

# Universidade de Pernambuco

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

### Proposta de Projeto de Doutorado

**Área: Computação Inteligente**

**Título: Combinação de *Extreme Learning Machine* e Regressão para problema de *big data* no combate a mortalidade infantil**

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

Na comunidade de aprendizagem de máquina, um dos tópicos de crescente interesse é a pesquisa de novas arquiteturas de aprendizado profundo (*Deep Learning*) para problemas de classificação e regressão em grandes massas de dados [1]. Nesse contexto, um dos principais gargalos para a utilização de redes de aprendizado profundo é o seu elevado tempo de aprendizado. O treinamento de unidades em diversas camadas empilhadas é um problema não trivial, devido à complexidade do espaço de busca dos parâmetros da rede. Uma possível solução para redução da complexidade é a utilização da *Extreme Learning Machine* (ELM) [2], onde um conjunto mínimo de parâmetros precisa ser ajustado durante o processo de aprendizado. Entretanto, uma das desvantagens do ELM é a escolha aleatória dos pesos da camada oculta, o que poderá ocasionar problemas de multicolinearidade e aumento da variância nas estimativas dos modelos gerados. Para isso, os modelos de regressão Ridge e LASSO [3] resolvem o problema de multicolinearidade acrescentando um pequeno vício ao estimador dos coeficientes de regressão encontrado por mínimos quadrados ordinários, afastando o sistema da singularidade.

Dessa forma, o objetivo deste projeto é propor modelos que combinam extensões do ELM com modelos de regressão: Ridge e LASSO, para caracterizar o problema a partir de uma visão de *big data analytics* integrando e estruturando, entre outras, as bases de dados: DataSUS, IPEA e IBGE que englobam aspectos desde os sanitários até os econômico-sociais da população do estado de Pernambuco. Assim, a partir dessa base integrada, identificar fatores que influenciam o problema da mortalidade infantil e apresente uma solução algorítmica para o mesmo [4][5], fornecendo propostas baseadas em dados que possam contribuir para a elaboração de políticas públicas para o estado de Pernambuco. Para isso, serão realizadas análises e extraído conhecimento dos dados visando a identificar os fatores que influenciam a mortalidade infantil e estimar o impacto que as políticas propostas produziram na taxa de mortalidade, caso viessem a ser implantadas pelo governo do estado de Pernambuco.

Espera-se que a aplicação da técnica de mineração de dados permite a criação de um modelo de previsão utilizando combinação ELM com regressores, a fim de promover avanços na área de saúde.

#### Referências Bibliográficas

- [1] Y. Bengio. Learning Deep Architectures for AI. in Foundations and Trends in Machine Learning, 2(1):1–127, 2009.
- [2] G. Huang, G. B. Huang, S. Song, K. You. Trends in Extreme Learning Machines: A Review. Neural Networks, 61, 2015, pp 32-48.
- [3] MONTGOMERY D.C.; PECK, E.A. e VINING, G. G. Introduction to Linear Regression Analysis, Wiley-Interscience, 2006.
- [4] Liu L, Johnson HL, Cousens S, Perin J, Scott S, Lawn JE, et al Global, regional, and national causes of child mortality: an updated systematic analysis for 2010 with time trends since 2000. Lancet. 2012 Jun;379(9832):2151-61.
- [5] Araújo BF, Zatti H, Madi JM, Coelho MB, Olmi FB, Canabarro CT. Analysis of neonatal morbidity and mortality in late-preterm newborn infants. J Pediatr. 2012 May-Jun;88(3):259-66.