

Predição de Eficiência Energética em Aplicações de HPC

Isabella Muniz^{1,2}, Micaella Coelho¹, Alexandre Porto¹,
Hiago Rocha¹, Kary Ocaña¹, Carla Osthoff¹

¹Laboratório Nacional de Computação Científica, LNCC, Petrópolis, RJ - Brasil

²Faculdade de Educação Tecnológica do Rio de Janeiro, FAETERJ, Petrópolis, RJ - Brasil

Motivação

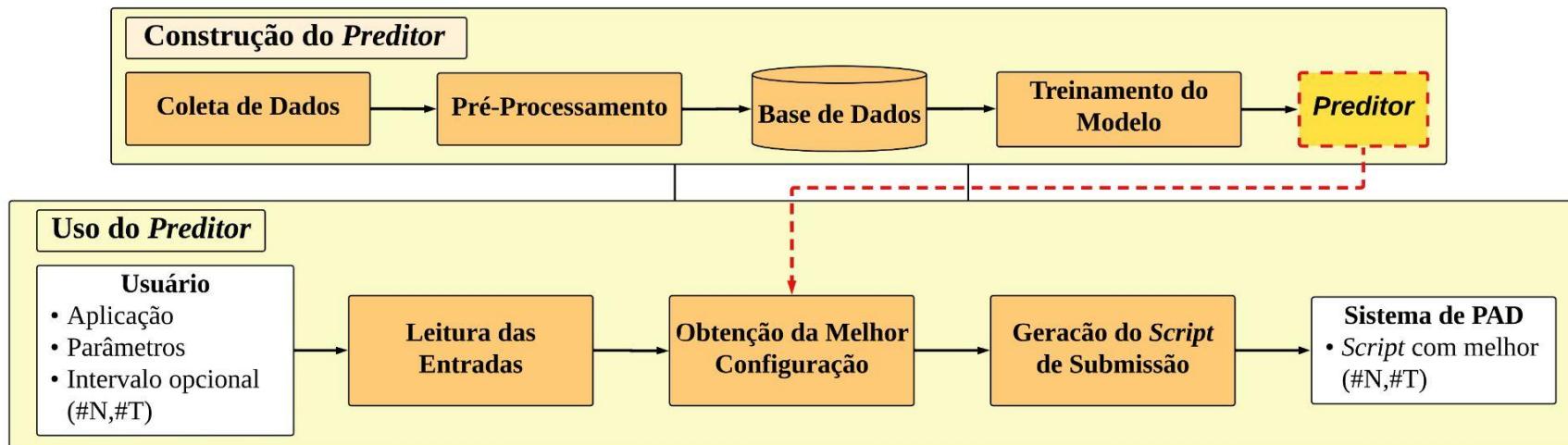
- O consumo de energia em ambientes de Computação de Alto Desempenho tem se consolidado como um desafio, em razão da crescente demanda por aplicações científicas de larga escala.
- Decisões relativas à alocação de recursos computacionais, tais como a definição do número de nós ou a escolha do número de threads por nó, podem exercer um impacto considerável no desempenho e na eficiência energética das aplicações.

Objetivos

- O objetivo deste trabalho é verificar a eficácia de uma estratégia preditiva baseado em Extra Trees Regressor (ETR) para estimar o consumo energético (EDP).
 - A avaliação é realizada com base no erro médio absoluto (MAE) e nas importâncias das variáveis, visando analisar a eficiência da abordagem.

Estratégia de Predição

- **Construção do Predictor:** coleta de dados, pré-processamento e treinamento do modelo ETR.
- **Uso do Predictor:** leitura de parâmetros de entrada, previsão da melhor configuração (número de nós e número de threads) e geração automática do script de submissão no SLURM.



Fonte: Adaptado de Coelho et al. (2025).

Estratégia de Predição

- **Energy-Delay-Product (EDP):** Combinação do consumo energético e tempo de execução, permitindo avaliar o trade-off desempenho-energia.
 - Com base nisso, EDP foi calculado como: $EDP(app) = \text{Energia}(app) \times \text{Tempo}(app)$.
- **Extra Trees Regressor (ETR):** É um algoritmo de ensemble que combina várias árvores de decisão totalmente expandidas.
- **Leave-One-Group-Out (LOGO):** Uma técnica de validação cruzada que divide os dados em grupos mutuamente exclusivos. Em cada iteração, um grupo é excluído do treinamento e usado para teste, enquanto os demais servem para treinar o modelo.

Metodologia Experimental

- **Ambiente computacional:** 2 Intel Xeon 6252 (48 núcleos), 384 GB RAM, InfiniBand, gerenciamento de jobs pelo SLURM.
- **Aplicação avaliada e execuções:** O estudo de caso utilizou a aplicação RAxML.
 - Variando: bootstrap {100, 1000}; 8 arquivos de entrada DENV com tamanhos variando de 187.026 a 484.520 bytes, número de threads {2, 24, 48} e número do nós {1 e 5}.
- **Medidas de desempenho:** Tempo de Execução coletado com o sacct (SLURM) e Consumo de Energia coletado com (RAPL).
- **Avaliação do modelo:** O modelo foi avaliado com validação cruzada LOGO.

