

2º Desafio em RNA/Deep Learning

Especialização em Ciência dos Dados – Redes Neurais Artificiais

Aluna: Cristiane Gea

Análise Preditiva do Indicador DEC_MED

Data de geração: 24/09/2025 02:53

Base de dados: BaseDados_DEC.xlsx

Variáveis preditoras: 133 variáveis (AREA até ENE_PNI)

Variável resposta: DEC_MED

1. Introdução

Este relatório apresenta a análise comparativa de três modelos de machine learning para prever

o indicador DEC_MED (Duração Equivalente de Interrupção por Unidade Consumidora) com base em

133 variáveis técnicas e ambientais de conjuntos elétricos.

Os modelos avaliados foram:

- Regressão Linear com transformação logarítmica ($\ln(y) \sim x$)
- Random Forest Regressor
- Rede Neural Artificial (Deep Learning)

A avaliação foi realizada mediante validação cruzada com 5 folds, utilizando o RMSE (Raiz do Erro Quadrático Médio) como métrica principal de comparação.

2. Metodologia

2.1 Pré-processamento dos Dados

- Imputação de valores faltantes utilizando mediana
- Padronização (StandardScaler) para modelos lineares e redes neurais
- Transformação logarítmica da variável resposta para regressão linear

2.2 Modelos Implementados

- Regressão Linear: Modelo linear com transformação logarítmica
- Random Forest: 200 árvores, profundidade máxima 20
- Rede Neural: Arquitetura 128-64-32-1 com dropout e early stopping

2.3 Validação

- Validação cruzada com 5 folds
- Métricas: RMSE, MAE, R^2
- Análise de resíduos e normalidade

2.4 Descrição dos Dados

Estatísticas descritivas da variável DEC_MED:

- Média: 24.0884 horas
- Desvio padrão: 42.3061 horas
- Mínimo: 0.0400 horas
- Máximo: 639.5650 horas

