

Previsão do consumo de eletricidade em uma instituição de ensino superior: um estudo de caso usando o WEKA

Electricity consumption forecast in a higher education: a case study using WEKA

Yasmin Chaucoski

Universidade Tecnológica Federal do Paraná
chaucoski@alunos.utfpr.edu.br

José Aírton Azevedo dos Santos

Universidade Tecnológica Federal do Paraná
professorjairton@gmail.com

Cidmar Ortiz dos Santos

Universidade Tecnológica Federal do Paraná
cidmar@utfpr.edu.br

Reginaldo Borges

Universidade Tecnológica Federal do Paraná
rborges@utfpr.edu.br

RESUMO

Este trabalho tem como objetivo comparar o desempenho de previsão de dois modelos de séries temporais: a Regressão Linear Múltipla (MLR) e as Redes Neurais Multilayer Perceptron (MLP). Para realizar esta comparação utilizou-se uma base de dados, do consumo de energia elétrica, disponibilizada pela UTFPR - Campus Medianeira. A base de dados apresenta uma série histórica, do consumo de energia, no período de dezembro de 2020 a maio de 2022. Modelos de previsão MLR e MLP foram implementados no software WEKA. Resultados obtidos, dos modelos de previsão, foram comparados por meio das métricas MAPE, MAE e RMSE. Verificou-se, para um horizonte de 4 dias, que, apesar dos efeitos da pandemia sobre o conjunto de dados, a Rede MLP apresentou um bom desempenho.

Palavras-chave: Consumo de energia. MLR. MLP. MAE. RMSE. MAPE.

ABSTRACT

This work aims to compare the forecasting performance of two time series models: Multiple Linear Regression (MLR) and Multilayer Perceptron Neural Networks (MLP). To carry out this comparison, we used a database of electrical energy consumption, made available by UTFPR - Campus Medianeira. The database presents a historical series of energy consumption from December 2020 to May 2022. MLR and MLP forecast models were implemented in the WEKA software. Results obtained from the prediction models were compared using the MAPE, MAE and RMSE metrics. It was verified, for a 4-day horizon, that, despite the effects of the pandemic on the data set, the MLP Network performed well.

Keywords: Energy consumption. MLR. MLP. MAE. RMSE. MAPE.

1. INTRODUÇÃO

A disponibilidade de energia elétrica está diretamente atrelada com a qualidade de vida de uma população, pois proporciona conforto nas atividades do dia a dia. A eletricidade, além de ser indispensável ao desenvolvimento humano, também é parte importante para o desenvolvimento econômico (Becker, 2014).

A previsão do consumo de eletricidade, a curto prazo, é um ótimo instrumento na função de conscientizar os consumidores sobre a economia de energia, porque se baseia no fornecimento de informações que ajudam na busca de padrões de consumo mais eficientes (Amaral, 2020; Kaytez, 2015).

O primeiro passo, para a previsão do consumo de eletricidade, em uma unidade consumidora, é conhecer o perfil de consumo da sua instalação. Os históricos de consumo, das unidades consumidoras, podem ser obtidos da concessionária de energia elétrica. Esses dados históricos são utilizados nos modelos de séries temporais, geralmente com outros indicadores independentes que se considerem influentes. O próximo passo é escolher qual será a metodologia adequada para desenvolver esses modelos (Kaytez, 2015).

Uma série temporal é uma sequência de observações de uma variável ao longo do tempo. Dentre as técnicas, de análise de séries temporais, tem-se a Regressão Linear Múltipla (MLR) e as Redes Neurais Artificiais (RNAs). A Regressão Linear Múltipla é uma técnica de análise de dados que contém uma variável dependente e várias variáveis independentes. Uma Rede Neural Artificial é um modelo, de aprendizado de máquina, que simula o funcionamento do cérebro humano (Haykin, 2001; Pinheiro et al., 2020).

Neste contexto, este trabalho tem como objetivo comparar o desempenho de previsão de dois modelos de séries temporais: a Regressão Linear Múltipla (MLR) e as Redes Neurais *Multilayer Perceptron* (MLP).

2. FUNDAMENTAÇÃO TEÓRICA

Nesta seção apresentam-se o *software* WEKA, os algoritmos de previsão e as métricas utilizadas neste trabalho.

