

<p>Universidade de Pernambuco Programa de Pós-Graduação em Engenharia da Computação (PPGEC)</p>
<p>Proposta de Tese de Doutorado</p>
<p>Área: Computação Inteligente</p>
<p>Título: Inteligência Artificial Explicável para Modelagem Preditiva em Dados Públicos Heterogêneos</p>
<p>Orientador(a): Roberta Andrade de Araújo Fagundes (Email): roberta.fagundes@upe.br</p>
<p>A crescente disponibilidade de dados públicos tem impulsionado a utilização de técnicas de Ciência de Dados [1] e Inteligência Artificial em diversas áreas, incluindo saúde, educação, segurança pública, mobilidade urbana e gestão governamental. Apesar dos avanços recentes, a aplicação de modelos de Inteligência Artificial em cenários reais ainda enfrenta desafios relacionados à heterogeneidade dos dados, mudanças temporais nos padrões observados e falta de interpretabilidade dos resultados. Modelos avançados de aprendizado de máquina frequentemente alcançam elevado desempenho preditivo, porém operam como "caixas-pretas", dificultando a compreensão dos fatores que influenciam suas decisões. Em aplicações governamentais e de interesse público, a transparência e a confiabilidade dos modelos são requisitos fundamentais para sua adoção. Problema: Como desenvolver modelos de Inteligência Artificial capazes de realizar previsões precisas em bases de dados públicas heterogêneas, mantendo níveis adequados de interpretabilidade, robustez e adaptação às mudanças presentes nos dados? Hipótese: A combinação de técnicas modernas de aprendizado de máquina, aprendizado profundo e métodos de Inteligência Artificial Explicável (XAI) pode produzir modelos preditivos mais confiáveis, interpretáveis e adaptáveis a cenários reais caracterizados por heterogeneidade e mudanças de distribuição dos dados. Objetivo: Propor uma arquitetura baseada em Inteligência Artificial Explicável para modelagem preditiva em dados públicos heterogêneos, considerando aspectos de desempenho, interpretabilidade e adaptação a mudanças nos dados. Metodologia: A pesquisa seguirá uma abordagem quantitativa e experimental. Inicialmente, serão coletados datasets públicos oriundos de portais governamentais, bases socioeconômicas, educacionais e de saúde. Em seguida, serão calculadas métricas estatísticas e estruturais para caracterizar os dados e identificar diferentes níveis de heterogeneidade. Serão avaliados algoritmos [2] de regressão e classificação. Paralelamente, serão empregados métodos de explicabilidade,[3] como SHAP e LIME, para interpretação das previsões. Posteriormente, será proposta uma arquitetura capaz de selecionar dinamicamente modelos mais adequados em função das características dos dados e de possíveis mudanças temporais identificadas. Os experimentos serão avaliados por meio de métricas de desempenho, estabilidade, interpretabilidade e custo computacional. Resultados: Desenvolvimento de uma taxonomia para caracterização de datasets públicos; Proposição de um framework inteligente para modelagem preditiva adaptativa; Melhoria da interpretabilidade dos modelos de Inteligência Artificial; Produção de ferramentas e artefatos reproduzíveis para a comunidade científica ; Publicação de artigos em periódicos e conferências de alto impacto nas áreas de Ciência de Dados e Inteligência Artificial.</p>
<p>Referências Bibliográficas:</p> <p>[1] JAMES, G.; WITTEN, D.; HASTIE, T.; TIBSHIRANI, R. <i>An Introduction to Statistical Learning</i>. Springer, 2021</p> <p>[2] MONTGOMERY, D. C.; PECK, E. A.; VINING, G. G. <i>Introduction to Linear Regression Analysis</i>. 6. ed. Hoboken: Wiley, 2021.</p>

[3] GUNNING, D.; AHA, D. DARPA's Explainable Artificial Intelligence (XAI) Program.
AI Magazine, v. 40, n. 2, p. 44-58, 2019.