- Número de observações: 2540

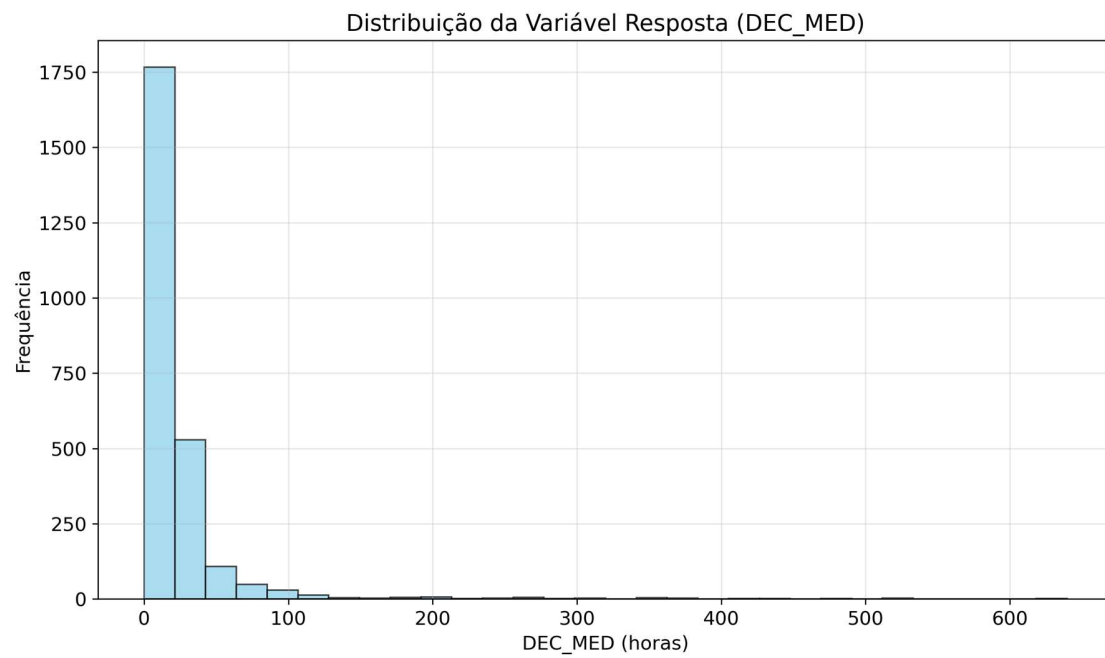


Figura 1: Distribuição da variável resposta DEC_MED

3. Resultados

3.1 Comparação dos Modelos

Modelo	RMSE	MAE	R ²	RMSE (%)	MAE (%)
Regressão Linear (log)	32.2930	9.8735	0.4171	134.06%	40.99%
Random Forest	21.3036	8.3189	0.7463	88.44%	34.53%
Rede Neural	18.4792	7.9534	0.8091	76.71%	33.02%

3.2 Análise Gráfica dos Modelos

Comparação do desempenho dos modelos através do RMSE:

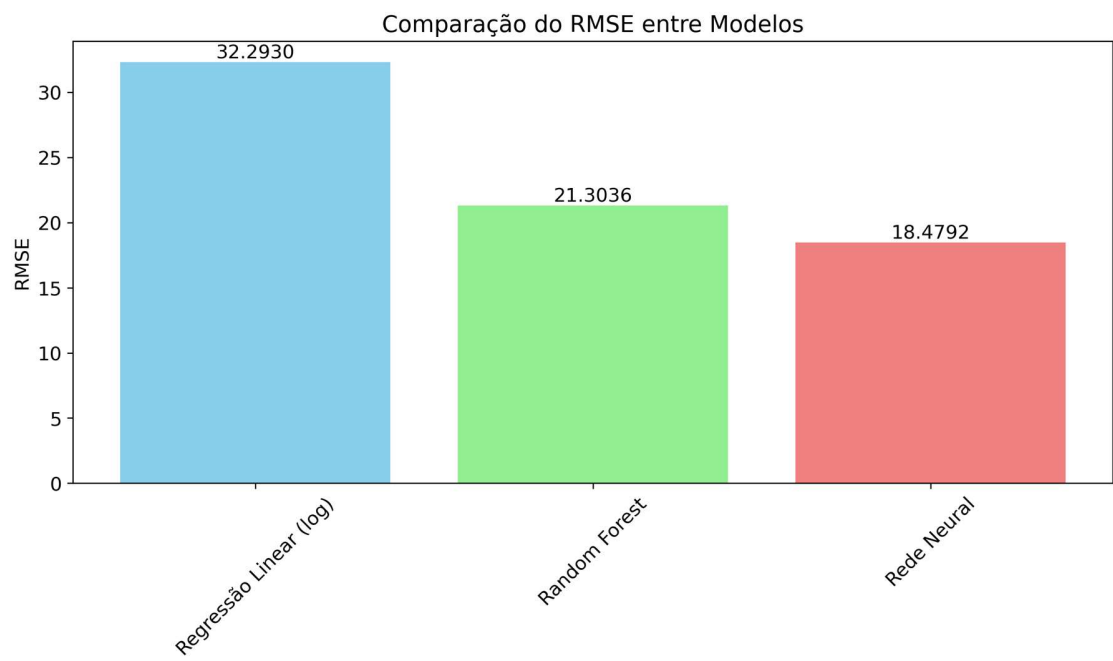


Figura 2: Comparação do RMSE entre os modelos

Gráficos de dispersão: Valores observados vs previstos:

