

Cluster MPI com Raspberry Pi para Treinamento Distribuído de Modelos LightGBM e Mapeamento de Sombreamento em Usinas Fotovoltaicas

Vagner Souza¹, Wendell Rodrigues¹, Rejane Sá²

¹Programa de Pós-Graduação em Ciência da Computação-IFCE

²Laboratório de Inovação Tecnológica-IFCE

vagner@lit.ifce.edu.br, wendell,rodrigues@ifce.edu.br

rejane.sa@ifce.edu.br

Abstract. *This paper proposes a distributed system using a Raspberry Pi cluster, each equipped with irradiance sensors and connected to inverters in a photovoltaic plant. The system processes data locally using MPI to train LightGBM models for real-time energy prediction and anomaly detection. Additionally, a virtual shading map is generated from irradiance data. This approach minimizes cloud processing, reducing latency and costs, while ensuring quick responses to environmental changes. The system is integrated with a CI/CD pipeline for continuous MLOps operations. Preliminary results demonstrate the solution's scalability and potential to enhance photovoltaic plant reliability.*

Resumo. *Este artigo propõe um sistema distribuído utilizando um cluster de Raspberry Pi, cada um com sensores de irradiância conectados a inversores em uma usina fotovoltaica. O sistema processa dados localmente com MPI para treinar modelos LightGBM, visando previsão de energia e detecção de anomalias em tempo real. Além disso, um mapa virtual de sombreamento é gerado com base nos dados de irradiância. A abordagem minimiza o processamento em nuvem, reduzindo latência e custos, enquanto assegura respostas rápidas às mudanças ambientais. O sistema é integrado a um pipeline CI/CD para operações contínuas de MLOps. Resultados preliminares demonstram a escalabilidade e o potencial da solução para melhorar a confiabilidade das usinas fotovoltaicas.*

1. Introdução

A crescente adoção de usinas fotovoltaicas tem impulsionado a necessidade de soluções tecnológicas avançadas que permitam o monitoramento e a otimização da geração de energia. No entanto, um dos grandes desafios enfrentados nesse contexto é a limitação de infraestruturas físicas e de comunicação [Yu et al. 2011], que dificultam a implementação de sistemas de processamento de alto desempenho. A análise e a previsão da geração de energia, bem como a detecção de anomalias, demandam um poder computacional significativo, que muitas vezes só pode ser alcançado através do uso de serviços de computação em nuvem.

Contudo, o uso da computação em nuvem em usinas fotovoltaicas pode ser inviável devido à necessidade de altas larguras de banda para a transmissão contínua de

grandes volumes de dados. A infraestrutura de comunicação disponível em áreas remotas, onde muitas dessas usinas estão localizadas, frequentemente não oferece suporte para esse tipo de operação. Como resultado, há uma latência significativa e uma redução na capacidade de resposta do sistema, o que pode comprometer a eficiência e a segurança operacional da usina.

Diante desse cenário, torna-se crucial a busca por alternativas que permitam a realização de processamento de alto desempenho diretamente nas usinas, minimizando a dependência de infraestrutura externa e garantindo uma operação mais autônoma e resiliente. Uma solução promissora envolve o uso de clusters de dispositivos de baixo custo, como os Raspberry Pi, organizados em uma arquitetura distribuída com suporte a paralelismo através da Interface de Passagem de Mensagens (MPI)[Ignacio and Martins 2002]. Essa abordagem possibilita o treinamento local de modelos de aprendizado de máquina, como o LightGBM, para a previsão da geração de energia e a detecção de anomalias em tempo real.

Durante o trabalho, foi implementado um sistema distribuído utilizando um cluster de Raspberry Pi para processar localmente os dados de irradiância e de inversores em uma usina fotovoltaica. O sistema visa a previsão e otimização da geração de energia e a geração de um mapa virtual de sombreamento da usina, permitindo uma análise detalhada do impacto das condições ambientais na eficiência dos painéis solares. A solução integrou-se ainda a um pipeline CI/CD, automatizando o ciclo de vida dos modelos de aprendizado de máquina, desde o treinamento até a implantação contínua em produção.

Os resultados preliminares indicam que a abordagem proposta é capaz de superar as limitações de infraestrutura de comunicação e oferecer uma alternativa viável e escalável para o processamento de alto desempenho em usinas fotovoltaicas, contribuindo para a melhoria da eficiência e da segurança operacional desses sistemas.

2. Metodologia

A metodologia adotada busca resolver desafios relacionados ao processamento de alto desempenho, necessários para a previsão em tempo real da geração de energia e detecção de anomalias, superando limitações impostas por infraestruturas de comunicação inadequadas, especialmente em áreas remotas.

