação do modelo de aprendizado de máquina, em que ocorre a classificação das ocorrências abertas no sistema. Depois que as ocorrências são classificadas e ordenadas, o resultado é enviado para um banco de dados de saída. Esse banco de dados se

comunica com o painel de controle para exibir o resultado aos operadores do sistema. As seções a seguir descrevem os detalhes adotados em cada etapa do processo.

### 2.1 Bases de Dados

Os conjuntos de dados são representados pelas abreviações AL, PA, MA e PI na Figura 2. Essas abreviações representam quatro regiões distintas no Brasil, a saber, Pará (PA), Alagoas (AL), Maranhão (MA) e Piauí (PI). Essas regiões têm características geográficas diferentes, como o tamanho do estado, o tamanho das áreas rurais em relação às áreas urbanas, o número de consumidores e o comportamento do consumidor. Essas diferenças se refletem na distribuição de dados, nos procedimentos e regras internas relativas à falta de energia e também nos indicadores de confiabilidade. Devido a essas diferenças, cada região foi analisada separadamente, com um certo nível de personalização da solução proposta para cada uma delas.

Para determinar se um evento de falta de energia está afetando um único consumidor ou vários, os operadores de rede precisam contar com as informações de, principalmente, três fontes de dados. A primeira fonte contém os dados de registro do consumidor, como sua localização e o equipamento elétrico mais próximo (como o transformador) ao qual ele está conectado. Da segunda fonte, o operador pode extrair informações sobre os próprios equipamentos elétricos. A terceira fonte de dados é composta por dados históricos e atuais, de notas de reclamação e eventos. Isso inclui, entre outras coisas, a identificação do reclamante, o canal que ele usou para abrir a nota de reclamação (telefone, Internet, aplicativo móvel), bem como qualquer informação adicional que ele possa ter fornecido ao abrir a nota de reclamação. Todas essas fontes de dados estão concentradas em um banco de dados completo para cada uma das quatro regiões analisadas.

### 2.2 Engenharia de Atributos

O banco de dados disponível tem um grande volume de informações que dificulta as implementações práticas e torna o estudo baseado em aprendizado de máquina muito lento. Nesse processo, o tempo é crítico, pois os consumidores ficam sem energia enquanto as regras atuais precisam acumular reclamações para identificar uma ocorrência coletiva. Diante disso, uma premissa importante foi elencada no desenvolvimento da solução: *acelerar* o processo de identificação de ocorrências.

Para dar prioridade à resposta rápida, buscou-se *minimizar* o conjunto de dados usados no aprendizado de máquina. Essa abordagem permite simplificar o modelo desenvolvido e obter tempos de resposta curtos, dada uma série de acessos aos bancos de dados em tempo real.

Assim, a partir dos bancos de dados citados na seção anterior, é aplicado um processo de análise e filtragem dividido em duas etapas:

- 1º estágio: pesquisa e redução do volume de informações a um conjunto potencialmente relevante, com base no conhecimento e experiência dos operadores;
- 2º estágio: a partir do conjunto obtido no estágio anterior, aplicou-se um estudo orientado por dados

para identificar os melhores atributos para classificar as ocorrências;

Como resultado disso, obteve-se um conjunto de atributos categóricos e numéricos que estão listados e detalhados nos três grupos abaixo.

O primeiro grupo de atributos está relacionado à localização da falha. Os locais têm um perfil probabilístico associado à recorrência de falta de energia. Os seguintes recursos fazem parte desse grupo:

- *Localizações geográficas*: tanto o bairro/distrito quanto a cidade onde o reclamante reside
- *Locais elétricos*: tanto o alimentador elétrico quanto a subestação.
- *ID do elemento de rede*: número de identificação do elemento de rede georreferenciado, como postes ou equipamentos.

O segundo conjunto está associado aos possíveis consumidores afetados pela interrupção, caso ela seja confirmada. Os atributos são:

- *Equipamento*: Tipo do equipamento elétrico a montante mais próximo do ponto de defeito. Normalmente, um transformador ou uma chave.
- *Consumidores conectados ao equipamento*: Quantidade total de consumidores nesse equipamento a montante. Esse recurso representa o potencial de consumidores afetados em caso de acionamento do equipamento de proteção.
- *Consumidores potencialmente afetados*: Quantidade total de consumidores conectados no ponto de defeito. Esse atributo representa os possíveis consumidores afetados pela interrupção. Um grande número de consumidores tende a gerar novas reclamações de interrupção e aumentar a probabilidade de agrupamento.