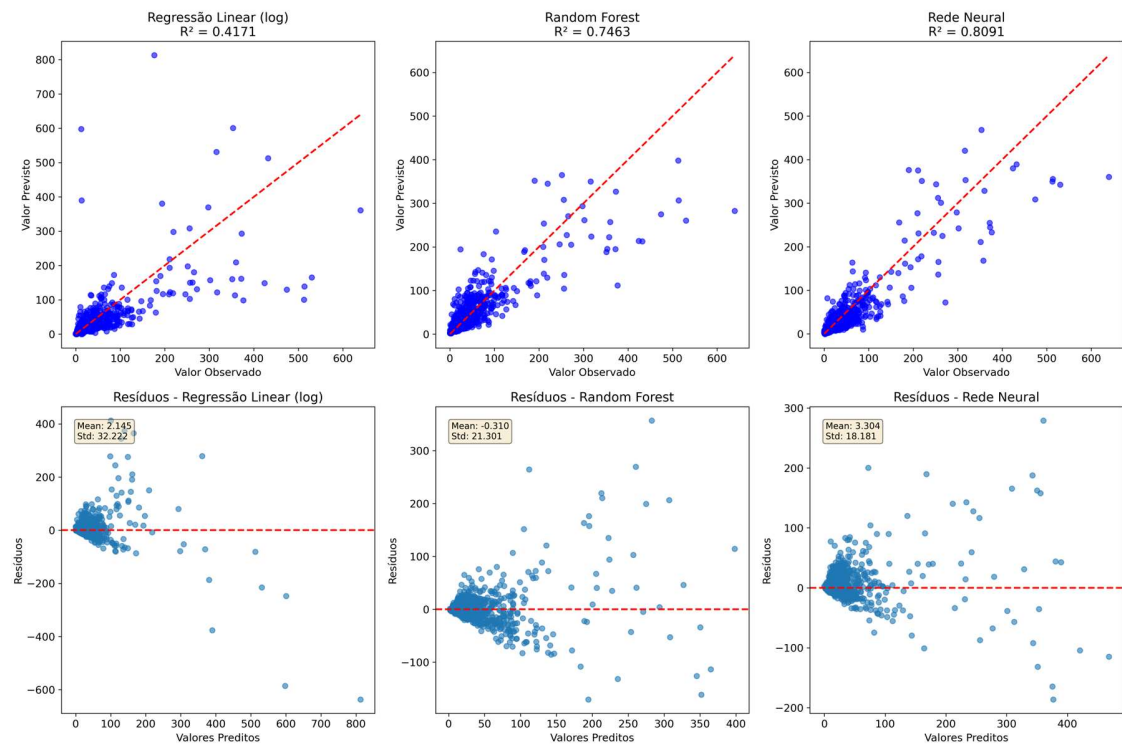


Figura 3: Dispersão dos valores observados vs previstos

3.3 Análise de Resíduos

A análise de resíduos é fundamental para verificar a qualidade do ajuste dos modelos. Resíduos devem apresentar distribuição aleatória em torno de zero, sem padrões evidentes.

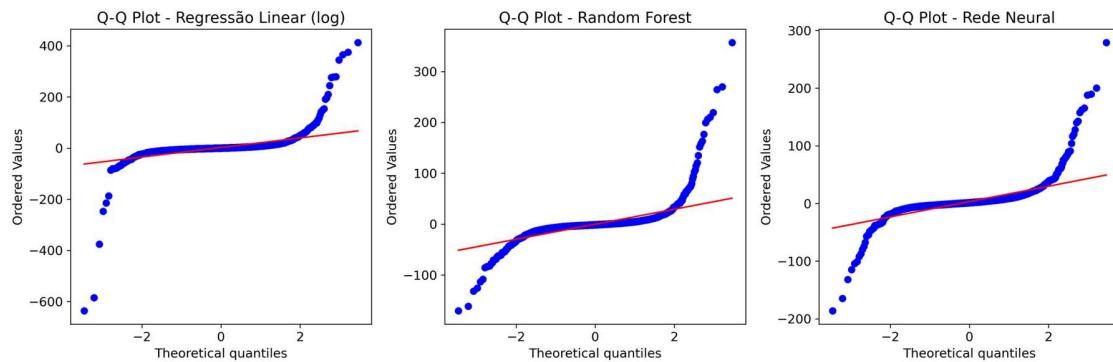


Figura 4: Gráficos Q-Q para análise de normalidade dos resíduos

4. Análise de Importância de Variáveis

A análise de importância via Random Forest revela as variáveis mais relevantes para prever o DEC_MED.