A arquitetura proposta, representada na Figura 1, consiste em um cluster MPI (Message Passing Interface) formado por vários dispositivos Raspberry Pi, cada um conectado a sensores de irradiância e a inversores distribuídos pela usina fotovoltaica. O uso do MPI permite a execução de tarefas paralelas e distribuídas, garantindo que o processamento de dados ocorra de forma eficiente e em tempo real. Cada Raspberry Pi no cluster é responsável por coletar dados locais, processá-los e participar do treinamento de modelos de aprendizado de máquina, especificamente utilizando o LightGBM [Ke et al. 2017][de Souza et al. 2023], um algoritmo de boosting de gradiente eficiente e escalável.

O cluster tem capacidade de promover dinamicamente qualquer nó do cluster a mestre, o que assegura a continuidade operacional mesmo em caso de falhas de hardware. A tolerância a falhas é uma característica importante do sistema, permitindo a substituição rápida e econômica de nós defeituosos por novos dispositivos de baixo custo, como os

próprios Raspberry Pi. Essa abordagem garante que o sistema permaneça operacional e confiável, mesmo em grandes projetos de usinas fotovoltaicas onde a continuidade do serviço é crítica.

O processamento de alto desempenho no cluster é parte de várias tarefas, incluindo o treinamento do modelo de aprendizado de máquina, a análise de grandes volumes de dados dos inversores e dos sensores de irradiância, e a geração de um mapa virtual de sombreamento da usina. No processo de treinamento do modelo LightGBM, os dados coletados dos inversores (como corrente e tensão) e dos sensores de irradiância são combinados com o mapa de sombreamento gerado em tempo real, que leva em consideração a cobertura de nuvens no céu. Essa combinação de dados permite que o modelo realize uma regressão mais precisa da produtividade da usina, ajustando-se dinamicamente às condições ambientais variáveis. A criação do mapa de sombreamento é particularmente intensiva em termos computacionais, pois requer a integração de dados coletados em múltiplos pontos distribuídos pela usina. Para garantir que esse processo seja viável e eficiente em um ambiente com recursos limitados, o cluster MPI distribui as tarefas de processamento entre os nós, permitindo que o sistema produza previsões precisas e em tempo real da geração de energia, adaptando-se rapidamente às mudanças nas condições de sombreamento.

Para validar a metodologia proposta, desenvolveu-se um código em Python disponível em um repositório GitHub¹ que emula o ambiente real de uma usina fotovoltaica. Esse código simula a coleta de dados de irradiância e as condições de sombreamento que podem afetar a eficiência dos painéis solares. O sistema implementado distribui as tarefas de processamento entre os nós do cluster, utilizando MPI para coordenar as operações, e integra-se a um pipeline CI/CD para automatizar o ciclo de vida do desenvolvimento e implantação dos modelos de aprendizado de máquina. A validação foi realizada em um ambiente controlado, simulando diferentes cenários de falhas de nós, o que permitiu testar a robustez e a capacidade de recuperação do sistema.

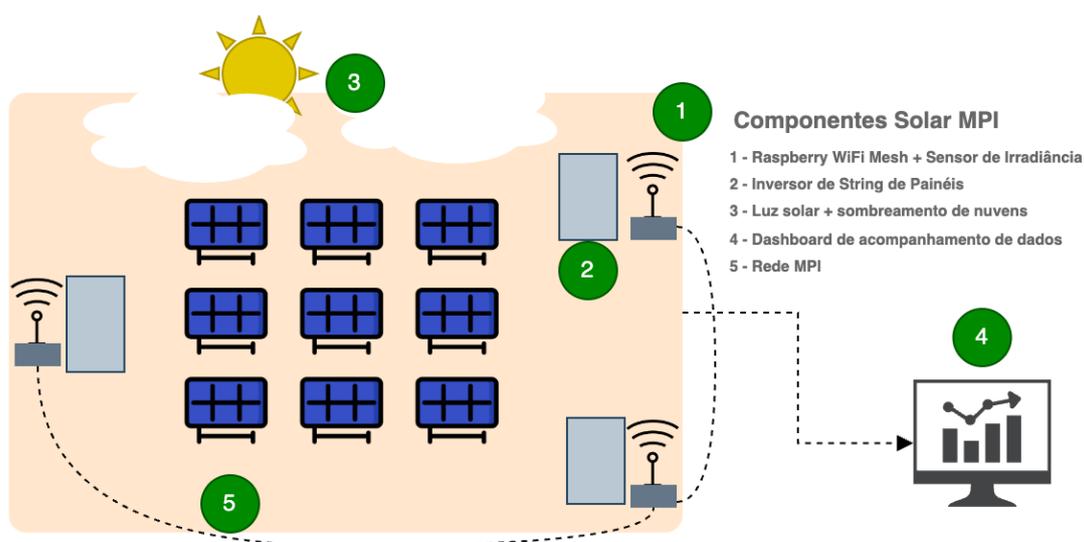


Figure 1. Arquitetura proposta para o cluster MPI em usinas fotovoltaicas.

¹Disponível em: <https://github.com/LIT-IFCE/solarmpi>

3. Resultados

Na simulação, considerou-se um cenário de 100 dispositivos Raspberry Pi distribuídos uniformemente pela área, cada um equipado com sensor de irradiância e conectados a inversores locais. A comunicação entre os dispositivos foi estabelecida através de uma rede WiFi Mesh [Oda et al. 2016], utilizando a tecnologia WiFi 6 para assegurar uma cobertura eficiente e uma comunicação de baixa latência.