O terceiro e último grupo consiste em informações associadas aos perfis de reclamações dos consumidores, como segue:

- *Identificação do consumidor*: Recurso útil no sentido de trazer as informações de chamadas anteriores feitas pelo consumidor, bem como uma taxa específica de quão precisas foram suas reclamações (em termos de uma falta de energia coletiva ou individual ou até mesmo a extensão da referida falta).
- *Informações adicionais*: Espaço para adicionar informações em formato de texto que podem ajudar no processo de tomada de decisão dos operadores. Às vezes, esse campo é usado para informar situações como “rua inteira sem energia”. Para o atributo, um valor numérico é obtido por um algoritmo de correspondência para palavras (ou frases) mais comuns relacionadas a uma falha individual ou coletiva.

### 2.3 Aprendizado Supervisionado

Neste trabalho, uma técnica de aprendizado de máquina é usada para classificar os casos derivados de cada nota de reclamação em individuais ou coletivas. Devido aos vastos dados históricos disponíveis com uma grande quantidade de exemplos rotulados, optou-se por uma abordagem de aprendizado supervisionado com um classificador *Random Forest* (Breiman, 2001).

O classificador *Random Forest* (RF) é um modelo de conjunto, ou seja, um modelo derivado de uma combinação de vários outros modelos de aprendizado de máquina. No caso específico do RF, os modelos combinados são os classificadores de árvore de decisão. Em outras palavras, para criar um modelo de RF, é preciso criar e treinar várias árvores de decisão e passar os dados por cada uma delas. A classe que será escolhida é aquela votada pela maioria dessas Árvores de Decisão.

Para escolher o conjunto de hiperparâmetros do modelo, foi utilizada a técnica de *Random Search* (Bergstra and Bengio, 2012). Esse método consiste em treinar vários modelos com hiperparâmetros escolhidos aleatoriamente (dentro de um espaço de busca previamente determinado) e selecionar o melhor (em termos de uma métrica selecionada).

Para o treinamento de cada modelo, em termos de hiperparâmetros, foi imposto um limite de 40 profundidades. A quantidade de árvores escolhidas após o *Random Search* foi de 43, 71, 59 e 73 para os modelos AL, MA, PI e PA, respectivamente. A quantidade mínima de amostras por folha foi de quatro para os modelos AL e PA e de duas para os demais modelos. Da mesma forma, a quantidade de amostras necessárias para uma divisão em cada árvore de decisão foi 10 para cada modelo, exceto PA, que foi 5.

Depois que os hiperparâmetros foram definidos, foi aplicada uma validação cruzada  $k$ -fold (Cherkassky and Muller, 2014). A validação cruzada  $k$ -fold é uma técnica que divide ainda mais o conjunto de dados de treinamento em subconjuntos  $k$ , cada um com uma parte de treinamento e uma parte de teste. Neste trabalho, selecionamos  $k = 5$ , o que significa que 20% do conjunto é alocado para teste e os 80% restantes para treinamento. O processo é repetido  $k = 5$  vezes, selecionando aleatoriamente uma porcentagem de padrões por classe, sem repetir os padrões de teste em cada subconjunto.

## 3. RESULTADOS E DISCUSSÃO

O método proposto foi implementado na distribuidora de energia e permaneceu em operação ininterrupta por pelo menos 4 meses até o encerramento deste estudo. Durante esse período, foram obtidos resultados relevantes associados à taxa de acerto, ao tempo de resposta, ao impacto nos indicadores e às experiências práticas com a equipe de operação. Esses resultados são apresentados nas seções a seguir.

### 3.1 Taxas de Acerto

A Figura 3 apresenta a matriz de confusão para as quatro regiões analisadas. Antes do uso do método proposto, a exatidão da classificação era de 59%, 72%, 56% e 52% para os estados AL, MA, PI e PA, respectivamente. O método proposto alcançou 72%, 71%, 74% e 70% para essas regiões. Portanto, houve uma melhoria significativa em três casos e resultado equivalente para MA. Isso pode ser explicado pelo fato de que a exatidão para essa região já era maior devido a algumas regras específicas usadas pela equipe.

