Instituto Federal de Educação, Ciência e Tecnologia de São Paulo
Campus Sertãozinho

Guilherme Santos da Silveira, Victoria de Oliveira Spagiari

Aplicação de Sistemas Inteligentes para a Detecção de Fraudes em Redes Elétricas

Sertãozinho

2024

1 INTRODUÇÃO

O roubo de energia elétrica é um problema crescente e desafiador que afeta a eficiência e a sustentabilidade dos sistemas de distribuição de energia (Stracqualursi et al., 2023). Estima-se que uma parte significativa das perdas não técnicas nas redes elétricas seja decorrente de práticas fraudulentas, as quais impactam negativamente tanto as concessionárias quanto os consumidores. A Associação Brasileira de Distribuidores de Energia Elétrica (Aneel) destacou em seu relatório de 2023 que cerca de 6,7% da energia elétrica gerada, equivalente a 38 TWh, foi perdida devido a fraudes e outras perdas não técnicas. Tais práticas representam um prejuízo de aproximadamente R$9,9 bilhões para o setor, afetando diretamente as tarifas cobradas da população.

Identificar essas práticas fraudulentas, no entanto, é uma tarefa complexa. Estudos conduzidos por Czechowski e Kosek (2016) sugerem que existem cerca de 300 métodos diferentes para cometer fraudes no consumo de energia elétrica. Apesar dos avanços tecnológicos em equipamentos de medição, como os medidores inteligentes, a distinção entre dados legítimos e fraudulentos ainda representa um grande desafio para as concessionárias e para os pesquisadores da área.

Diante desse cenário, o problema de pesquisa deste trabalho é: treinar um sistema inteligente capaz de identificar eficientemente os casos de furto de energia elétrica a partir da vasta quantidade de dados gerados pelos sistemas de medição.

Para a solução de tal problema sugere-se aqui o uso do método *Support Vector Machine* (SMV) (Cortes, 1995), tendo como hipótese que o mesmo tem a capacidade de identificar padrões de consumo suspeitos que indiquem fraudes (Jindal et al., 2016). Outra hipótese é que a combinação de SVM com outras técnicas de análise de dados, como a MANOVA (*Multivariate Analysis of Variance*) (Hotelling, 1933; Karhunen, 1947; Loève, 1977), poderia aumentar a compreensão da dinâmica dos dados, o que por sua vez aumentaria a precisão na detecção de fraudes.

O objetivo geral deste trabalho é desenvolver um modelo de detecção de fraudes em redes elétricas utilizando o método *Support Vector Machine*. Os objetivos específicos incluem: (1) revisar a literatura existente sobre técnicas de detecção de fraudes em redes elétricas; (2) implementar o modelo de SVM utilizando dados de uma concessionária de energia; (3) avaliar o desempenho do modelo em termos de precisão e capacidade de generalização; (4) utilizar a MANOVA para compreender a dinâmica dos dados e (5) extrair conclusões a partir dos resultados obtidos.

Este trabalho é relevante tanto para a comunidade científica quanto para o setor elétrico, pois oferece uma solução potencial para um problema que causa bilhões de reais

em prejuízos anualmente. A implementação de um sistema eficaz de detecção de fraudes não só ajudará as concessionárias a reduzir perdas financeiras, mas também contribuirá para a justiça tarifária, garantindo que os custos da energia não sejam indevidamente elevados devido às práticas fraudulentas.

2 METODOLOGIA

O objetivo deste trabalho é treinar e aplicar o método inteligente *Support Vector Machine* (SVM) (Cortes, 1995) para identificar clientes fraudulentos em um banco de dados de uma companhia distribuidora de energia elétrica. O SVM será treinado para diferenciar entre padrões de consumo legítimos e fraudulentos, auxiliando na detecção de fraudes com maior precisão.

A pesquisa adota uma abordagem quantitativa, focada na análise de dados históricos fornecidos por uma companhia distribuidora de energia elétrica e disponibilizados gratuitamente na plataforma *Kaggle* (Samoshyn, 2019). Essa abordagem permite a aplicação de técnicas de aprendizado de máquina para modelagem preditiva e identificação de padrões associados ao furto de energia elétrica.

