

Universidade de Pernambuco

Programa de Pós-Graduação em Engenharia da Computação (PPGEC)

Proposta de Tese de Doutorado

Área: Computação Inteligente

Título: Integração de Machine Learning e Métodos de Causa e Efeito para Análise de Dados

Orientadora: Roberta Andrade de Araújo Fagundes

Nome (Email): roberta.fagundes@upe.br

Descrição: A crescente disponibilidade de dados temporais em domínios como saúde, finanças, indústria e sistemas socioeconômicos tem ampliado o uso de técnicas de *Machine Learning* (ML) [1] para previsão e detecção de padrões. No entanto, modelos predominantemente preditivos apresentam limitações relevantes quanto à explicabilidade e à inferência causal, dificultando a compreensão dos mecanismos subjacentes aos fenômenos observados e a tomada de decisão informada em contextos dinâmicos e incertos. Esta pesquisa propõe o desenvolvimento de uma abordagem integrada que combine técnicas de *Machine Learning* com métodos formais de análise de causa e efeito [2] para o tratamento de dados. O objetivo central é avançar além da correlação, incorporando estruturas causais capazes de modelar dependências temporais, efeitos retardados (*lagged effects*) e relações não lineares, preservando desempenho preditivo e ampliando interpretabilidade e robustez analítica. Metodologicamente, o estudo investigará a integração de modelos de séries temporais baseados em ML, como redes neurais recorrentes, *temporal convolutional networks* e modelos baseados em atenção, com frameworks de inferência causal, incluindo grafos causais dinâmicos, modelos estruturais e abordagens inspiradas em *Granger causality* [3] e *causal Discovery* [4]. Serão propostas estratégias híbridas para aprendizado de estruturas causais em ambientes de alta dimensionalidade e não estacionariedade. A avaliação empírica será conduzida por meio de estudos de caso e experimentos com conjuntos de dados reais e sintéticos, considerando métricas de acurácia preditiva, estabilidade temporal, capacidade de recuperação de relações causais e utilidade para apoio à decisão. Como contribuição científica a proposição de um arcabouço metodológico que una desempenho computacional, rigor causal e explicabilidade, ampliando o potencial analítico de dados temporais complexos e fortalecendo a confiabilidade de sistemas inteligentes em contextos críticos. Como resultados esperados, pretende-se desenvolver um framework inovador para análise causal de dados; Avanços na interpretabilidade e confiabilidade de modelos de ML; Evidências empíricas da viabilidade de integrar ML e inferência causal em contextos reais e subsídios teóricos e práticos para aplicações em áreas que demandam decisões baseadas em dados.

Referências Bibliográficas:

- [1] RUSSELL, S.; NORVIG, P. *Artificial Intelligence: A Modern Approach*. 4. ed. Hoboken: Pearson, 2021.
- [2] List, J. A., Muir, I., & Sun, G. (2024). **Using machine learning for efficient flexible regression adjustment in economic experiments**. *Econometric Reviews*, 44(1), 2-40.
- [3] Granger, C. W. J. (1969). *Investigating causal relations by econometric models and cross-spectral methods*. *Econometrica*, 37(3), 424–438.
- [4] Spirtes, P., Glymour, C., & Scheines, R. (2000). *Causation, Prediction, and Search*. MIT Press.