



Inteligência Artificial para otimizar o tratamento de incidentes na distribuição de eletricidade

Tema: Qualidade de Energia

Autores: Martín Ariel Hunziker, Laura Alonso Alemany, Carla Rosetti, Felipe Arruda, Ronaldo Ribeiro de Almeida Junior; Paulo De Moraes Oliveira Filho; Antonio Manoel Matta Dos Santos Lameirao E Roberto Farias De Toledo

Co-Autores: -

Empresa: Calden

Resumo

Uma parte muito importante da qualidade do serviço é a resolução rápida de incidentes. As equipes técnicas designadas para lidar com incidentes têm uma alta porcentagem de deslocamentos improcedentes, ou seja, casos em que, quando a equipe chega ao local, o incidente já foi resolvido ou a equipe não tinha a capacidade de resolvê-lo. Portanto, a disponibilidade das equipes pode ser melhorada com a detecção prévia do atendimento insatisfatório, que permite evitar o deslocamento da equipe. A partir do banco de dados histórico de incidentes da Light, obtivemos um modelo preditivo que identifica corretamente até 90% dos incidentes que não deveriam ser tratados por uma equipe técnica, com apenas 6,7% de erros em que são despriorizados os casos que realmente necessitam de uma equipe técnica.

1. Introdução

Uma parte muito importante da qualidade do serviço é a resolução rápida de incidentes. Para isso, é essencial otimizar o uso das equipes técnicas, e um dos principais obstáculos ao uso eficiente das equipes são os deslocamentos improcedentes: aqueles casos em que uma equipe foi designada para atender a um incidente, mas o incidente já havia sido resolvido, não era de responsabilidade da distribuidora, poderia ter sido resolvido sem a necessidade de uma equipe, ou era necessário atender ao incidente com outro tipo de equipe ou de forma não técnica (administrativa, comercial etc.). Ferreira et al. (2018) abordam esse problema com resultados promissores.

Na Light, os incidentes considerados improcedentes representam cerca de 80% do atendimento a incidentes, o que significa que grande parte das horas da equipe técnica é ocupada por atendimentos desnecessários. Portanto, evitar pelo menos parte desses deslocamentos pode resultar em uma melhoria significativa no tempo de resolução de incidentes, contribuindo para uma melhoria geral na qualidade do serviço.

Neste trabalho, aplicamos técnicas de Inteligência Artificial para identificar possíveis deslocamentos improcedentes, alcançando uma precisão entre 88% e 90%, que aumenta à medida que o tempo passa desde o início do incidente.

As soluções orientadas por dados geralmente têm uma margem de erro, mas, nesse caso, os erros mais graves constituem 6,7% dos casos. Esses erros representam os casos em que um incidente que requer uma equipe técnica é identificado como improcedente e, portanto, não é priorizado para atenção por uma equipe técnica. No entanto, mesmo esses tipos de erros não teriam um impacto muito negativo sobre o processo de produção do sistema, pois seriam identificados manualmente como casos genuínos no processo de chamada e teriam sua prioridade alterada. Além disso, o tempo médio de atenção diminuiria, de modo que o tempo de atenção para esses erros também diminuiria, mesmo levando em conta o tempo em que eles são despriorizados, aguardando uma chamada de verificação.

2. Desenvolvimento

Conjunto de dados

Aplicamos diferentes abordagens de machine learning em um conjunto de dados da operação da Light, composto por 201.362 atendimentos de incidentes, registrados entre 1º de junho e 30 de outubro de 2024. Desses casos, 164.064 (81%) são considerados improcedentes, pois não exigiram a atenção de uma equipe técnica. Os seguintes tipos de casos são considerados improcedentes:

- Nenhum endereço encontrado devido a erro no cadastro ou ninguém no local para recebê-los (24%).
- O caso foi tratado por ligações telefônicas (23%),
- Quando uma equipe técnica chegou para atender ao incidente, o incidente já havia sido resolvido (22%),
- O caso foi um defeito elétrico dentro da residência (7%)
- Não se tratava de um problema técnico (3%).

A partir desse conjunto de dados, foram obtidas três partições: 20% dos casos foram reservados como partição de teste, para avaliar o desempenho dos diferentes modelos. Dos 80% restantes, 80% foram usados para treinamento e 20% para validação, a fim de encontrar os melhores valores para os hiperparâmetros das abordagens de aprendizado de máquina.

