APLICAÇÃO DE ANÁLISE DE DADOS PARA A IDENTIFICAÇÃO DE FURTO DE ENERGIA ELÉTRICA

Palavras-Chave: SISTEMAS DE DISTRIBUIÇÃO DE ENERGIA ELÉTRICA, CIÊNCIA DE DADOS, FURTO DE ENERGIA ELÉTRICA

Autores(as):

Ivan de Sousa Oliveira, FEEC - UNICAMP

Profª. Drª. Fernanda Caseño Trindade Arioli (orientadora), FEEC - UNICAMP

INTRODUÇÃO:

Em uma economia que busca a transição energética e sua eletrificação, a eficiência da distribuição é de suma importância, de forma que tratar as perdas de energia é fundamental para reduzir as perdas econômicas para a sociedade. Dentre os tipos de perdas de energia elétrica, se enquadram como perdas técnicas as perdas associadas a processos físicos na distribuição de energia, as quais são inevitáveis, enquanto perdas não técnicas são aquelas originadas por adulteração de medidores, furto e erros de medição. O furto de energia é um sério problema de energia em países emergentes como o Brasil, que pode vir a resultar em uma atribulação econômica que vai além das perdas do valor furtado, podendo alcançar danos a níveis infra estruturais ao sistema de distribuição devido à sobrecarga desse. De acordo com o relatório sobre perdas de energia elétrica da ANEEL [1], as perdas não técnicas alcançam cerca de 40,2 TWh, o que corresponde a 6,8% da energia injetada na rede de distribuição. Contudo, devido à metodologia adotada pela agência reguladora, é estabelecido um limite de repasse das perdas na tarifa ao consumidor, em que é repassado o que é considerado prejuízo dentro de uma medição eficiente.

Entre as fraudes que podem ser detectadas na inspeção, encontra-se a adulteração do medidor de energia, que pode ser realizada de diversas maneiras. A partir dessa visão sobre perdas não técnicas, este trabalho buscou analisar e trabalhar com os dados reais de uma distribuidora de energia elétrica, tentando levantar e estabelecer relações que envolvam a energia consumida, modelos e marcas de medidores utilizados, informações geográficas sobre a localização.

METODOLOGIA:

Durante a realização dos estudos, as atividades foram segmentadas em duas diferentes partes que foram realizadas ao longo do ano: Modelo baseado na análise de

medidores e Modelo baseado na análise de consumo. A análise de medidores ocorreu em um primeiro momento, e posteriormente ocorreu a análise sobre o consumo utilizando ferramentas já estabelecidas durante a primeira parte.

1. Análise de Medidores

Primeiramente foi realizada a separação das informações relacionadas a cada unidade consumidora conforme a figura 1, obtendo os modelos de medidor, coordenadas em X e Y, classificação econômica (CNAE) e histórico de inspeção. As coordenadas foram transformadas em retângulos de Geohash, uma forma de representar áreas segundo um sistema numérico de base 32 com extensão variável, de forma que quanto maior a extensão do código, maior será a precisão e menor será a área representada. Visando agrupar um número mínimo de unidades consumidoras por área, utilizou-se uma atribuição dinâmica para garantir que ao menor 70% dos Geohash tivessem pelo menos 15 unidades consumidoras.

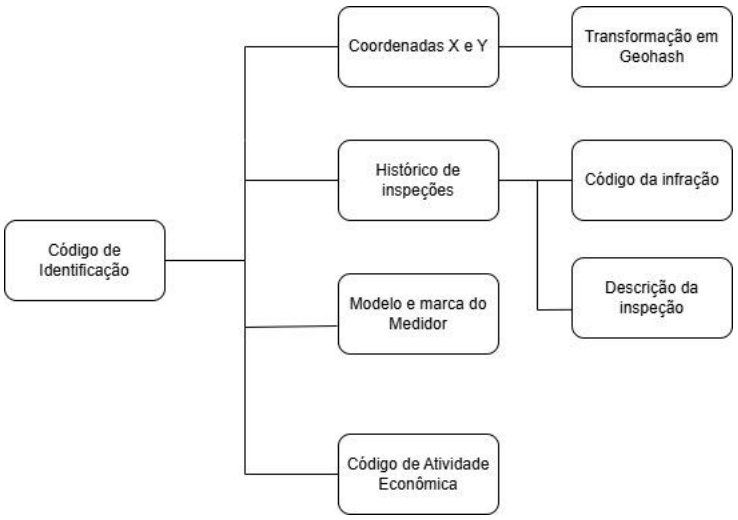

Figura 1 – Diagrama de representação dos dados para cada unidade consumidora.

