

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

**Proposta de Tese de Doutorado**

**Área: Computação Inteligente**

**Título:** Estratégias de IA Explicável para Compreensão do Desempenho, Avaliação e Otimização de Modelos de Regressão

**Orientador(a): Roberta Andrade de Araújo Fagundes**

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

**Descrição:** Modelos de regressão [1] são a espinha dorsal de inúmeras aplicações em diversos domínios, desde a previsão de preços de imóveis e demanda de energia até a estimativa de risco em saúde e finanças. Embora algoritmos complexos como *Gradient Boosting Machines* (GBMs) e Redes Neurais Profundas (DL) frequentemente atinjam desempenhos preditivos superiores, eles o fazem às custas da interpretabilidade. Essa característica de "caixa-preta" impede que usuários, reguladores e até mesmo os próprios desenvolvedores compreendam por que uma previsão específica foi feita, onde o modelo falha consistentemente ou como ele pode ser sistematicamente melhorado. Essa falta de transparência gera vários problemas: a dificuldade em depurar modelos, a impossibilidade de garantir conformidade com regulamentações (como a LGPD ou GDPR, que exigem "direito à explicação"), a desconfiança dos usuários e a incapacidade de extrair *insights* valiosos para o domínio de aplicação. A Inteligência Artificial Explicável (XAI) [2] surge como a ponte entre o alto desempenho preditivo e a necessidade de compreensão, mas ainda existem lacunas em como as técnicas de XAI podem ser sistematicamente integradas ao ciclo de vida completo de desenvolvimento e avaliação de modelos de regressão para otimizá-los de forma prática e eficaz. **Justificativa:** este projeto de doutorado se justifica pela crescente demanda por sistemas de IA não apenas precisos, mas também **confiáveis e auditáveis**. A integração profunda da XAI nas fases de compreensão do desempenho, avaliação e otimização de modelos de regressão representa uma fronteira de pesquisa. **Objetivos:** desenvolver, implementar e avaliar um *framework* sistemático para integrar estratégias de Inteligência Artificial Explicável (XAI) [3] nas fases de compreensão do desempenho, avaliação e otimização de modelos de regressão complexos. **Metodologia:** será experimental e de desenvolvimento, focada na construção e validação do *framework* proposto. **Resultados Esperados:** Um *framework* metodológico documentado para integrar XAI nas fases de compreensão, avaliação e otimização de modelos de regressão; Novas métricas ou abordagens para diagnosticar e visualizar padrões de erro e vieses em modelos de regressão utilizando XAI; Estratégias de otimização de modelo baseadas em XAI validadas experimentalmente, demonstrando melhorias no desempenho preditivo ou na robustez; Estudos de caso detalhados que comprovem a aplicabilidade e o valor do *framework* proposto em problemas reais de domínios críticos; Contribuições científicas através da publicação de artigos em periódicos de alto impacto e conferências renomadas nas áreas de Machine Learning, XAI, Ciência de Dados Aplicada ou domínios de aplicação; Um repositório de código contendo as implementações do *framework* e das estratégias propostas.

**Referências Bibliográficas:**

- [1] Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction* (2nd ed.). Springer.
- [2] Molnar, C. (2020). *Interpretable Machine Learning: A Guide for Making Black Box Models Explainable* (2nd ed.). Leanpub.
- [3] Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "Why Should I Trust You?": Explaining the Predictions of Any Classifier. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*, 1135-1144.