

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

**Proposta de Projeto de Doutorado**

**Área: Inteligência Computacional**

**Título: Uma abordagem para solucionar problemas de multicolinearidade em modelos de regressão semi-supervisionada com *ensemble***

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

**Co-orientador – Renata Maria Cardoso Rodrigues de Souza ([rmcrs@cin.ufpe.br](mailto:rmcrs@cin.ufpe.br))**

**Descrição**

Análise de Regressão, segundo Montgomery et al. [1], é uma ferramenta estatística que tem como objetivo realizar a estimação de valores ( $y$ ) e explicar o relacionamento entre variáveis através de modelos matemáticos. Essas variáveis são divididas em dois grupos: variáveis independentes ( $x$ ); e variáveis dependentes ( $y$ ). Uma das principais dificuldades encontradas por pesquisadores em análise de regressão é conseguir trabalhar com conjuntos de dados que possuem sérios problemas de multicolinearidade [2] e/ou de alta dimensionalidade, em que o número de co-variáveis é maior do que o número de observações. Há multicolinearidade em um modelo de regressão ocorre quando duas ou mais variáveis independentes ( $x$ ) são fortemente relacionadas linearmente entre si. Estas situações podem causar resultados instáveis nos métodos de regressão por mínimos quadrados ordinários, além do aumento da variância dos coeficientes estimados. Assim, modelos de regressão semi-supervisionado com *ensemble* [3] se tornou, recentemente, uma boa alternativa para aumentar a capacidade de generalização de modelos de aprendizagem de máquinas. Esses modelos são resultantes da integração de dois ou mais algoritmos de tipos semelhantes ou diferentes com o objetivo de criar um sistema mais robusto que incorpore o resultado de todas as técnicas utilizadas, aumentando a robustez e acurácia.

Esse projeto visa propor uma abordagem para resolver problemas de multicolinearidade baseados em técnicas de Estatística e Aprendizagem de Máquina usando regressão semi-supervisionada com *ensemble*.

**Referências**

1. Douglas C Montgomery, Elizabeth A Peck, and G Georey Vining. Introduction to linear regression analysis, volume 821. John Wiley & Sons, 2012.
2. Duzan, H.; Shariff, N. S. B. M. Ridge Regression for Solving the Multicollinearity Problem: Review of Methods and Models, Journal of Applied Sciences 15 (3): 392-404, 2015.
3. Joao Mendes-Moreira, Carlos Soares, Alípio Mário Jorge, and Jorge Freire De Sousa. Ensemble approaches for regression: A survey. ACM Computing Surveys (CSUR), 45(1):10, 2012. [1]  
[SEP]