|     |          | predição |      | Acerto geral: |
|-----|----------|----------|------|---------------|
|     |          | Ind.     | Col. |               |
| AL: | objetivo | Ind.     | 31%  | 16%           |
|     | Col.     | 13%      | 41%  | 72%           |

|     |          | predição |      | Acerto geral: |
|-----|----------|----------|------|---------------|
|     |          | Ind.     | Col. |               |
| MA: | objetivo | Ind.     | 21%  | 17%           |
|     | Col.     | 11%      | 50%  | 71%           |

|     |          | predição |      | Acerto geral: |
|-----|----------|----------|------|---------------|
|     |          | Ind.     | Col. |               |
| PA: | objetivo | Ind.     | 37%  | 15%           |
|     | Col.     | 15%      | 34%  | 70%           |

|     |          | predição |      | Acerto geral: |
|-----|----------|----------|------|---------------|
|     |          | Ind.     | Col. |               |
| PI: | objetivo | Ind.     | 19%  | 17%           |
|     | Col.     | 9%       | 55%  | 74%           |

Figura 3. Matriz de confusão para os modelos treinados.

É importante ressaltar que, na nova metodologia proposta, cada ocorrência é analisada individualmente com base em seu conjunto de características. Em contraste, no método tradicional adotado pela companhia de energia, várias ocorrências são acumuladas para obter a classificação manual. Essa abordagem maximiza a precisão da classificação manual sob a penalidade de levar mais tempo para classificar, um tópico discutido na próxima seção.

### 3.2 Tempo de Resposta

O método precedente para essa tarefa de classificação foi bastante manual, o que faz com que, em alguns casos, possa levar horas para ser feita, especialmente em períodos de crise, como os causados por chuvas fortes, tempestades e fenômenos semelhantes.

Além disso, é necessário mais reclamações associadas a uma mesma ocorrência para se concluir que se trata de uma ocorrência coletiva. Todos esses fatores implicam tempos de espera de mais de 150 minutos em alguns casos, conforme exemplos na Tabela 1. Em contraste com isso, a proposta apresenta tempos de resposta de alguns segundos a menos de 5 minutos.

Tabela 1. Comparação de tempo de resposta.

| Caso | Número de clientes | Tempo de predição (min.) |          | Tempo poupado |
|------|--------------------|--------------------------|----------|---------------|
|      |                    | original                 | proposta |               |
| #1   | 20                 | 106                      | 0.6      | 99.4%         |
| #2   | 40                 | 159                      | 0.6      | 99.6%         |
| #3   | 378                | 27                       | 4.8      | 82.2%         |
| #4   | 63                 | 144                      | 0.6      | 99.6%         |
| #5   | 215                | 83                       | 2.4      | 97.1%         |
| #6   | 217                | 39                       | 2.4      | 93.8%         |

O tempo de resposta do método proposto está associado principalmente ao acesso ao banco de dados. Com um número maior de consumidores potencialmente afetados por uma falha, mais dados são solicitados do banco de dados. Por outro lado, grandes volumes de consumidores facilitam a classificação manual, pois mais reclamações são recebidas pelo sistema. Mesmo assim, a nova metodologia fornece classificações 80-99% mais rápidas do que a abordagem tradicional.

### 3.3 Impacto no Indicador de Confiabilidade

Apesar da rápida classificação das ocorrências, o processo de restabelecimento da energia é mais demorado, pois depende do trabalho das equipes de manutenção em campo. Por isso, avaliamos o impacto da metodologia proposta no indicador de tempo de interrupção do consumidor, denominado CHI (Consumidor Hora Interrompido). O resultado comparativo é apresentado na Tabela 2.

Tabela 2. Comparação do indicador CHI.

| Caso | número de clientes | CHI      |          | CHI poupado |
|------|--------------------|----------|----------|-------------|
|      |                    | original | proposta |             |
| #1   | 20                 | 37.40    | 2.00     | 94.7%       |
| #2   | 40                 | 377.60   | 271.61   | 28.1%       |
| #3   | 378                | 1924.49  | 1754.47  | 8.8%        |
| #4   | 63                 | 348.39   | 197.20   | 43.4%       |
| #5   | 215                | 1324.40  | 1025.59  | 22.6%       |
| #6   | 217                | 809.41   | 668.40   | 17.4%       |

O CHI é obtido pelo número de horas de interrupção no fornecimento de energia multiplicado pelo número de consumidores. Portanto, quanto mais consumidores forem afetados por um evento, maior será o CHI.

Na metodologia apresentada, o tempo economizado na identificação de coletivas não depende de muitos consumidores e com isso o impacto relativo no CHI é menor, como caso 3 na Tabela 2. Por outro lado, poucos consumidores são mais difíceis de detectar manualmente e o impacto relativo da proposta é maior, como no caso 1.