Características

Cada registro de atendimento é descrito pelas seguintes características dos bancos de dados da própria Light:

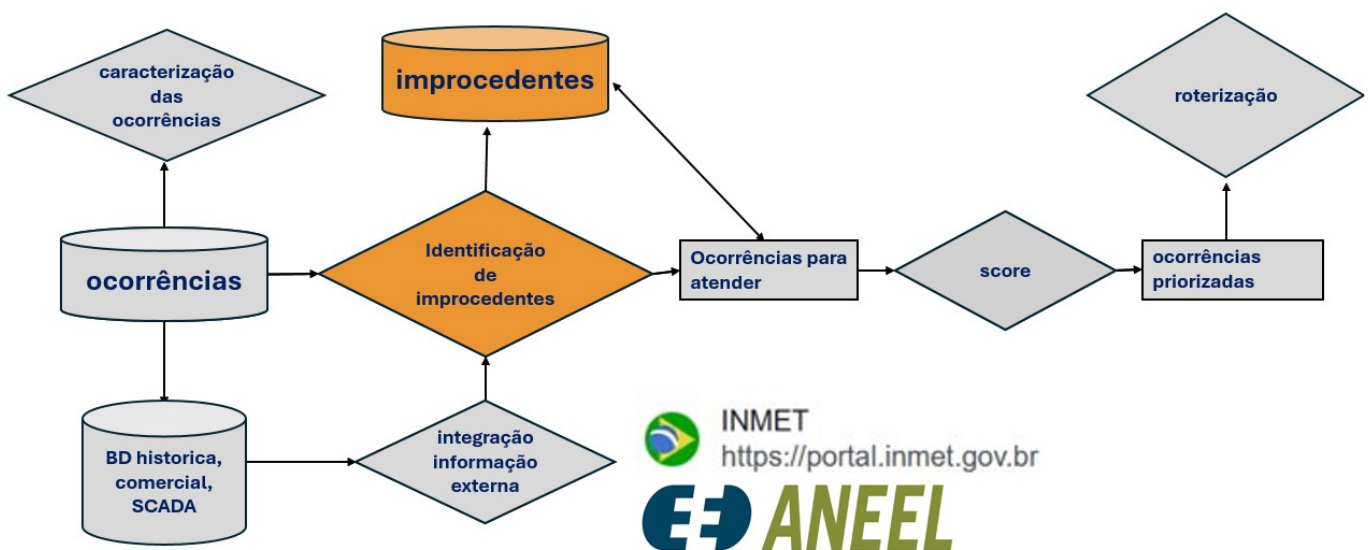
- classificação como área de risco
- tipo de área (urbana/rural),
- tipo de serviço com o qual o incidente é caracterizado quando aberto (relatado pela central de atendimento),
- região, bairro,
- indicação se há informações complementares para localizar o endereço,

- indicação se o incidente está localizado em uma favela,
- indicação de que a emergência é uma reincidência (nas últimas 72 horas),
- para casos em que o cliente tenha uma emergência anterior nos últimos 7 dias:
- código de fechamento da emergência anterior
- quantidade de emergências anteriores.

Além disso, informações meteorológicas correspondentes às horas imediatamente anteriores ao incidente foram associadas a cada caso. Esses dados foram obtidos do banco de dados disponível no Instituto Nacional de Meteorologia (<https://portal.inmet.gov.br/>).

As informações sobre a região também foram integradas a cada registro, especialmente as características da rede elétrica, com base nas informações fornecidas pela Aneel (<https://dadosabertos.aneel.gov.br>). Dados históricos relacionados ao consumo de energia do cliente, número e duração das interrupções de serviço também foram incluídos.

A figura a seguir mostra o fluxo de trabalho para a caracterização de ocorrências.



