

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

Proposta de Dissertação de Mestrado

Área: Computação Inteligente

Título: Aplicação de Modelos de Regressão Baseados em Machine Learning para Predição em Dados Públicos

Orientador(a): Roberta Andrade de Araújo Fagundes
(Email): roberta.fagundes@upe.br

Descrição: A crescente disponibilização de dados públicos por órgãos governamentais tem criado oportunidades para o desenvolvimento de soluções analíticas capazes de apoiar a tomada de decisão baseada em evidências. Nesse cenário, técnicas de Machine Learning [1], especialmente modelos de regressão, têm sido amplamente utilizadas para identificar padrões complexos e realizar previsões precisas em diferentes domínios, como saúde, educação, segurança pública e economia. Entretanto, a diversidade e a heterogeneidade dos dados públicos ainda representam desafios para a construção de modelos preditivos robustos e generalizáveis. **Problema:** Qual é o desempenho de diferentes algoritmos de regressão baseados em Machine Learning na predição de variáveis de interesse a partir de bases de dados públicas brasileiras? **Objetivo:** Investigar e comparar o desempenho de modelos de regressão baseados em Machine Learning aplicados a dados públicos, visando identificar técnicas mais adequadas para tarefas de previsão em cenários reais. **Objetivos específicos:** Coletar e integrar dados provenientes de portais de dados abertos governamentais; Realizar pré-processamento, limpeza e transformação dos dados;; Implementar modelos de regressão linear e algoritmos avançados de Machine Learning; Comparar o desempenho preditivo dos modelos utilizando métricas estatísticas e Identificar fatores que influenciam a qualidade das previsões. **Metodologia:** Inicialmente, serão selecionadas bases de dados públicas disponibilizadas por órgãos governamentais, como o Portal Brasileiro de Dados Abertos, DATASUS e IBGE. Após a etapa de preparação dos dados, serão aplicadas técnicas de análise exploratória e engenharia de atributos. Serão implementados e comparados diferentes algoritmos de regressão[2]. A avaliação dos modelos será realizada por meio de validação cruzada e métricas como Erro Médio Absoluto (MAE), Raiz do Erro Quadrático Médio (RMSE), Coeficiente de Determinação (R^2) e Erro Percentual Absoluto Médio (MAPE). **Resultados:** Espera-se identificar quais modelos de regressão apresentam melhor capacidade preditiva em diferentes tipos de dados públicos, bem como compreender o impacto das características dos dados sobre o desempenho dos algoritmos. Os resultados poderão subsidiar aplicações práticas em gestão pública e apoiar futuras pesquisas na área de Ciência de Dados. **Contribuições:** A pesquisa contribuirá para a avaliação sistemática de técnicas modernas de regressão em cenários reais, fornecendo evidências experimentais sobre sua aplicabilidade em bases governamentais abertas. Além disso, pretende-se disponibilizar um protocolo experimental que possa ser reproduzido em diferentes domínios de aplicação.

Referências Bibliográficas:

[1] JAMES, G.; WITTEN, D.; HASTIE, T.; TIBSHIRANI, R. *An Introduction to Statistical Learning*. Springer, 2021

[2] MONTGOMERY, D. C.; PECK, E. A.; Vining, G. G. *Introduction to Linear Regression Analysis*. 6. ed. Hoboken: Wiley, 2021.