### 3.4 Experiências Práticas

A solução desenvolvida é inserida no meio de um processo multitarefa e compartilhada com os operadores, levantando questões sobre a aplicabilidade da solução. Portanto, esta seção tem como objetivo descrever as experiências obtidas durante o período de 4 meses de testes no setor de operações da empresa.

Para monitorar o sistema de classificação e fornecer os resultados aos operadores da rede elétrica, foi desenvolvida a interface mostrada na Figura 4. Como uma fase de testes, os operadores precisam visualizar os resultados nessa interface para apoiar a tomada de decisões.

Apesar dos resultados promissores apresentados nas seções anteriores, a equipe de operação relata picos de trabalho devido à avaliação de resultados que são entregues muito mais rapidamente em comparação com a metodologia antiga. Na metodologia anterior, havia um acúmulo de ocorrências para a formação do coletivo, portanto, o trabalho de manutenção era distribuído ao longo das horas. Com o sistema automatizado, o trabalho do operador foi reduzido significativamente, mas a demanda de trabalho de manutenção ficou concentrada em alguns períodos.

Além disso, o restabelecimento do fornecimento de energia ainda depende do desempenho das equipes de manutenção. Portanto, apesar da resposta mais rápida na identificação de ocorrências coletivas, a redução dos indicadores pode esbarrar nas limitações das equipes de campo.

| Ocorrência | Mês | Abrangência | Data                | Consumidor | Subestação | Alimentador | Previsão   | Cientes |
|------------|-----|-------------|---------------------|------------|------------|-------------|------------|---------|
| 16603      | 3   | CR          | 17/03/2023 15:10:09 | 14.466.252 | RLU        | Y1          | INDIVIDUAL | 109     |
| 13969      | 3   | CR          | 15/03/2023 19:10:11 | 3.863.221  | TBM        | Y3          | INDIVIDUAL | 106     |
| 16413      | 3   | CR          | 17/03/2023 12:28:16 | 8.266.158  | ARD        | M5          | COLETIVA   | 106     |
| 16154      | 3   | CR          | 17/03/2023 09:08:01 | 9.726.314  | CPJ        | Y3          | INDIVIDUAL | 103     |
| 16379      | 3   | CR          | 17/03/2023 11:23:15 | 10.407.278 | MCI        | Y2          | COLETIVA   | 103     |
| 16614      | 3   | CR          | 17/03/2023 15:50:45 | 6.962.084  | PCA        | Y5          | COLETIVA   | 102     |
| 16705      | 3   | CR          | 17/03/2023 17:14:26 | 16.035.380 | PMB        | Y4          | COLETIVA   | 102     |
| 16790      | 3   | CR          | 17/03/2023 18:38:13 | 15.930.955 | PMB        | Y4          | INDIVIDUAL | 102     |
| 15219      | 3   | CR          | 16/03/2023 11:00:56 | 13.003.356 | CZA        | Y4          | INDIVIDUAL | 100     |
| 16408      | 3   | CR          | 17/03/2023 12:29:44 | 1.575.538  | RLU        | Y5          | INDIVIDUAL | 99      |
| 15754      | 3   | CR          | 16/03/2023 18:34:53 | 8.227.438  | MCI        | Y2          | INDIVIDUAL | 98      |
| 15943      | 3   | CR          | 16/03/2023 22:53:10 | 17.316.499 | TBM        | W1          | INDIVIDUAL | 97      |
| 16516      | 3   | CR          | 17/03/2023 06:34:19 | 9.138.927  | MTB        | Y2          | COLETIVA   | 97      |
| 16544      | 3   | CR          | 17/03/2023 14:17:46 | 176.141    | PJA        | Y4          | INDIVIDUAL | 97      |
| 16485      | 3   | CR          | 17/03/2023 17:03:52 | 15.248.607 | ARD        | M6          | INDIVIDUAL | 97      |
| 16360      | 3   | CR          | 17/03/2023 11:38:29 | 9.488.812  | PMB        | Y7          | INDIVIDUAL | 96      |
| 13910      | 3   | CR          | 15/03/2023 12:34:05 | 6.814.053  | CRD        | M5          | COLETIVA   | 95      |
| 16604      | 3   | CR          | 17/03/2023 15:27:05 | 12.262.935 | TBM        | Y9          | INDIVIDUAL | 95      |
| 15819      | 3   | CR          | 16/03/2023 19:23:08 | 15.145.611 | UPM        | Y2          | INDIVIDUAL | 91      |

Figura 4. Dashboard desenvolvido para monitoramento do sistema de classificação e fornecimento de resultados à equipe de operação.