O código de emulação foi projetado para simular a coleta de dados de corrente e tensão dos inversores, bem como os valores de irradiância capturados pelos sensores distribuídos na usina. Esses dados foram integrados em tempo real para a geração de um mapa virtual de sombreamento, que leva em consideração a cobertura de nuvens e outros fatores ambientais. A rede WiFi Mesh neste cenário deve permitir que os dispositivos mantenham uma comunicação constante e estável, mesmo em uma área extensa como a de 10 hectares.

O treinamento do modelo LightGBM foi realizado de forma distribuída entre os 100 dispositivos Raspberry Pi, com cada nó processando um subconjunto dos dados coletados em sua região específica da fazenda. Esse treinamento distribuído permitiu uma abordagem eficiente, onde os dados locais de cada nó foram utilizados para ajustar o modelo de predição de produtividade. O tempo médio de treinamento por nó foi de aproximadamente 12 minutos para um conjunto de dados de 8.000 instâncias, com 100 iterações de *boosting*. A validação do modelo, conduzida em paralelo, demonstrou uma precisão robusta, com uma diferença mínima entre as predições individuais dos nós e o modelo global agregado.

A comunicação dentro da rede WiFi Mesh, gerenciada pelo protocolo MPI, foi essencial para a sincronização dos processos de treinamento e para a troca de informações necessárias para a construção do mapa de sombreamento. A latência média na comunicação entre nós foi de aproximadamente 150 milissegundos para pacotes de dados de até 1 MB, o que é aceitável para a aplicação em questão, permitindo uma atualização constante e em tempo real do modelo LightGBM e do mapa de sombreamento.

Além disso, a resiliência do sistema foi testada em cenários de falha de nós dentro da rede WiFi Mesh. Quando um ou mais nós foram desconectados ou apresentaram falhas, o sistema conseguiu redistribuir as tarefas de processamento para os nós restantes sem impacto significativo no tempo total de treinamento ou na precisão das predições. Essa capacidade de auto-recuperação e redistribuição de carga demonstra a viabilidade e robustez da solução em um ambiente de grande escala como o de uma fazenda fotovoltaica.

Os resultados obtidos indicam que o sistema proposto é capaz de realizar o processamento distribuído necessário para predições precisas da produtividade da usina fotovoltaica, mesmo em um ambiente extenso e com recursos computacionais limitados. A integração eficiente entre a coleta de dados, o treinamento distribuído do modelo, e a comunicação em uma rede WiFi Mesh assegura a viabilidade da solução para aplicações reais em grandes fazendas fotovoltaicas, onde a confiabilidade e a precisão são fundamentais para otimizar a operação e a manutenção do sistema de geração de energia.

4. Conclusão e Trabalhos Futuros

Este trabalho apresentou a implementação de um sistema distribuído utilizando um cluster MPI composto por dispositivos Raspberry Pi, voltado para a predição em tempo real da produtividade de uma usina fotovoltaica. Ao integrar dados de sensores de irradiância, informações dos inversores e a geração de um mapa virtual de sombreamento, foi possível demonstrar a viabilidade do uso do algoritmo LightGBM em um ambiente de alto desempenho. A abordagem proposta mostrou-se eficaz, superando as limitações de infraestrutura de comunicação em áreas remotas e garantindo a continuidade operacional mesmo em cenários de falha de nós. Os resultados indicam que o sistema é capaz de executar tarefas complexas de aprendizado de máquina de forma distribuída, mantendo alta precisão e desempenho, o que o torna aplicável em fazendas solares de grande escala. Como trabalho futuro, pretende-se explorar a integração de técnicas de aprendizado profundo para aprimorar ainda mais a precisão das predições, além de investigar o uso de hardware mais avançado para otimizar o desempenho do cluster em cenários com maior complexidade e variabilidade ambiental.

References

- de Souza, V. H., Rodrigues, W., and Lima, M. (2023). Smartgrid improvements on photovoltaic systems by analyzing energy and weather measures. In *Proceedings of the Advance 2023*.
- Ignacio, A. and Martins, V. (2002). Mpi: uma ferramenta para implementação paralela. *Pesquisa Operacional*, 22.
- Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., and Liu, T.-Y. (2017). Lightgbm: A highly efficient gradient boosting decision tree. In *Neural Information Processing Systems*.
- Oda, T., Yamada, M., Obukata, R., Barolli, L., Woungang, I., and Takizawa, M. (2016). Experimental results of a raspberry pi based wireless mesh network testbed considering tcp and los scenario. In *2016 10th International Conference on Complex, Intelligent, and Software Intensive Systems (CISIS)*, pages 175–179.
- Yu, F., Zhang, P., Xiao, W., and Choudhury, P. (2011). Communication systems for grid integration of renewable energy resources. *Network, IEEE*, 25:22 – 29.