2.1 O WEKA

O Weka é uma coletânea de algoritmos de aprendizado de máquina para funções de mineração de dados. Nele estão contidas várias ferramentas para o processo de mineração de dados, incluindo a preparação dos dados de entrada (Menotti, 2017).

O *software* foi criado na Universidade de Waikato, na Nova Zelândia. Seu nome significa *Waikato Environment for Knowledge Analysis*, ou em português, Ambiente Waikato para Análise de Conhecimento. O sistema é escrito em Java e funciona em várias plataformas (Witten et al., 2016).

Ao iniciar o programa, na versão 3.8.6, é possível escolher entre quatro campos: *Explorer*, *Knowledge Flow*, *Experimenter* e *Workbench*. A interface *Experimenter*, utilizada neste trabalho, auxilia na escolha de quais métodos e valores atuam melhor no problema dado, além de coletar estatísticas de desempenho e realizar testes de significância (Witten et al., 2016).

Segundo Witten et al., (2016) o WEKA pode ser utilizado:

- ✓ Na comparação de desempenho de diversos métodos de aprendizagem;
- ✓ No uso dos modelos de aprendizagem para previsão de novos casos e;
- ✓ Na adoção de métodos de aprendizagem para inspeção das saídas e coleta de informações.

2.2 ALGORITMOS

Os algoritmos utilizados neste trabalho são:

- ✓ Rede Neural *Multilayer Perceptron* (MLP): As Redes MLPs são sistemas compostos por neurônios artificiais. Os neurônios são distribuídos em camadas e ligados por muitas conexões, geralmente associadas a pesos sinápticos, que guardam a informação do modelo e atribuem as entradas aos neurônios (Braga et al., 2000). As redes neurais, mediante um processo de treinamento, são capazes de aprender por meio da experiência (Santos & Chaukoski, 2020; Santos, 2022).
- ✓ Regressão Linear Múltipla (MLR): A técnica de regressão linear múltipla, segundo Krajevski et al. (2009), é uma técnica bastante utilizada para estimar, por meio de um modelo matemático, o comportamento de um sistema. A regressão linear baseia-se na hipótese da existência de uma relação linear entre as variáveis dependentes e independentes (Antoniolli, 2021).

2.3 MÉTRICAS

Os modelos obtidos foram avaliados pelas métricas (Antoniolli, 2021; Cankurt & Subasi, 2015; Machado et al., 2015):

- ✓ RMSE – *Root Mean Squared Error* – A métrica RMSE indica a falha do modelo em estimar a variabilidade em torno da média;
- ✓ MAE - *Mean Absolute Error* – A métrica MAE indica o desvio dos valores previstos a partir dos valores observados;
- ✓ MAPE - *Mean Absolute Percentage Error* – Essa métrica aponta qual a porcentagem do erro sobre os valores reais.

As equações, das métricas MAE, RSME e MAPE, são apresentadas na Tabela 1.

Tabela 1 – Métricas MAE, RMSE e MAPE

Sigla	MAE	RMSE	MAPE
Equação	$\frac{1}{n} \sum_{i=1}^n y_i - \hat{y}_i $	$\sqrt{\frac{1}{n} \sum_{i=1}^n ((y_i - \hat{y}_i)^2)}$	$\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)/y_i \times 100$

Fonte: Antoniolli (2021).

Onde: y_i é o valor real do período i , $\hat{y}_i$ é a previsão para o período i e n é o número de observações.

3. MATERIAIS E MÉTODOS

Nesta seção serão abordados os procedimentos adotados para realização do trabalho.

3.1 A UNIDADE CONSUMIDORA

O presente trabalho foi realizado na Universidade Tecnológica Federal do Paraná - UTFPR, campus Medianeira (Figura 1). Inicialmente intitulada Centro Federal de Educação Tecnológica – CEFET, e instalada no município de Medianeira desde 1987, a instituição passou a ser universidade no ano de 2005, e atualmente oferece cursos de bacharelado, tecnologia e licenciatura (Becker, 2014).

Figura 1 – Campus Medianeira (Vista Superior)



Fonte: Google Maps (2023).

