

# Classificação de Níveis de Carga de Energia Diária na Região Norte do Brasil Utilizando Aprendizado de Máquina

Lais Sansara S. Silva, Luan Evaristo M. Lindolfo

<sup>1</sup>Faculdade de Computação (Facomp) – Universidade Federal do Pará (UFPA)  
Castanhal – PA – Brasil

**Abstract.** *The growing demand for electricity in Brazil makes it essential to monitor and forecast consumption to avoid system overloads. This work focuses on analyzing the daily average energy consumption in the North region of Brazil, using public data from the National Electric System Operator (ONS). The main objective is to classify days based on their energy consumption level, categorizing them as 'Low', 'Medium', or 'High'. To achieve this, five distinct machine learning algorithms were implemented and compared: Random Forest, Support Vector Machine (SVM), a Multi-Layer Perceptron (MLP) Artificial Neural Network, K-Nearest Neighbors (KNN), and Logistic Regression. The models were trained and evaluated, and the results indicate that the MLP Neural Network achieved the best performance among the tested approaches, with a high recall for the 'High' consumption class, demonstrating the viability of using predictive models to assist in the proactive management of the electrical grid.*

**Keywords:** *Energy Consumption, Machine Learning, Classification, Predictive Analysis, North Region.*

**Resumo.** *A crescente demanda por energia elétrica no Brasil torna essencial o monitoramento e a previsão do consumo para evitar sobrecargas no sistema. Este trabalho foca na análise da média diária de consumo de energia na região Norte do Brasil, utilizando dados públicos do Operador Nacional do Sistema Elétrico (ONS). O objetivo principal é classificar os dias quanto ao seu nível de consumo de energia, categorizando-os em 'Baixa', 'Média' ou 'Alta'. Para isso, foram aplicados e comparados cinco algoritmos de aprendizado de máquina: Floresta Aleatória (Random Forest), Máquinas de Vetores de Suporte (SVM), uma Rede Neural Artificial (Multi-Layer Perceptron), K-Vizinhos Mais Próximos (KNN) e Regressão Logística. Os modelos foram treinados e avaliados, e os resultados indicam que a Rede Neural Artificial obteve o melhor desempenho entre as abordagens testadas, com destaque para a capacidade de identificar dias de alto consumo, conforme demonstrado pelo recall da classe 'Alta', reforçando a viabilidade do uso de modelos preditivos para auxiliar no gerenciamento proativo da rede elétrica.*

**Palavras-chave:** *Consumo de Energia, Aprendizado de Máquina, Classificação, Análise Preditiva, Região Norte.*

## 1. Introdução

O monitoramento e a previsão da demanda por energia elétrica são componentes críticos para a estabilidade da infraestrutura energética de um país. No Brasil, um território de

dimensões continentais, a gestão do sistema elétrico enfrenta desafios complexos, especialmente em regiões com características geográficas e climáticas distintas, como a Região Norte. O aumento do consumo, se não for adequadamente previsto, pode levar a sobrecargas na rede, resultando em interrupções no fornecimento que afetam tanto a população quanto a economia local. É nesse contexto que se torna essencial a aplicação de técnicas de inteligência computacional para otimizar o planejamento e a operação da rede elétrica, mitigando os riscos associados à demanda crescente por energia.

Diante dessa problemática, este trabalho investiga a aplicação de técnicas de aprendizado de máquina como uma ferramenta para a análise de picos de consumo. O estudo utiliza um conjunto de dados público, fornecido pelo Operador Nacional do Sistema Elétrico (ONS), para construir e avaliar modelos de classificação capazes de identificar dias com diferentes níveis de consumo de energia. A análise se concentra no consumo diário (em MWmed) da Região Norte, onde os dias foram rotulados em três classes ('Baixa', 'Média', 'Alta'). O desempenho de cinco algoritmos distintos — Random Forest (Floresta Aleatória), SVM, Rede Neural, KNN e Regressão Logística — é comparado, buscando identificar o modelo mais eficaz para auxiliar no gerenciamento proativo da rede elétrica.

## 2. Trabalhos Relacionados

A aplicação de técnicas de Inteligência Computacional (IC) no setor elétrico brasileiro tem sido amplamente explorada, com foco especial na previsão de demanda e consumo de energia. Trabalhos particularmente relevantes para esta pesquisa são:

[[Barros et al. 2022](#)] desenvolveram um estudo utilizando Redes Neurais Artificiais (MLP) para previsão de curto prazo da carga de energia no Sistema Interligado Nacional (SIN), demonstrando a eficácia dessa abordagem para capturar padrões complexos e não-lineares presentes nas séries temporais de consumo energético. Seus resultados mostraram que modelos neurais superaram abordagens tradicionais, especialmente na identificação de picos de consumo, o que corrobora com os achados do presente trabalho.