Modelos

Diferentes abordagens de aprendizado de máquina foram avaliadas para obter modelos preditivos do conjunto de dados: regressão logística, árvores de decisão e *random forests*. Conforme desenvolvido por Barja Martínez et al. (2021), Yang et al. (2022), os métodos de *ensemble*, como os *random forests*, são particularmente adequados para lidar com problemas complexos e não lineares, como o que estamos tratando. De fato, Kurian et al. (2020) mostram que os *random forests* têm bons recursos para modelagem de incidentes. Nos últimos anos, os sistemas de deep learning foram investigados, aplicados e avaliados como modelos preditivos para diferentes aspectos do gerenciamento de eletricidade (Kardi et al. 2021, Elmasry 2022, Ammar and Kora 2023, Akhtar et al. 2023, entre muitos outros). Como os *random forests*, esses modelos

são flexíveis o suficiente para modelar os fenômenos complexos envolvidos em incidentes de eletricidade. No entanto, conforme argumentado por Coma-Puig et al. (2016, 2023), é preferível aplicar modelos interpretáveis quando o objetivo é integrá-los à operação manual, pois isso confere maior autonomia aos especialistas. Diferentes abordagens de machine learning interpretáveis foram investigadas, como *k-nearest neighbours* (Kumar and Vidya 2020) o *graph neural networks* (Kosasih et al. 2022), mas neste primeiro teste de viabilidade nos limitamos a explorar florestas aleatórias como um método de conjunto robusto e interpretável.

Métricas

Foram aplicadas métricas padrão baseadas em erros: accuracy, precision e recall por classe. Dessas métricas, a de interesse é o recall na classe de improcedentes, ou seja, quantos deslocamentos improcedentes são identificados automaticamente, para evitar desperdiçar tempo com tais deslocamentos. Uma segunda métrica de interesse é a precisão em improcedentes, para identificar os deslocamentos procedentes que são erroneamente identificados como improcedentes, o que significa sua despriorização e, portanto, um aumento no tempo necessário para resolver o incidente.

Análise dos resultados

Os resultados obtidos pelo melhor classificador, o *random forest*, foram os seguintes (o valor da métrica é relatado no momento 0 e com todas as informações disponíveis no final do evento):

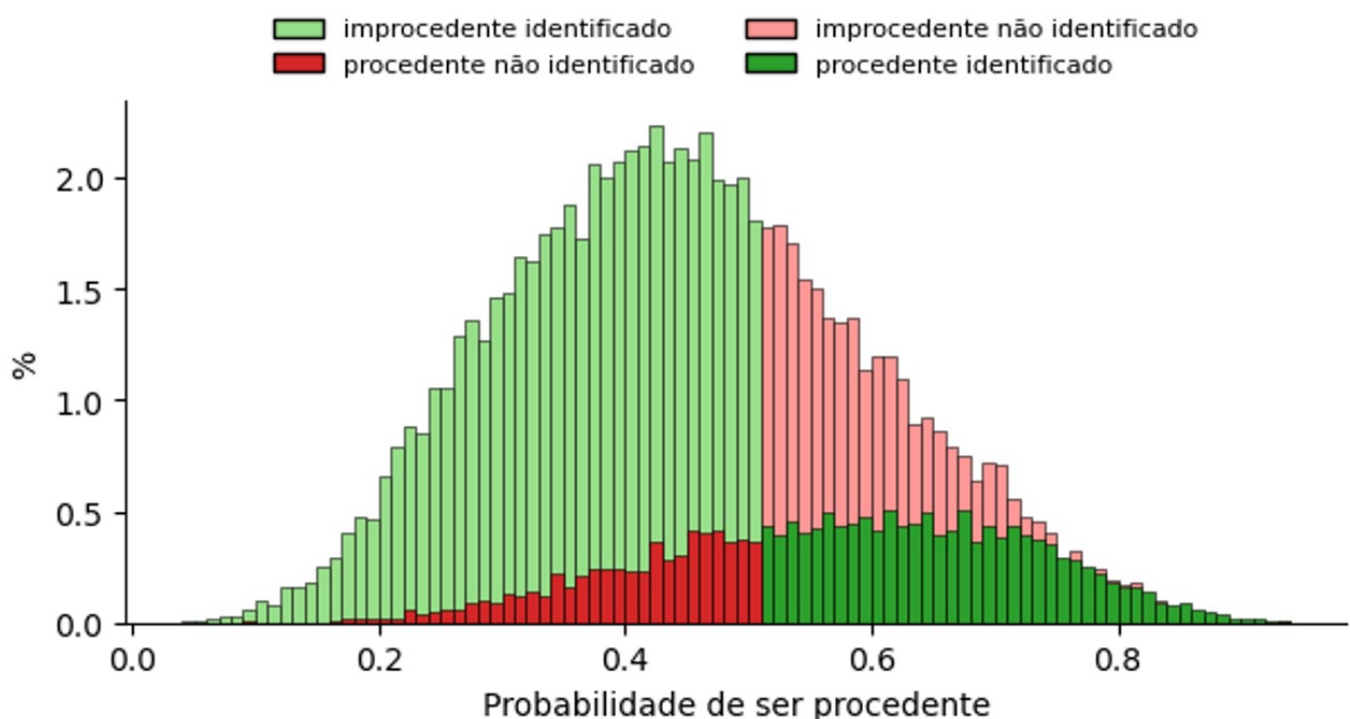
- *Precision* média: 63% - 68%.
- *Precision* deslocamento improcedente: 88% - 90%.
- *Precision* deslocamento procedente: 29% - 33%
- *Recall* deslocamento improcedente: 63% - 69%
- *Recall* deslocamento procedente: 64% - 66%