3.2 TIPO DE PESQUISA

Neste estudo, a metodologia utilizada pode ser considerada como (Gil, 2010; Antonioli, 2021):

- ✓ Aplicada: porque tem como objetivo contribuir para fins práticos e busca uma solução para problemas encontrados na realidade;
- ✓ Descritiva: porque descreve as características de uma determinada população, procurando detectar potenciais relações entre as variáveis;
- ✓ Quantitativa: por apresentar dados históricos do consumo de energia da empresa.

3.3 BASE DE DADOS

Os dados utilizados, para realização deste trabalho, foram dados diários de consumo médio de energia, da UTFPR *campus* Medianeira, no decorrer de 18 meses, correspondendo ao período de dezembro de 2020 a maio de 2022.

Na Tabela 2, são apresentados os doze primeiros dados do conjunto utilizado.

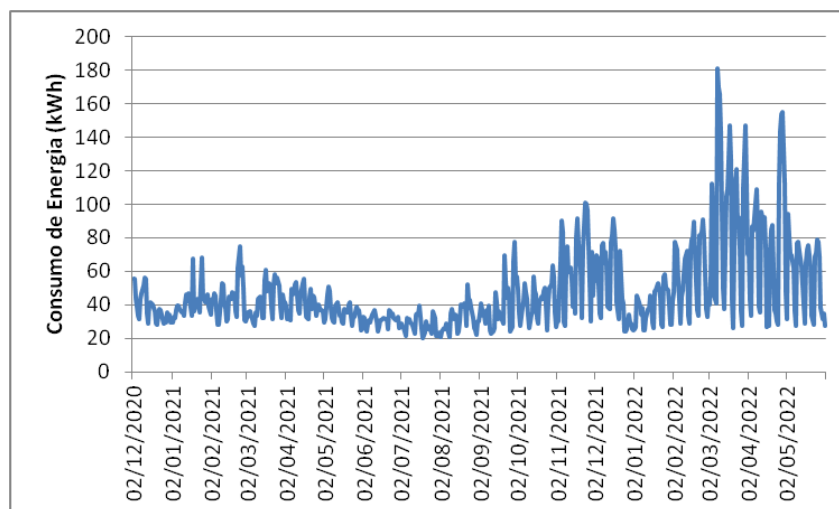
Tabela 2 – Valores iniciais da base de dados

Data	Consumo (kWh)
02/12/2020	55,41
03/12/2020	55,24
04/12/2020	43,37
05/12/2020	39,04
06/12/2020	33,39
07/12/2020	31,34
08/12/2020	43,46
09/12/2020	47,07
10/12/2020	51,58
11/12/2020	56,1
12/12/2020	55,36
13/12/2020	35,06

Fonte: Elaborado pelos autores (2023).

A série histórica, do consumo de energia da UTFPR, é apresentada na Figura 2.

Figura 2 – Série do consumo de energia - UTFPR



Fonte: Elaborado pelos autores (2023).

Na Figura 2, é possível observar uma grande variabilidade nos dados. Isso acontece porque o consumo está relacionado com o acesso de pessoas à Universidade, o que depende muito da época do ano, período, entre outras causas.

3.4 ETAPAS DO TRABALHO

Para comparar os resultados dos modelos, MLR e MLP, foram realizadas, neste trabalho, três etapas (Figura 3):

- ✓ Limpeza e transformação de dados: Inicialmente, na etapa de limpeza, foi executada a limpeza dos dados obtidos da UTFPR. Nesta etapa, dados duplicados e incompletos foram eliminados. Essa é uma parte importante, porque geralmente bons dados trazem bons resultados.

Na sequência, na etapa de transformação dos dados, estes, após processados e limpos, foram exportados para um arquivo no formato aceito pelo WEKA.

- ✓ Treinamento e Validação: Na segunda etapa, Treinamento e Validação, foram selecionados, por meio de métricas, os melhores modelos de Regressão Linear e Redes Neurais.
- ✓ Teste: Na última etapa, etapa de Teste, os modelos foram testados para dados que não participaram da etapa de Treinamento e Validação.

Figura 3 – Etapas do Trabalho



Fonte: Elaborado pelos autores (2023).

4. RESULTADOS E DISCUSSÕES

Inicialmente, neste trabalho, realizou-se uma análise descritiva dos dados (Tabela 3).

