

Regressão Linear de Múltiplas Variáveis: Aplicada ao Consumo de Energia

Andrey Huanik A. da S. Ribeiro, Laiane da Silva Sousa, Mateus Herivelton O. da Silva,
Paulo Rodrigo P. da Silva

¹Engenharia de Computação – Universidade Federal do Pará (UFPA)
Caixa Postal – 68746-630 – Castanhal– PA – Brasil

ahuanik2003@gmail.com, laianesilvasousa03@gmail.com, mateusbrdev@gmail.com

Resumo. *Séries temporais desempenham um papel essencial na análise de dados ao longo do tempo, fornecendo insights sobre tendências e padrões que impactam decisões em várias áreas, incluindo energia, economia e saúde. No contexto do consumo de energia, compreender as relações entre variáveis como temperatura, ocupação e uso de dispositivos é vital para otimizar recursos e reduzir custos.*

A regressão linear de múltiplas variáveis surge como uma ferramenta poderosa para modelar essas relações, permitindo a previsão de variáveis-alvo com base em fatores explicativos. Este trabalho explora a aplicação desse modelo no consumo de energia, considerando variáveis como dia, hora, mês, semana, temperatura, ocupação e iluminação.

Abstract. *Time series play an essential role in analyzing data over time, providing insights into trends and patterns that impact decisions in various areas, including energy, economics, and healthcare. In the context of energy consumption, understanding the relationships between variables such as temperature, occupancy, and device usage is vital to optimizing resources and reducing costs. Multivariable linear regression emerges as a powerful tool for modeling these relationships, allowing the prediction of target variables based on explanatory factors. This work explores the application of this model to energy consumption, considering variables such as day, hour, month, week, temperature, occupancy, and lighting.*

1. Fundamentação Teórica

A regressão linear é um modelo matemático que busca descrever a relação entre uma variável dependente e uma ou mais variáveis independentes. Quando há múltiplos fatores explicativos, a regressão é chamada de “múltipla” e assume a forma:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n + \varepsilon \quad (1)$$

Onde:

- β_0 : intercepto;
- β_i : coeficientes que medem a influência de cada variável independente X_i ;
- ε : erro (ou termo residual).

Este modelo é amplamente utilizado devido à sua simplicidade e capacidade de interpretação, sendo adequado para prever comportamentos com base em padrões históricos.

2. Metodologia

2.1. Dados e Variáveis

Foi utilizado um conjunto de dados com as seguintes variáveis:

- **Dia da semana:** categórica, mapeada para valores numéricos (ex.: segunda-feira = 2);
- **Hora:** variável numérica representando o período do dia;
- **Mês:** categórica, indicando o mês do ano;
- **Temperatura:** medida em graus Celsius;
- **Ocupação:** binária, indicando a presença de pessoas no ambiente;
- **Iluminação:** binária, representando se as luzes estavam ligadas ou desligadas.

2.2. Etapas do Processo

1. Pré-Processamento:

- Conversão de variáveis categóricas em numéricas;
- Análise exploratória com mapas de calor e histogramas;
- Exclusão de variáveis irrelevantes.

2. Divisão dos Dados:

- Conjunto de treino (80)
- Conjunto de teste (20)

3. Treinamento do Modelo:

- Regressão linear múltipla aplicada para prever consumo de energia;
- Comparação com o modelo de florestas aleatórias.

2.3. Avaliação

As métricas de desempenho utilizadas foram:

- **Erro Médio Quadrado (MSE):** mede a discrepância entre valores reais e previstos;
- **Coefficiente de Determinação (R^2):** avalia a proporção da variância explicada pelo modelo.

3. Resultados e Discussão

3.1. Visualizações

1. **Distribuição do Consumo de Energia:** O histograma revelou comportamento assimétrico, com picos em períodos específicos;
2. **Relação Entre Temperatura e Consumo:** Maiores consumos foram observados em temperaturas intermediárias;
3. **Padrão por Hora do Dia:** O consumo aumentou nas horas de maior ocupação.

3.2. Avaliação do Modelo

- **Regressão Linear Múltipla:**
 - MSE: 30.45;
 - R^2 : 0.82.
- **Florestas Aleatórias:**
 - R^2 : 0.85, apresentando menor erro absoluto.