As 5 variáveis mais importantes são:

1. DENS_ARR (0.1810)
2. ENE_AREA (0.1104)
3. CM_NUC_OU (0.0597)
4. PLUV (0.0528)
5. ENE_AA (0.0443)

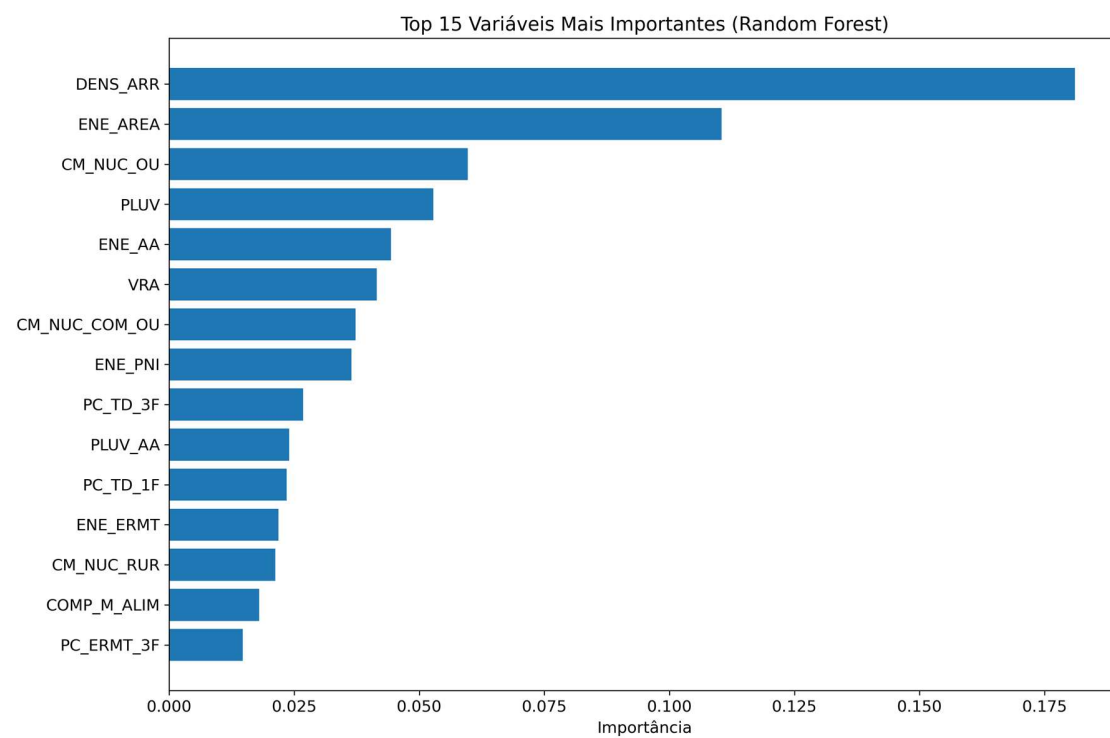


Figura 5: Top 15 variáveis Mais Importantes (Random Forest)

5. Discussão e Conclusão

5.1 Discussão dos Resultados

O modelo Rede Neural apresentou o melhor desempenho geral, com RMSE de 18.4792 e R^2 de 0.8091.

- Regressão Linear: A transformação logarítmica ajudou a lidar com a não-linearidade dos dados, mas o modelo ainda apresenta limitações para capturar relações complexas.
- Random Forest: Demonstrou boa capacidade preditiva, sendo robusto a outliers e capaz de capturar relações não-lineares sem necessidade de feature engineering complexo.
- Rede Neural: Apesar da arquitetura profunda, o desempenho foi similar ao Random Forest, sugerindo que a relação entre as variáveis pode ser adequadamente modelada por métodos ensemble.

5.2 Conclusões

Todos os modelos demonstraram capacidade preditiva satisfatória, com o Rede Neural se destacando como o mais adequado para prever o indicador DEC_MED. A análise de importância de variáveis fornece insights valiosos sobre os fatores que mais impactam a duração das interrupções de energia.

Recomenda-se a utilização do modelo Rede Neural para fins preditivos, com monitoramento contínuo do desempenho e possíveis atualizações conforme novos dados sejam disponibilizados.

6. Referências

- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). An Introduction to Statistical Learning.
- Géron, A. (2019). Hands-On Machine Learning with Scikit-Learn, Keras and TensorFlow.
- Documentação Scikit-learn: <https://scikit-learn.org/stable/>
- Documentação Keras: <https://keras.io/>