O desenvolvimento do projeto será conduzido utilizando a linguagem de programação *Python*, escolhida por sua robustez e extensa biblioteca de ferramentas para análise de dados e aprendizado de máquina. O método *Support Vector Machine* será implementado com a biblioteca *Scikit-Learn* (Pedregosa et al., 2011). A metodologia possui as seguintes etapas: tratamento de dados, treinamento do modelo, avaliação da assertividade e escrita do trabalho.

Inicialmente, será realizado o tratamento do banco de dados com informações do consumo de energia, uma vez que é comum que esses dados apresentem problemas como informações faltantes, valores discrepantes e outras inconsistências. Essa etapa é essencial para garantir a qualidade do treinamento do modelo.

Após o tratamento, o banco de dados será dividido em duas partes: 80% dos dados serão utilizados para o treinamento do SVM, enquanto os 20% restantes serão reservados para a validação do modelo, cuja eficácia será avaliada utilizando o coeficiente de determinação R^2 (Wright, 1921; Berk, 1977).

Além disso, será realizada uma Análise Multivariada de Variância (MANOVA) (Hotelling, 1933; Karhunen, 1947; Loève, 1977) para entender melhor os principais fatores que influenciam o furto de energia elétrica, permitindo uma análise mais detalhada dos dados e dos padrões detectados.

Por fim, a escrita da monografia será realizada com base nos dados e resultados obtidos ao longo do projeto. Para garantir a qualidade do texto, será utilizada uma ferramenta de Inteligência Artificial para correções ortográficas e ajustes gramaticais, proporcionando maior precisão e clareza na apresentação dos resultados.

REFERÊNCIAS

BERK, K. N. Tolerance and condition in regression computations. **Journal of the American Statistical Association**, Taylor & Francis, v. 72, n. 360a, p. 863–866, 1977. Citado na página 3.

CORTES, C. Support-vector networks. **Machine Learning**, 1995. Citado 2 vezes nas páginas 1 e 3.

CZECHOWSKI, R.; KOSEK, A. M. The most frequent energy theft techniques and hazards in present power energy consumption. In: IEEE. **2016 joint workshop on cyber-physical security and resilience in smart grids (CPSR-SG)**. [S.l.], 2016. p. 1–7. Citado na página 1.

HOTELLING, H. Analysis of a complex of statistical variables into principal components. **Journal of educational psychology**, Warwick & York, v. 24, n. 6, p. 417, 1933. Citado 2 vezes nas páginas 1 e 3.

JINDAL, A. et al. Decision tree and svm-based data analytics for theft detection in smart grid. **IEEE Transactions on Industrial Informatics**, v. 12, n. 3, p. 1005–1016, 2016. Citado na página 1.

KARHUNEN, K. Under lineare methoden in der wahr scheinlichkeitsrechnung. **Annales Academiae Scientiarun Fennicae Series A1: Mathematia Physica**, v. 47, 1947. Citado 2 vezes nas páginas 1 e 3.

LOÈVE, M. **Elementary probability theory**. [S.l.]: Springer, 1977. Citado 2 vezes nas páginas 1 e 3.

PEDREGOSA, F. et al. Scikit-learn: Machine learning in python. **Journal of machine learning research**, v. 12, n. Oct, p. 2825–2830, 2011. Citado na página 3.

SAMOSHYN, A. **Fraud Detection in Electricity and Gas Consumption**. 2019. Atualizado à 4 anos. Acesso em: 04-09-2024. Disponível em: <https://www.kaggle.com/datasets/mrmorj/fraud-detection-in-electricity-and-gas-consumption?select=SampleSubmission+%282%29.csv>. Citado na página 3.

STRACQUALURSI, E. et al. Systematic review of energy theft practices and autonomous detection through artificial intelligence methods. **Renewable and Sustainable Energy Reviews**, v. 184, p. 113544, 2023. ISSN 1364-0321. Citado na página 1.

WRIGHT, S. Correlation and causation. **Journal of agricultural research**, v. 20, n. 7, p. 557, 1921. Citado na página 3.