Tabela 3 – Análise descritiva de dados – UTFPR

Resumo descritivo	Consumo
Quantidade	543
Média (kWh)	47,27
Mínimo (kWh)	20,34
Máximo (kWh)	180,93
Desvio Padrão (kWh)	24,77
Coefficiente de Variação (%)	52,4

Fonte: Elaborado pelos autores (2023).

Observa-se, dos dados apresentados na Tabela 3, que o consumo ficou, para o período em estudo, com média de 47,27kWh. Apresentando também, neste intervalo de tempo, consumo mínimo e máximo de 20,34kWh e 180,93kWh, respectivamente. Nota-se ainda, da Tabela 3, um coeficiente alto de variação dos dados (52,4%).

4.1 TREINAMENTO E VALIDAÇÃO

Neste trabalho, para obter a melhor modelagem possível do consumo de energia, diferentes arquiteturas e hiperparâmetros foram avaliados. Os algoritmos foram treinados com 377 amostras (70%) e validados com 162 amostras (30%).

Os melhores parâmetros, encontrados para os modelos MLR e MLP são apresentados na Figura 4.

Figura 4 – Parâmetros de configuração dos modelos MLR e MLP

The image shows two side-by-side panels from the Weka software interface. The left panel is for 'weka.classifiers.functions.LinearRegression' and the right panel is for 'weka.classifiers.functions.MultilayerPerceptron'. Both panels have an 'About' section at the top. The Linear Regression panel lists parameters: attributeSelectionMethod (M5 method), batchSize (100), debug (False), doNotCheckCapabilities (False), eliminateColinearAttributes (True), minimal (False), numDecimalPlaces (4), outputAdditionalStats (False), ridge (1.0E-8), and useQRDecomposition (False). The Multilayer Perceptron panel lists parameters: GUI (False), autoBuild (True), batchSize (100), debug (False), decay (False), doNotCheckCapabilities (False), hiddenLayers (8), learningRate (0.01), momentum (0.1), nominalToBinaryFilter (True), normalizeAttributes (True), normalizeNumericClass (True), numDecimalPlaces (2), reset (True), resume (False), seed (0), trainingTime (500), validationSetSize (0), and validationThreshold (20).

Fonte: Elaborado pelos autores (2023).

Os resultados das métricas, para os parâmetros apresentados na Figura 4, são apresentados na Tabela 3.

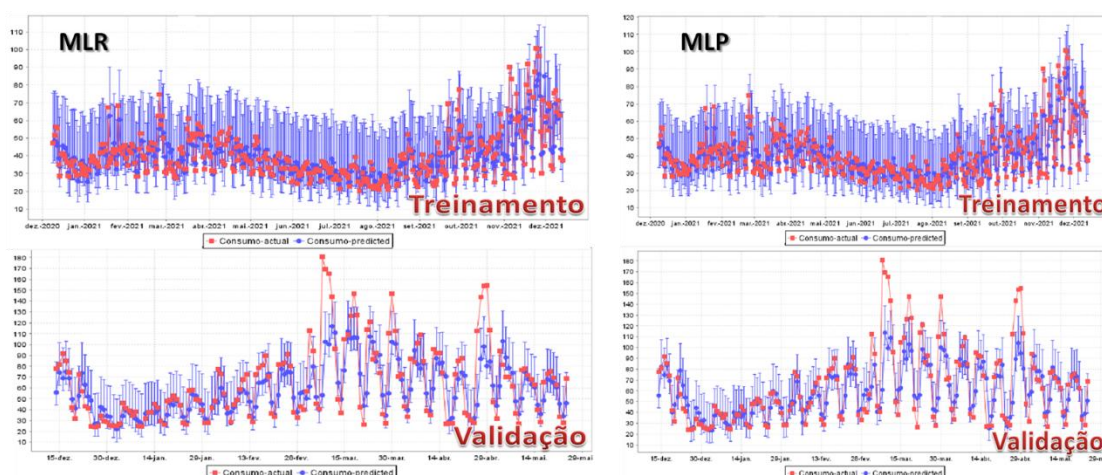
Tabela 3 – Resultados das Métricas – Conjunto de Validação

MÉTRICAS	MLR	MLP
MAE (kWh)	14,66	13,77
MAPE (%)	21,71	20,99
RMSE (kWh)	22,77	22,03

Fonte: Os autores.