Esses resultados indicam que, quando o modelo identifica um caso como improcedente, a cada 10 casos, em 9 ele está correto (precisão) e pode economizar o deslocamento da equipe técnica para atender o caso. De todos os deslocamentos improcedentes, 6 a 7 em cada 10 são identificados corretamente (*recall*), de modo que 3 a 4 em cada 10 casos que são improcedentes não seriam identificados. O modelo poderia ser aprimorado para reduzir essa margem, incorporando informações adicionais ou mais dados históricos.

Por outro lado, o tratamento dos deslocamentos procedentes obtém métricas piores, o que é de se esperar, uma vez que se trata da classe minoritária, e os modelos de *machine learning* tendem a ter um desempe-

ho inferior para classes minoritárias. De fato, dos casos identificados como procedentes, apenas 30% são realmente procedentes. No entanto, esse resultado não nos parece particularmente prejudicial, pois uma melhora nesse indicador provavelmente significaria uma piora em outro indicador que consideramos mais prejudicial, por exemplo: o complemento de *recall* na classe de procedentes (silêncio, os casos procedentes que não conseguimos identificar como tal) ou o complemento de precisão na classe de improcedentes (erro, os casos procedentes que identificamos como improcedentes). Como já dissemos, esse tipo de erro, que nos parece ser o mais grave, afeta 6,7% do número total de casos, que são despriorizados para atenção por uma equipe técnica, embora não deversem. Esse tipo de erro é o que precisamos melhorar com a mais alta prioridade nas próximas etapas do projeto.

A figura mostra graficamente a relação entre o volume de casos corretamente identificados como improcedentes (verde claro) - que economizariam tempo para a equipe técnica - os casos improcedentes que não são identificados (vermelho claro) - que representam um uso ineficiente das equipes - os casos corretamente identificados (verde escuro) e os casos que não são identificados (vermelho escuro), que são os erros mais graves - casos que seriam despriorizados para o envio de uma equipe, embora realmente precisem disso.



Desenvolvimentos futuros

Nas fases seguintes deste projeto, integraremos as informações textuais associadas aos registros de serviço, principalmente a partir das observações dos operadores da central de atendimento. Além disso, também será realizada uma análise de *clustering* para identificar subtipos de incidentes que exigem tratamento diferenciado.

Processo de produção do sistema

Propõe-se que o sistema seja colocado em produção em um fluxo de trabalho no qual os casos detectados como improcedentes não sejam descartados, mas tratados com um *callback*, para verificar as causas mais prováveis da improcedência: se o incidente já foi resolvido, se diz respeito a áreas não técnicas, se requer um tipo específico de equipe, se requer detalhes sobre o endereço físico ou se requer especificar os

requisitos para o atendimento da equipe técnica. Dessa forma, não são descartados os erros mais graves - aqueles em que um caso que exige uma equipe técnica é identificado como improcedente.

A probabilidade de improcedência também é um dado valioso para priorizar os incidentes a serem atendidos. Em um trabalho futuro, integraremos essas informações ao método de priorização de incidentes.

3. Conclusão

Neste artigo, mostramos que uma abordagem baseada em machine learning pode contribuir para um melhor uso das equipes técnicas, identificando em um estágio inicial os casos em que a intervenção de uma equipe técnica não é necessária ou apropriada. De fato, um modelo preditivo inferido a partir de dados históricos do tratamento de incidentes da Light identifica corretamente 88% dos casos que não exigem a intervenção de uma equipe técnica no momento em que o incidente é relatado e 90% imediatamente antes do despacho. Essa porcentagem é particularmente valiosa, considerando que os casos improcedentes representam 80% dos incidentes relatados. Por outro lado, esse modelo comete apenas 6,7% de erros graves, nos quais os casos que necessitam de uma equipe técnica são despriorizados.