2. Análise do Consumo

Utilizados os dados do modelo anterior em junção com as informações mensais de consumo coletadas mensalmente em cada unidade consumidora. Como nem todos os consumidores tiveram seus consumos coletados mensalmente, o que aconteceu levou a adoção de diferentes tipos de janelas, englobando alguns meses de consumo, para que seja possível corrigir as distorções causadas pela ausência de medição em determinados meses. Durante a análise foram utilizadas janelas de 3,4,5 e 6 meses de duração.

Foram empregados dois tipos de modelos para a análise visando tentar identificar anomalias no padrão de consumo de energia. Em um primeiro momento foi adotada um método de classificação chamado *Isolation Forest* utilizando entropia como peso, após isso foi adotada uma metodologia usando regressão LASSO e Ridge. Em ambas as análises se

utilizou agrupamento em uma determinada área, além de também utilizar uma separação entre os perfis de consumo de uma mesma atividade comercial.

O funcionamento da Isolation Forest [2] se baseia em árvores binárias com dois pontos extremos e os compara, eliminando o ponto mais isolado, desdobrando até varres todos os pontos ou atingir uma altura pré-definida para a arvore. Como incremento, são atribuídos pesos as entradas, pesos esses que são as somas das entropias do consumo das janelas temporais, dessa forma a densidade dos pontos começa a ser um fator a ser levado em ponto, dado que o tamanho dos pontos começa a ser um fator que os diferencia. Para o cálculo da entropia, ocorreu a normalização das janelas de consumo e a partir da densidade de probabilidade do consumo obteve-se a probabilidade, e então a entropia é calculada conforme a equação 1.

Como metodologia alternativa visando aumentar a precisão, utilizou-se um procedimento em duas etapas conforme a figura 2. Na primeira etapa, por meio de uma regularização do tipo LASSO [3], utilizando uma função *Softmax* como base, realiza-se uma seleção de variáveis para cada Geohash, assumindo que as influências variam por região. Na segunda etapa, realiza-se uma regularização do tipo Ridge utilizando as variáveis selecionadas anteriormente, ainda utilizando a função *Softmax* como base, em que seus valores de saída próximos a 0 correspondem a uma unidade de consumo regular enquanto suas saídas próximas a 1 sinalizam uma unidade fraudulenta.

$$H_s = \sum_{k=0}^{m} p_k \, log \frac{1}{p_k} \qquad \text{(Equação 1)}$$

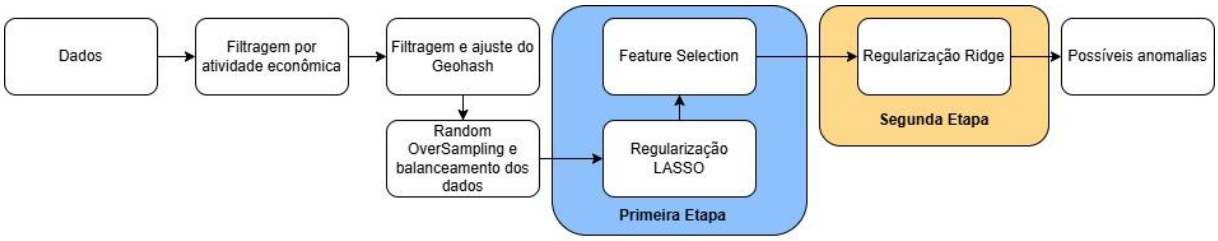

Figura 2 – Fluxograma referente ao procedimento em duas etapas.

RESULTADOS E DISCUSSÃO:

1. Análise de Medidores

Levantando as informações a respeito dos medidores e sua incidência por Geohash, conforme a figura 3, considerando diferentes tipos de medidores. Em um primeiro momento, realizando uma análise visual da concentração espacial dos medidores verificou-se que pontos de maior incidência. Ao passar para uma análise quantitativa, conclui-se que as regiões de maior incidência estão isoladas e fazem vizinhança apenas com Geohash de baixa incidência de adulterações.