Observa-se, dos resultados, apresentados na Tabela 3, que o algoritmo MLP obteve o menor erro para as métricas MAE, MAPE e RMSE. Os resultados obtidos, pelos dois modelos para os conjuntos de Treino e Validação, são apresentados na Figura 5.

Figura 5 – Treinamento e Validação (MLR e MLP)

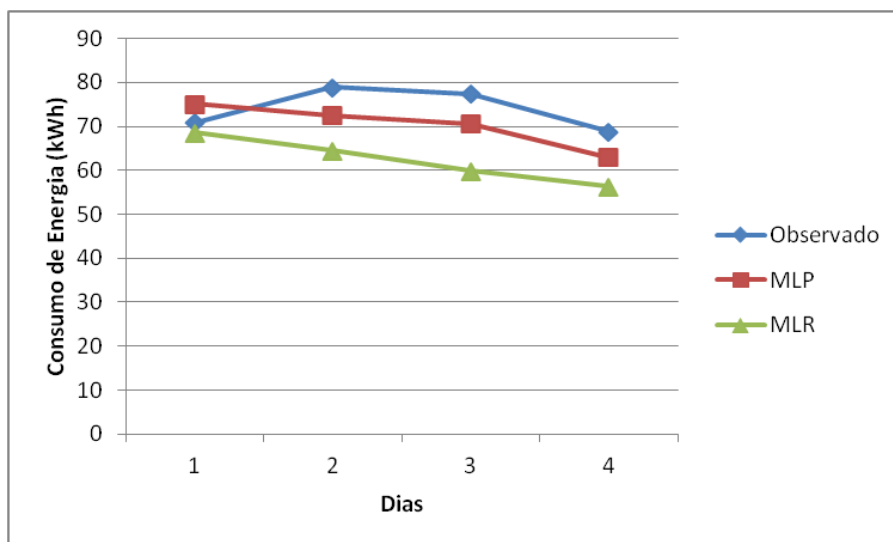


Fonte: Elaborado pelos autores (2023).

4.2 TESTE

Realizou-se, na Etapa de Teste, uma comparação entre os modelos MLR e MLP, para 4 dias que não participaram das Etapas de Treinamento e Validação (Conjunto de Teste: 24/05/2022, 25/05/2022, 26/05/2022 e 27/05/2022). Na Figura 6 apresentam-se, em termos gráficos, os valores observados (reais - UTFPR) e os resultados das previsões, dos modelos MLP e MLR, para os quatro dias do Conjunto de Teste.

Figura 6 – Previsões do conjunto de Teste



Fonte: Elaborado pelos autores (2023).

Na Tabela 4 apresentam-se os resultados das métricas MAE, RSME e MAPE para o Conjunto de Teste.

Tabela 4 – Resultados das Métricas – Conjunto de Teste

MÉTRICAS	MLR	MLP
MAE (kWh)	11,65	5,81
MAPE (%)	15,59	7,83
RMSE (kWh)	13	5,9

Fonte: Elaborado pelos autores (2023).

Por meio dos resultados apresentados, na Tabela 4, concluiu-se que o modelo MLP, com relação às métricas MAE, RMSE e MAPE, obteve, para o Conjunto de Teste, melhor desempenho que o modelo MLR. Pode-se também observar, por meio da Figura 6, que a curva do modelo MLP se ajustou melhor a curva dos dados Observados (reais).

5. CONSIDERAÇÕES FINAIS

Foram comparados, neste trabalho, dois algoritmos de previsão de séries temporais do software WEKA, sendo eles os algoritmos de Regressão Linear Múltipla e Redes Neurais Artificiais do tipo *Multilayer Perceptron*. Todos os dois algoritmos passaram pelas fases de preparação de dados, definição das estruturas, validação, testes e avaliação dos resultados.

