



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

Proposta de Dissertação de Mestrado

Area: Banco de Dados e Inteligência Artificial

Título: Análise e Descoberta de Padrões em Dados Abertos

Orientador: Dimas Cassimiro do Nascimento Filho (email:) dimas.cassimiro@ufape.edu.br

Co-orientador (se existir): Bruno Costa e Silva Nogueira (email:) bruno@ic.ufal.br

Resumo

O crescente volume de dados disponíveis publicamente representa uma fonte de dados relevante para a descoberta de padrões e análises inteligentes que podem beneficiar a sociedade em diversas áreas, tais como saúde, educação, orçamento público e meio ambiente. A análise de padrões em dados abertos é crucial para compreender fenômenos complexos, identificar tendências emergentes e apoiar a tomada de decisões. Este tema de pesquisa visa contribuir para o avanço da capacidade de explorar e interpretar conjuntos de dados abertos disponíveis, promovendo assim a utilização eficaz dessas informações para benefício da sociedade.

O objetivo deste tema consiste em investigar métodos avançados de análise de dados, técnicas de mineração de dados e algoritmos de Inteligência Computacional para identificar padrões significativos e tendências em conjuntos de dados abertos. Além disso, pretende-se desenvolver novas estratégias para visualização de dados que possam auxiliar na compreensão e interpretação desses dados. Contribuindo assim, para a disseminação de dados disponíveis em bases de dados públicas e para aprimorar o processo decisório baseado em dados abertos disponíveis.

Dentre as contribuições específicas que podem ser desenvolvidas, podem ser destacadas:

- 1) A criação de ferramentas para visualização de dados abertos;
- 2) A aplicação e adaptação de técnicas de Ciência de Dados para análise descritiva de dados abertos;
- 3) A aplicação e combinação de modelos de Inteligência Computacional para prever



resultados em dados abertos;

- 4) A aplicação e adaptação de técnicas de mineração de dados para a descoberta de padrões não triviais em dados abertos;
- 5) Modelagem de bancos de dados para representação e otimização de consultas sobre dados abertos.

A identificação e compreensão de padrões não triviais em dados abertos não apenas aprimora a utilidade dos dados disponibilizados, mas também impulsiona a inovação, permitindo a formulação de políticas mais informadas e a criação de soluções personalizadas para desafios complexos. Além disso, a aplicação de técnicas de ciência de dados no contexto de dados abertos poderá produzir novas contribuições científicas, por meio da análise de uma ampla variedade de dados abertos em diferentes contextos.

Referências Bibliográficas

[1] SANTOS, Philippe; FERREIRA, Rafael; MIRANDA, Péricles. Dados abertos educacionais: Uma revisão da literatura brasileira. In: Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação-SBIE). 2017. p. 11.

[2] FERREIRA, Leandro A.; RODRIGUES, Rodrigo L.; DE SOUZA, Rodrigo NPM. Dados abertos educacionais brasileiros: Um mapeamento sistemático da literatura. In: Anais do XXXII Simpósio Brasileiro de Informática na Educação. SBC, 2021. p. 1186-1195.

[3] BITTENCOURT, Cinthya; ESTIMA, Jacinto; PESTANA, Gabriel. Open Data Initiatives in Brazil. In: 2019 14th Iberian Conference on Information Systems and Technologies (CISTI). IEEE, 2019. p. 1-4.

[4] FERNANDES, Eduardo et al. Educational data mining: Predictive analysis of academic performance of public school students in the capital of Brazil. Journal of business research, v. 94, p. 335-343, 2019.

[5] DA FONSECA SILVEIRA, Rodrigo et al. Educational data mining: Analysis of drop out of engineering majors at the UnB-Brazil. In: 2019 18th IEEE International Conference On Machine Learning And Applications (ICMLA). IEEE, 2019. p. 259-262.

