

# Universidade de Pernambuco

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

### Proposta de Tese de Doutorado

**Área: Inteligência Computacional**

**Título: Plataformas de Dados para o Ensino Superior: Proposta de uma Arquitetura Educacional Aberta e Escalável**

**Orientador – Alexandre Magno Andrade Maciel (amam@ecomp.poli.br)**

#### Descrição

Em uma era em que decisões estratégicas devem ser fundamentadas em dados, é essencial o uso de ferramentas que justifiquem as ações implementadas, definam métricas de desempenho e, principalmente, possibilitem o acompanhamento dos efeitos futuros das intervenções propostas[1].

A implementação de uma arquitetura de dados em instituições de ensino superior é alternativa para integrar e organizar informações provenientes de diferentes sistemas acadêmicos, administrativos e fontes externas. Essa estrutura permite não somente a padronização e confiabilidade dos dados, mas também estabelece uma base sólida para o desenvolvimento de políticas institucionais baseadas em evidências, promovendo maior transparência nos processos decisórios, otimização de recursos e apoio ao planejamento acadêmico e orçamentário[2].

Este projeto propõe desenvolver uma arquitetura de dados no contexto de dados educacionais do ensino superior, com ênfase na construção de uma arquitetura como plataforma para o monitoramento de indicadores educacionais. Essa arquitetura deve integrar dados institucionais e dados abertos governamentais, viabilizando a aplicação de técnicas de Machine Learning (ML) e Mineração de Dados Educacionais (Educational Data Mining – EDM). O objetivo é extrair conhecimento, gerar previsões sobre o desempenho estudantil, identificar padrões de evasão por meio de clustering e interpretar os fatores relevantes para tais previsões com base em abordagens de Inteligência Artificial Explicável (Explainable AI – XAI)[3][4].

A implementação dessa arquitetura em larga escala envolve desafios como a escolha de tecnologias escaláveis e de fácil manutenção, especialmente em contextos públicos, bem como a integração de diferentes fontes de dados — tanto internos (como Moodle, SIGA/SIGAA) quanto públicas (como os dados e indicadores do INEP). Como contribuição, o processo completo de arquitetura, ingestão, preparação e disponibilização de dados, acompanhado de metadados adequados, cria um ambiente propício para a aplicação de técnicas avançadas de análise de dados.

Este projeto objetiva que tal estrutura contribua com a gestão educacional, promovendo intervenções baseadas em evidências e o monitoramento contínuo por meio de dashboards, além de possibilitar a aplicação de algoritmos supervisionados e não supervisionados para apoio à tomada de decisão.

#### Referências Bibliográficas

- [1] Baker, R.S., Inventado, P.S. (2014). **Educational Data Mining and Learning Analytics**. In: Larusson, J., White, B. (eds) Learning Analytics. Springer, New York, NY.
- [2] Borges, Giuliano Alves, et al. **Parceria entre governo local e universidades: formação do Observatório da Cidade de Macaé no contexto de rede colaborativa interinstitucional e intersetorial**. Revista Vértices 22.3 (2020): 412-428.
- [3] Baepler, P., e Murdoch, C. J. "Academic analytics and data mining in higher education." International Journal for the Scholarship of Teaching and Learning 4.2 (2010): 17.
- [4] Marques, Ebony, et al. **Sabia: Uma plataforma para auxiliar a gestão baseada em evidências nas instituições de ensino superior**. Workshop de Aplicações Práticas de Learning Analytics em Instituições de Ensino no Brasil (WAPLA). SBC, 2023.