Inicialmente, por meio das métricas *Root Mean Squared Error* (RMSE), *Mean Absolute Error* (MAE) e *Mean Absolute Percentage Error* (MAPE), foram selecionados os melhores modelos no Conjunto de Validação. Em seguida, foram estimados os valores para quatro dias do mês de maio de 2022 (Conjunto de Teste). Estes valores foram também confrontados, por meio das métricas, aos valores reais. Observou-se, para um horizonte de 4 dias, que somente o modelo MLP obteve um valor de MAPE menor que 10%, não ficando longe dos valores fornecidos pela Universidade. Portanto, se pôde concluir que o modelo MLP demonstrou uma boa capacidade de generalização para este horizonte de previsão.

Um fator importante a ser observado é a ocorrência da pandemia, que impactou grandemente o consumo de eletricidade e consequentemente na variação dos dados, pois nesse período o acesso à Universidade esteve muito restrito, reduzindo assim também, a utilização de diversos equipamentos responsáveis por manter o consumo nos níveis considerados normais para o dia a dia.

Para trabalhos futuros sugere-se a utilização destes mesmos métodos para os dados seguintes aos aqui manipulados, ou ainda uma comparação dos resultados da rede MLP com algoritmos de aprendizagem profunda, como os algoritmos LSTM, BILSTM, GRU E BIGRU.

REFERÊNCIAS

Amaral, H. L. M. (2020) *Desenvolvimento de uma nova metodologia para previsão de consumo de energia elétrica de curto prazo utilizando redes neurais artificiais e decomposição de séries temporais*. Tese de Doutorado, Universidade de São Paulo, São Paulo.

Antoniolli, E. A. (2021) *Previsão do consumo de energia elétrica em um frigorífico: um estudo de caso utilizando regressão linear, redes neurais e máquinas de vetores de suporte*. Dissertação de Mestrado, Universidade Tecnológica Federal do Paraná, Medianeira, Brasil.

Becker, T. V. (2014) *Otimização da demanda e consumo de energia elétrica da Universidade Tecnológica Federal do Paraná – Campus Medianeira*. Trabalho de Conclusão de Curso, Universidade Tecnológica Federal do Paraná, Medianeira, 2014.

Braga, A. P., Carvalho, A. P. L. F., & Ludermir, T. B. (2000) *Redes Neurais Artificiais: Teoria e Aplicações*. 1. ed. Rio de Janeiro: Editora LTC.

Cankurt, S., & Subasi, A. (2015) Comparasion of linear regression and neural network models forecasting tourist arrivals to turkey. *Eurasian Journal of Science & Engineering*.

Centrais Elétricas Brasileiras. (2005) *Gestão energética*. Rio de Janeiro: Eletrobrás.

Gil, A. C. (2010) *Como Elaborar Projetos de Pesquisa*. São Paulo: Atlas.

Haykin, S. (2001) *Redes neurais: princípios e prática*. 2 ed. Porto Alegre: Bookman.

Krajewski, L., Ritzman, L., & Malhotra, M. (2009) *Administração de produção e operações*. 8. ed. São Paulo: Pearson Prentice Hall.

Kaytez, F. (2015) Forecasting electricity consumption: A comparison of regression analysis, neural networks and least squares support vector machines. *International Journal of Electrical Power Energy Systems*, 67(1).

Machado, n. g., Ventura,b T. M., Danelichen, V. H. M., Querino, C. A., & Biudes, M. S. (2015) Estimation of rainfall by neural network over a neotropical region. *Revista Brasileira de Climatologia*, 11(1).

Menotti, D. (2017) *Introdução ao Weka*. Monografia de Especialização. Universidade Federal do Paraná, Curitiba.

Pinheiro, T. C., Santos, J. A. A., & Pasa, L. A. (2020) *Gestão da produção de frangos de corte por meio de redes neurais artificiais*. *Revista Holos*, 2(1).

Santos, J. A. A. (2022) Previsão do preço do café arábica: uma aplicação de redes neurais CNN-BLSTM. *Research, Society and Development*, 11(1).

Santos, J. A. A., & Chaukoski, Y. (2020) Previsão do Consumo de Energia Elétrica na Região Sudeste: Um Estudo de Caso Usando SARIMA e LSTM. *Revista Cereus*, 12(4).

Witten, I. H., Frank, E., & Hall, M. A. (2016) *Data mining: practical machine learning tools and techniques*. New York: Morgan Kaufmann.