

Universidade de Pernambuco

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

Proposta de Tese de Doutorado

Área: **Computação Inteligente**

Título: **Uma Abordagem de *Machine Learning* para aplicações e processo de análise de dados**

Orientador – **Roberta Andrade de A. Fagundes** (roberta.fagundes@upe.br)

Data Science e *Machine Learning* são dois termos abordados em todos os âmbitos de desenvolvimento e evolução de soluções. *Data Science* [1] diz respeito a trabalhar com grandes volumes de dados, checando a veracidade, fazendo análises e descobrindo novas informações. *Machine Learning* [2][3] trata da classificação e/ou predição de resultados, gerando novos dados e informações através do aprendizado e análise de padrões a partir de dados históricos. Esta linha investiga técnicas, métodos e ferramentas que permitam a extrair valor agregado de grandes volumes de dados bruto disponíveis em diversas fontes e de vários formatos. A linha desenvolve pesquisas voltadas: i) ao desenvolvimento de técnicas, métodos e ferramentas de extração de conhecimento a partir de grandes volumes de dados; ii) desenvolvimento de aplicações de integração, mineração e análise de dados em diferentes domínios; iii) a análise, descoberta e compreensão de fenômenos através da ciência de dados; iv) a técnicas e modelos computacionais que tornem o processo de análise e dados escalável e reproduzíveis.

- **Processo de Descoberta de Conhecimento:** desenvolve pesquisa exploratória acerca da extração de conhecimento a partir de grandes volumes de dados. O processo envolve a coleta, integração, pré-processamento, mineração e interpretação dos resultados;
- **Ciência de Dados (data science):** investiga o uso de técnicas estatísticas, algoritmos de aprendizado de máquina, analítica visual, entre outros, para compreensão de problemas e fenômenos a partir de grandes volumes de dados;
- **Escalabilidade em tempo e espaço da análise de dados:** investiga técnicas estatísticas de agregação de dados e de otimização de código para aumentar a quantidade de dados que podem ser analisados ou para diminuir o tempo de análise com ferramental existente;

O objetivo da abordagem proposta é trazer uma imersão em conceitos relacionados a *Machine Learning*, aplicado à análise de dados, de modo a capacitar os participantes a trabalharem com essa tecnologia criando soluções baseadas em dados, aplicando métodos e processo amplamente validados pelo mercado. Durante a pesquisa um ambiente experimental para avaliação dessa abordagem composto por modelos de inteligência computacional serão implementados, como também, a utilização de dados (artificiais/reais).

Referências Bibliográficas

- [1] IWASHITA, Adriana Sayuri; PAPA, João Paulo. An overview on concept drift learning. **IEEE Access**, v. 7, p. 1532-1547, 2018.
- [2] NORVIG P. e RUSSELL S. **Inteligência Artificial**, 3ª Edição, 2013.
- [3] MONTGOMERY D.C.; PECK, E.A. e VINING, G. G. **Introduction to Linear Regression Analysis**, Wiley-Interscience, 2006.