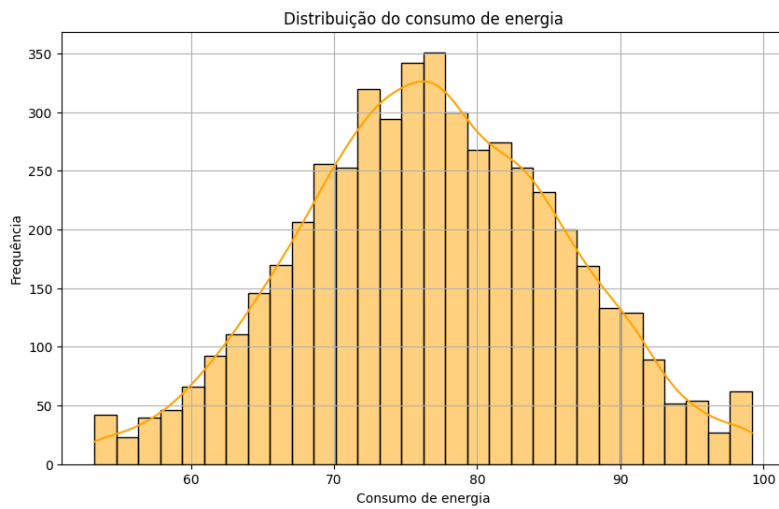


Figura 1. Distribuição do consumo de energia ao longo do tempo.

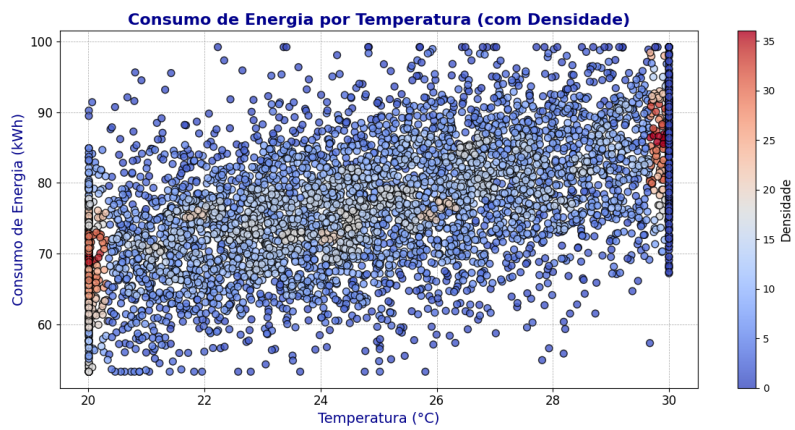


Figura 2. Consumo de energia por temperatura.