Em complemento, [[Silva et al. 2023](#)] aplicaram múltiplos algoritmos de aprendizado de máquina, incluindo Random Forest e SVM, para classificação de níveis de consumo energético na região Nordeste do Brasil. Seu trabalho evidenciou que características temporais como dia da semana e mês são preditores cruciais para a precisão dos modelos, achado que se alinha com os resultados obtidos em nossa análise de importância de *features*.

Além desses estudos específicos, a literatura aponta uma vasta gama de aplicações de IC no setor elétrico. Com a crescente automação e instalação de medidores inteligentes, o volume de dados gerados demanda algoritmos sofisticados para resolver problemas complexos de gerenciamento, otimização e manutenção [[Oliveira et al. 2024](#)]. Nesse contexto, destacam-se aplicações desde a detecção de anomalias em módulos fotovoltaicos até a gestão de redes elétricas inteligentes (*smart grids*), fundamentais para a integração de fontes renováveis e criação de um sistema elétrico mais flexível e eficiente.

A capacidade de prever com acurácia a geração de fontes intermitentes, como solar e eólica, e o consumo de energia são tarefas essenciais para o planejamento energético e estabilidade do sistema. Estudos nacionais frequentemente utilizam modelos de aprendi-

zado de máquina para prever carga e consumo de energia, validando o uso de Redes Neurais Artificiais como ferramenta eficaz para modelar a complexidade e não-linearidades das séries temporais de consumo [Lima et al. 2024].

Este trabalho insere-se neste contexto ao aplicar e comparar modelos de aprendizado de máquina para a tarefa específica de classificação de consumo de energia na Região Norte, utilizando um conjunto de dados ampliado e metodologia rigorosa de parametrização.

### 3. Metodologia

Para este estudo, seguimos um processo estruturado, incluindo coleta de dados, pré-processamento, análise exploratória e modelagem. Todo o código-fonte e scripts utilizados nesta pesquisa estão disponíveis publicamente em um repositório GitHub: [Classificação de Energia](#).

#### 3.1. Base de Dados

Foi utilizado o conjunto de dados "Carga de Energia", disponibilizado pelo ONS, que contém a média diária de consumo de energia em MWmed por subsistema do Brasil, abrangendo um período estendido de janeiro de 2022 a agosto de 2025. Para este trabalho, foi aplicado um filtro para obter apenas os registros referentes à região Norte (identificada como subsistema 'N'), totalizando mais de 1300 registros. Os atributos utilizados foram:

- *id\_subsistema*: Identificador único do subsistema
- *nom\_subsistema*: Nome do subsistema (Norte, Nordeste, etc.)
- *din\_instante*: Data da medição
- *val\_cargaenergiamwmed*: Valor da carga energética em MWmed

#### 3.2. Pré-processamento e Análise Exploratória

As colunas *id\_subsistema*, *nom\_subsistema*, *din\_instante* e *val\_cargaenergiamwmed* foram utilizadas. A coluna *din\_instante*, que representa a data, foi convertida para o formato *datetime*, e a partir dela foram criadas novas variáveis: *dia\_semana*, *mes* e *ano*.

Foi identificado e removido um valor ausente na coluna *val\_cargaenergiamwmed*. Para transformar o problema em uma tarefa de classificação, a variável alvo contínua (*val\_cargaenergiamwmed*) foi discretizada em três classes nominais — 'Baixa', 'Média' e 'Alta' — com base nos quantis (0,33 e 0,66) da distribuição dos dados.

#### 3.3. Configuração Experimental

Conforme as diretrizes do trabalho, foram implementados no mínimo três algoritmos de classificação. Os modelos selecionados para o *benchmarking* final foram Regressão Logística, K-Vizinhos Mais Próximos (KNN), Floresta Aleatória (Random Forest), Máquinas de Vetores de Suporte (SVM) e Rede Neural (MLP).