Com base nessa experiência e nos resultados apresentados, no futuro, avaliam-se ajustes para aumentar a precisão da classificação, mesmo que isso signifique sacrificar o tempo. Com isso, o resultado obtido deve reduzir o trabalho do operador e otimizar a utilização das equipes.

#### 4. CONCLUSÕES

Neste trabalho, é proposto um novo método para classificação de casos de interrupções elétricas individuais e coletivas. Esse método baseado em aprendizado de máquina mostra uma melhoria significativa em termos de precisão para três das quatro regiões brasileiras abordadas neste documento, com um aumento de 13%, 18% e 20% nas precisões de AL, PI e PA, bem como resultados semelhantes para o estado do MA. Além da precisão, o método proposto é significativamente mais rápido, levando um minuto para a classificação, ao contrário do método original, que pode levar até horas para ser executado (com precisão semelhante ou até menor). Além disso, o método é totalmente automático e não requer intervenção humana, o que faz com que haja uma redução significativa nas atividades de mão de obra para os operadores de rede, que agora podem se concentrar no gerenciamento das equipes de manutenção. De modo geral, essas vantagens fazem com que o consumidor tenha um serviço melhor, o operador da rede tenha um trabalho menos estressante e mais eficiente e a empresa de eletricidade tenha uma redução nas multas por tempo de interrupção elétrica devido a uma classificação mais rápida e precisa e, portanto, uma ação efetiva sobre casos coletivos.

#### AGRADECIMENTOS

Os autores agradecem o apoio financeiro da Equatorial Energia e da ANEEL (Agência Nacional de Energia Elétrica) – Programa de P&D, projeto PD-00037-0044/2022.

#### REFERÊNCIAS

Al-Kanj, L., Bouzaiene-Ayari, B., and Powell, W.B. (2017). A probability model for grid faults using incomplete information. *IEEE Transactions on Smart Grid*, 8(2), 956–968.

ANEEL - Agência Nacional de Energia Elétrica (2021). Módulo 8 – qualidade da energia elétrica. *Procedimentos*

de Distribuição de Energia Elétrica no Sistema Elétrico Nacional – PRODIST, 8, 1–90.

Bahmanyar, A., Jamali, S., Estebarsari, A., and Bompard, E. (2017). A comparison framework for distribution system outage and fault location methods. *Electric Power Systems Research*, 145, 19–34.

Bergstra, J. and Bengio, Y. (2012). Random search for hyper-parameter optimization. *Journal of machine learning research*, 13(2).

Breiman, L. (2001). Random forests. *Machine Learning*, 45, 5–32.

Chen, P.C., Dokic, T., and Kezunovic, M. (2014). The use of big data for outage management in distribution systems. In *International Conference on Electricity Distribution (CIRED) Workshop*.

Cherkassky, V. and Mulier, F.M. (2014). *Classical mechanics: the theoretical minimum*. Penguin Random House, New York, NY.

Dong, Y., Aravinthan, V., Kezunovic, M., and Jewell, W. (2009). Integration of asset and outage management tasks for distribution systems. In *2009 IEEE Power & Energy Society General Meeting*, 1–8.

Dunn, L.N., Sohn, M.D., LaCommare, K.H., and Eto, J.H. (2019). Exploratory analysis of high-resolution power interruption data reveals spatial and temporal heterogeneity in electric grid reliability. *Energy Policy*, 129, 206–214.

Guang, O. (2019). Research and analysis of power outage dispatch control system based on customer zero-blackout awareness. *IOP Conference Series: Materials Science and Engineering*, 677(4), 042065.

IEEE, S.A. (2012). IEEE guide for electric power distribution reliability indices. *IEEE Std 1366-2012 (Revision of IEEE Std 1366-2003)*, 1–43.

Jooshaki, M., Abbaspour, A., Fotuhi-Firuzabad, M., Moeini-Aghaie, M., and Lehtonen, M. (2018). Designing a new procedure for reward and penalty scheme in performance-based regulation of electricity distribution companies. *International Transactions on Electrical Energy Systems*, 28(11), e2628.

Rosales-Asensio, E., Elejalde, J.L., Pulido-Alonso, A., and Colmenar-Santos, A. (2022). Resilience framework, methods, and metrics for the prioritization of critical electrical grid customers. *Electronics*, 11(14).

Uzunoglu, B. (2017). Locating distribution power system fault employing bayes theorem with subjective logic. In *2017 2nd International Conference on System Reliability and Safety (ICSRS)*, 130–134.