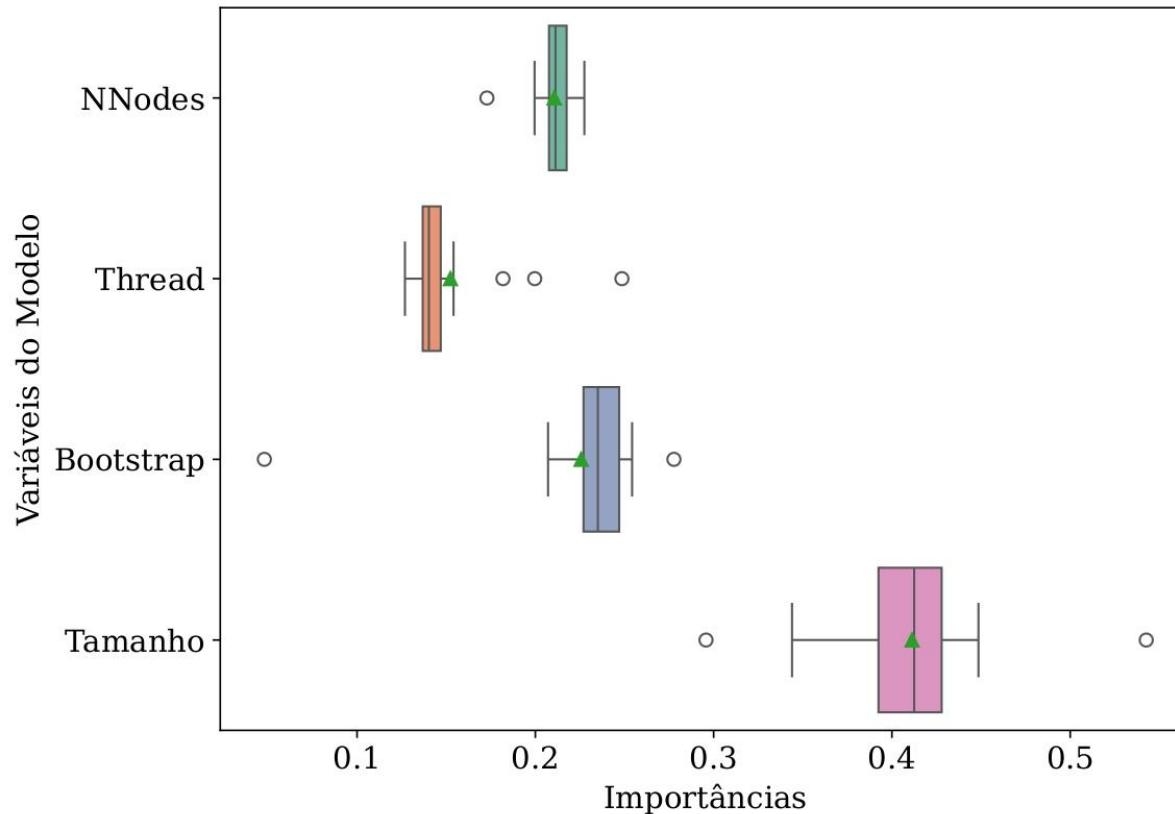
Resultados

- A análise do desempenho do modelo através do erro médio absoluto: o **MAE** global foi **0,05**.

Métrica	Valor
count	16.00
mean	0.05
std	0.10
min	0.00
25%	0.00
50%	0.01
75%	0.02
max	0.29

Resultados

- A variável com maior importância é o Tamanho da entrada.
- A variável com menor importância é a Thread.



Conclusão e Trabalhos Futuros

- Este trabalho avaliou uma metodologia para predição da eficiência energética de execuções em sistemas de HPC.
- Os resultados indicaram que a abordagem é capaz de fornecer previsões com baixo erro, obtendo um MAE global de 0,05 em diferentes cenários de entrada.
- Como trabalhos futuros, pretende-se ampliar a avaliação do modelo com benchmarks representativos de HPC.

Referências

- Lorenzon, A. F., Beck, A. C. S., Navaux, P. O., and Messer, B. (2025). Energy-efficient gpu allocation and frequency management in exascale computing systems. In ISC High Performance 2025 Research Paper Proceedings (40th International Conference), pages 1–11. Prometheus GmbH.
- Porto, A. H., Coelho, M., Rocha, H. M., Osthoff, C., Ocaña, K., and Cardoso, D. O. (2026). Assuming the best: Towards a reliable protocol for resource usage prediction for high-performance computing based on machine learning. Future Generation Computer Systems, 175:108070.
- Coelho, M., Porto, A., de A. Rocha, H. M. G., Muniz, I., Cardoso, D. O., Ocaña, K., Lorenzon, A., Navaux, P. O. A., and Osthoff, C. (2025). Modelagem preditiva de energy-delay product para otimização de submissões em supercomputadores. In Anais do Simpósio em Sistemas Computacionais de Alto Desempenho (SSCAD), Brasil. Sociedade Brasileira de Computação (SBC).

Obrigada!