Nas próximas etapas do projeto, incorporaremos informações dos registros textuais do serviço da central de atendimento, bem como o texto de diferentes observações no processo. Além disso, também será realizada uma análise de *clustering* para identificar subtipos de incidentes que exigem tratamento diferenciado.

4. Referências bibliográficas

Ammar Mohammed, Rania Kora. (2023). A comprehensive review on ensemble deep learning: Opportunities and challenges, *Journal of King Saud University - Computer and Information Sciences*, Volume 35, Issue 2.

Saima Akhtar, Muhammad Adeel, Muhammad Iqbal, Abdallah Namoun, Ali Tufail, Ki-Hyung Kim (2023). Deep learning methods utilization in electric power systems, *Energy Reports*, Volume 10, Pages 2138-2151, ISSN 2352-4847.

Barja-Martinez, S., Aragüés-Peñalba, M., Munné-Collado, Í., Lloret-Gallego, P., Bullich-Massagué, E., & Villafafila-Robles, R. (2021). Artificial intelligence techniques for enabling Big Data services in distribution networks: A review. *Renewable and Sustainable Energy Reviews*, 150, 111459.

Coma-Puig B, Carmona J, Gavalda R, Alcoverro S, Martin V (2016) Fraud detection in energy consumption: a supervised approach. In: 2016 IEEE International Conference on Data Science and Advanced Analytics (DSAA). IEEE, pp 120–129

Coma-Puig, B., Calvo, A., Carmona, J. et al. A case study of improving a non-technical losses detection system through explainability. *Data Min Knowl Disc* 38, 2704–2732 (2024). <https://doi.org/10.1007/s10618-023-00927-7>

Wisam Elmasry, Mohammed Wadi, (2022). EDLA-EFDS: A Novel Ensemble Deep Learning Approach For Electrical Fault Detection Systems, *Electric Power Systems Research*, Volume 207, ISSN 0378-7796.

Raul S. Ferreira, Mauricio P. Dal Pont, Bruno M. A. da Silva, and Wendell W. Teixeira. 2018. RDI: A Real-time Decision Support System Applied to Dispatch Decision Problem. In Anais da V Escola Regional de Sistemas de Informação do Rio de Janeiro, outubro 16, 2018, Nova Friburgo, Brasil. SBC, Porto Alegre, Brasil, 108 - 115. DOI: <https://doi.org/10.5753/ersirj.2018.4664>.

Kardi, M., AlSkaif, T., Tekinerdogan, B., & Catalao, J. P. S. (2021). Anomaly detection in electricity consumption data using deep learning. In 2021 IEEE International Conference on Environment and Electrical Engineering and 2021 IEEE Industrial and Commercial Power Systems Europe (EEEIC / I&CPS Europe).

Edward Elson Kosasih, Fabrizio Margaroli, Simone Gelli, Ajmal Aziz, Nick Wildgoose & Alexandra Brintrup (2022) Towards knowledge graph reasoning for supply chain risk management using graph neural networks, International Journal of Production Research

Kumar, A., & Vidya, H. A. (2020). Application of k-Nearest Neighbor (kNN) Machine Algorithm for Fault Classification. International Journal of Advanced Science and Technology, 29, 8441-8448.

Daniel Kurian, Yongsheng Ma, Lianne Lefsrud, Fereshteh Sattari. (2020). Seeing the forest and the trees: Using machine learning to categorize and analyze incident reports for Alberta oil sands operators, Journal of Loss Prevention in the Process Industries, Volume 64, ISSN 0950-4230.

Portal do Instituto Nacional de Meteorologia (<https://portal.inmet.gov.br/>).

Portal de Dados Abertos da Aneel. <https://dadosabertos.aneel.gov.br>.

Yongquan Yang, Haijun Lv, and Ning Chen. 2022. A Survey on ensemble learning under the era of deep learning. Artif. Intell. Rev. 56, 6 (Jun 2023), 5545–5589. <https://doi.org/10.1007/s10462-022-10283-5>