A **Regressão Logística** é um modelo linear, simples e eficaz para problemas de classificação. O algoritmo **KNN** é um classificador baseado na proximidade, que faz previsões com base nos vizinhos mais próximos da amostra no espaço de características. O **Random Forest** é um método que combina múltiplas árvores de decisão para melhorar

a precisão e controlar overfitting. As **Máquinas de Vetores de Suporte (SVM)** buscam encontrar um hiperplano ótimo que maximize a margem entre classes em um espaço de alta dimensão. Por fim, a **Rede Neural (MLP)** é um modelo não-linear inspirado no funcionamento do cérebro humano, composto por múltiplas camadas de neurônios interconectados. Para cada algoritmo, os seguintes hiperparâmetros foram utilizados:

- **Random Forest:** 40 árvores (n\_estimators=40)
- **SVM:** Kernel RBF com C=2.0
- **MLP:** 4 camadas ocultas com 70 neurônios cada, função de ativação tanh, solver adam
- **KNN:** 5 vizinhos (n\_neighbors=5) com métrica Minkowski (p=3)
- **Regressão Logística:** Parâmetros padrão com random\_state=1

As variáveis categóricas ('nom\_subsistema', 'dia\_semana' e 'mes') e a variável alvo foram transformadas em numéricas utilizando a técnica de *Label Encoding*. Em seguida, os dados das features foram escalonados com 'StandardScaler' para padronizar suas escalas. Para a parametrização dos algoritmos, foi utilizado o método de busca em grade ('GridSearchCV') com validação cruzada, garantindo que os melhores hiperparâmetros fossem selecionados de forma sistemática para otimizar o desempenho dos modelos. Por fim, a base foi dividida em conjuntos de treino (75%) e teste (25%) para avaliação dos modelos.

#### 4. Resultados e Discussão

Os modelos de classificação foram treinados e avaliados utilizando métricas como Acurácia, F1-Score e Recall. A Tabela 1 apresenta o desempenho comparativo dos algoritmos no conjunto de teste.

**Tabela 1. Comparativo de Desempenho dos Modelos.**

Modelo	Acurácia	F1 Score	Recall
Regressão Logística	0.5215	0.5208	0.5215
KNN	0.5512	0.5534	0.5512
Random Forest	0.6552	0.6555	0.6552
SVM	0.5898	0.5855	0.5898
Rede Neural (MLP)	0.6897	0.6923	0.6897

O modelo de Rede Neural (MLP) obteve o melhor desempenho entre as abordagens testadas, com um 'recall' significativamente superior para a classe de consumo 'Alta'. O modelo Random Forest apresentou um desempenho competitivo, enquanto os modelos mais simples, como Regressão Logística e KNN, obtiveram as performances mais baixas. O desempenho superior da Rede Neural e do Random Forest, que são capazes de capturar relações complexas e não-lineares, sugere que o problema de classificação não é linearmente separável.

A matriz de confusão e o 'recall' para a classe 'Alta' demonstraram que o modelo MLP se mostrou eficaz em identificar a maioria dos dias de alto consumo, o que é crucial para o objetivo de prevenir sobrecargas. A análise de importância de features (realizada

com uma Árvore de Decisão auxiliar) indicou que ‘dia\_semana’ e ‘mes’ foram os preditores mais influentes. Este resultado, é considerado moderado, indicando que a capacidade de generalização do modelo é limitada pelas variáveis temporais sozinhas.

## 5. Considerações Finais

Este trabalho demonstrou a aplicação de técnicas de aprendizado de máquina para a classificação de dias com diferentes níveis de consumo de energia na região Norte do Brasil. A utilização de um conjunto de dados ampliado e a parametrização rigorosa dos modelos permitiram a criação de modelos preditivos mais eficazes que podem servir como ferramenta de apoio para o planejamento e a operação do sistema elétrico.

O *Comparativo* entre os algoritmos indicou que a Rede Neural Artificial foi a abordagem mais eficaz para este problema, dentre as testadas. Como trabalhos futuros, sugere-se a inclusão de outras variáveis, como dados meteorológicos e ocorrência de feriados, e a aplicação de modelos mais complexos para séries temporais, como LSTMs, para aprimorar a capacidade preditiva.

## Referências

- Barros, C. A. M., Santos, R. P., and Oliveira, L. F. (2022). Previsão de carga de energia no sin utilizando redes neurais artificiais: Uma abordagem de curto prazo. *Revista Brasileira de Energia*, 28(3):45–62.
- Lima, E. C., Souza, M. A., and Almeida, R. F. (2024). Previsão de séries temporais de consumo de energia utilizando modelos híbridos. In *Congresso Brasileiro de Automação Inteligente*, pages 78–92, Belo Horizonte.
- Oliveira, F. R., Pereira, A. B., and Rodrigues, T. M. (2024). *Smart Grids e Inteligência Artificial: Desafios e Oportunidades para o Setor Elétrico Brasileiro*. Editora Energia Moderna, Rio de Janeiro.
- Silva, M. A., Costa, P. R., and Lima, J. S. (2023). Machine learning para classificação de consumo energético: Um estudo de caso na região nordeste. In *Simpósio Brasileiro de Sistemas Elétricos*, pages 112–125, São Paulo.