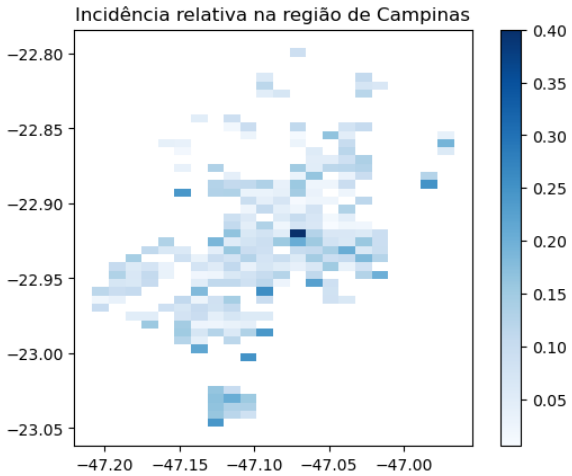

Figura 3 – Heat map relativo à incidência de irregularidades na região analisada.

Analisando o total dos medidores, verifica-se que existe uma predominância de fraude de medidores eletromecânicos, o que se pode explicar pela maior disponibilidade desses para instalação residencial ao mesmo tempo que existe uma maior facilidade de adulteração, em que não é necessário adulterar o circuito interno [4], diferente dos medidores eletrônicos. Ao analisar pela incidência, esse padrão de predominância eletromecânica persiste para medidores monofásicos e bifásicos, no entanto, para medidores trifásicos existe um maior equilíbrio e uma maior tendência recente na instalação de medidores eletrônicos em unidades com perfil de consumo mais alto, o que viabiliza operações mais custosas como a alteração da placa eletrônica.

2. Análise de Consumo

Para análise dos dois modelos, foram utilizadas as métricas F1-Score, acurácia e precisão para validação, as tabelas 1 e 2 apresentam os resultados médios obtidos tanto para o meio urbano quanto para o meio rural.

Tabela 1 – Resultados médios para zona rural.

Metodologia	F1 - score	Acurácia	Precisão
Isolation Forest	0,06	0,88	0,03
Duas etapas	0,23	0,86	0,18

Tabela 2 – Resultados médios para zona urbana.

Metodologia	F1- score	Acurácia	Precisão
Isolation Forest	0,02	0,90	0,01
Duas etapas	0,18	0,83	0,11

Em termos de precisão, os resultados obtidos por meio do procedimento por duas etapas foram superiores aos obtidos com Isolation Forest, mesmo que ambos os resultados resultem em médias baixas. Isso ocorre porque para ambos existem áreas irregulares, em que não são

obtidos nenhum verdadeiro positivo, resultando uma precisão igual a zero. A acurácia do Isolation Forest foi 2% e 7% maior para os casos de zona rural e urbana, respectivamente.

CONCLUSÕES:

Primordialmente, a maneira como a fraude é executada se diferencia entre regiões, Geohash, classificação de consumo, atividade econômica e modelo do medidor. Logo, a alteração de desempenho entre diferentes Geohash ocorre porque podem ter sidos realizados procedimentos diferentes que resultam em padrões opostos, de forma que mesmo realizando seleção de variáveis o resultado continua nulo. Além disso, foi utilizado dados reais de medição, que foram captados de maneira manual e estavam sujeitos a erros locais.

Dessa forma, é possível concluir que fatores materiais, geográficos e também humanos influenciam na precisão dos modelos adotados. O combate a perdas não técnicas é um problema de grande dificuldade de solução porque possui dados extremamente desbalanceados e não é raro ter dados rotulados erroneamente, indicando que o consumidor é regular, enquanto na verdade é irregular.

BIBLIOGRAFIA

[1] ANEEL, **Relatório de Perdas de Energia Elétrica na Distribuição**, 2025. [Online]. Disponível em https://git.aneel.gov.br/publico/centralconteudo/-/raw/main/relatorioseindicadores/tarifaeconomico/Relatorio_Perdas_Energia.pdf

[2] SANTANA, L. **"Detecção de Fraude com Isolation Forest"**. [Online]. Disponível em: https://medium.com/@lamartine_sl/detec%C3%A7%C3%A3o-de-fraude-com-isolation-forest-9ea50e4fe01b

[3] HAHN, G., LUTZ, S., LAHA, N., LANGE, C., **A framework to efficiently smooth L1 penalties for linear regression**, 2020. [Online]. Disponível em https://www.biorxiv.org/content/10.1101/2020.09.17.301788v3.full.pdf

[4] MIYASAKA, G., **"Análise do Desempenho de Medidores de Energia Elétrica Ativa em Condições Distorcidas e Desequilibradas"**, 2020. [Online]. Disponível em https://repositorio.ufu.br/bitstream/123456789/29009/3/An%C3%A1liseDesempenhoMedidores.pdf