

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

**Proposta de Dissertação de Mestrado**

**Área: Inteligência Computacional**

**Título: Ciência dos Dados: Uma abordagem para Análise de Indicadores Sociais usando métodos de *machine learning***

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

A preocupação latente em nossa sociedade devido às suas implicações para a população exposta e aos possíveis problemas de educação, habitação e saúde podem causar [4][5]. Nesse sentido, esta pesquisa contribui para a identificação quais desses setores de acordo com as características dos moradores, apresentadas por indicadores sociais (ou seja, idade, escolaridade, renda, entre outros) classificados em três categorias de avaliação (educação, habitação e saúde). Em seguida, esses dados [6] serão processados e analisados por meio de dois algoritmos inteligentes. A inteligência computacional [1] é uma técnica tem como objetivo fazer as máquinas resolverem problemas por meio da observação das pessoas ou apenas por meio do estudo de seus próprios métodos como por exemplo, regressão [3], redes neurais, árvores de decisão, algoritmos genéticos etc. Na Inteligência Artificial [2], por sua vez, destacam-se os estudos em processamento de linguagem natural (*natural language processing*), reconhecimento de padrões (*pattern recognition*) e aprendizado de máquina (*machine learning*) que podem ajudar a extrair dos grandes volumes de dados (estruturados ou não-estruturados) conhecimento para auxiliar a tomada de decisões. Nessa proposta, algoritmos inteligentes serão utilizados para compor a abordagem da análise de indicadores sociais facilitando a tomada de decisão. Durante o estudo um ambiente experimental para avaliação dos algoritmos implementados. Nesse ambiente experimental serão utilizados dados reais do IBGE (Instituto Brasileiro de Geografia e Estatística) [6] e o desempenho do modelo será medido pelo erro de predição estimado pelo método Monte Carlo no conjunto de teste.

**Referências Bibliográficas**

1. ALPAYDIN, E. (2020). Introduction to machine learning. MIT press.
2. Rodrigo F Mello, Moacir Antonelli Ponti. Machine Learning: A Practical Approach on the Statistical Learning Theory, 2018.
3. Montgomery D.C.; Peck, E.A. ; Vining,G. Introduction to Linear Regression Analysis, Wisley Interscience, 2006.
4. Luis Izquierdo Horna, Miker Damazo, Deyvis Yanayaco. Identification of urban sectors prone to solid waste accumulation: A machine learning approach based on social indicators. (<https://doi.org/10.1016/j.compenvurbsys.2022.101834>).
5. Sergey Smetanin; Mikhail Komarov. Misclassification Bias in Computational Social Science: A Simulation Approach for Assessing the Impact of Classification Errors on Social Indicators Research (10.1109/ACCESS.2022.3149897).
6. <https://www.ibge.gov.br/estatisticas/sociais/populacao/9221-sintese-de-indicadores-sociais.html?=&t=resultados>.