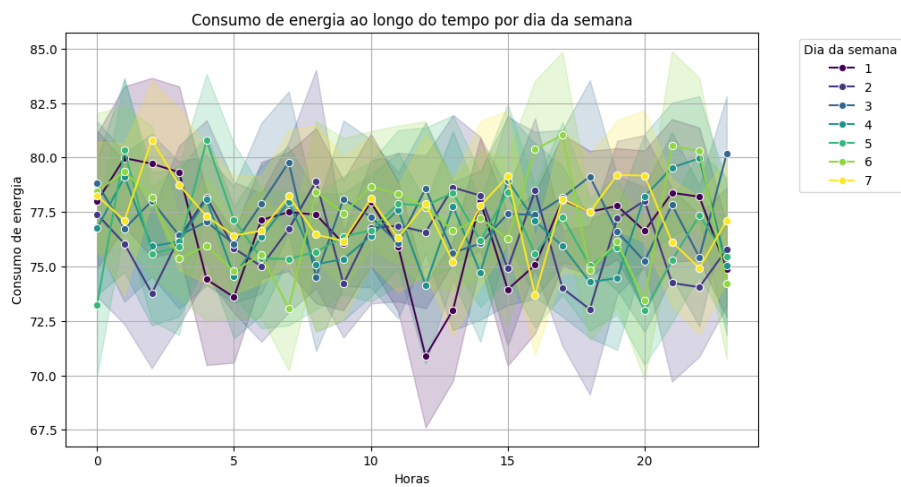


Figura 3. Consumo de energia ao longo do tempo por dia da semana

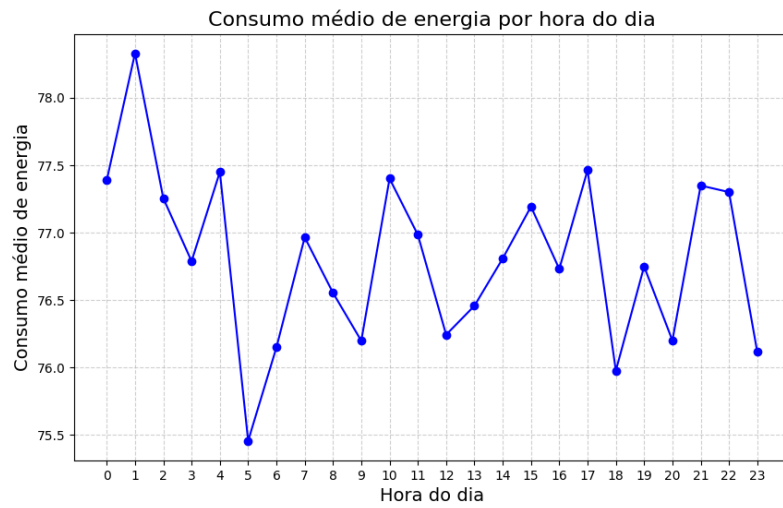


Figura 4. Consumo médio de energia por hora do dia

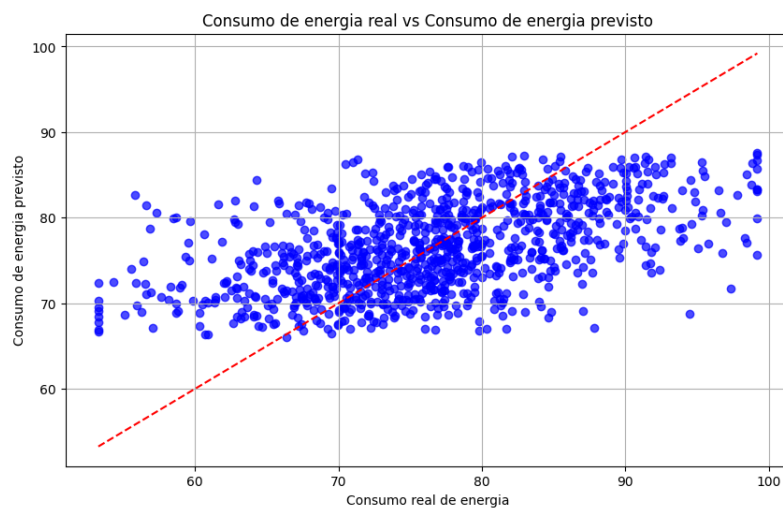


Figura 5. Análise do consumo de energia real vs Consumo de energia previsto.

3.3. Previsão

Uma nova observação ("segunda-feira, 27.73°C, ocupação ativa, luz ligada") resultou em um consumo previsto de 5.82 kWh, reforçando a aplicabilidade do modelo.

4. Conclusão

O uso da regressão linear múltipla para analisar o consumo de energia destacou a relação entre variáveis ambientais e comportamentais, oferecendo insights valiosos para a otimização de recursos energéticos. Embora simples, o modelo apresentou boa capacidade explicativa, com R^2 de 0.82, e de fácil interpretação.

A inclusão de florestas aleatórias melhorou a precisão, alcançando R^2 de 0.85. Futuramente, estudos podem explorar o impacto de mais variáveis ou aplicar técnicas avançadas, como redes neurais, para alcançar maior sofisticação nas previsões.

Referências

[1] B. Damásio, "Regressão linear múltipla: um guia completo!", Blog Psicometria Online, 03-mar-2021. [Online]. Disponível em: <https://www.blog.psicometriaonline.com.br/oque-e-regressao-linear-multipla/>. [Acesso em: 28-jan-2025].