

## Análise comparativa do desempenho de três modelos de previsão de séries temporais do software WEKA

*Comparative analysis of the performance of three time series forecasting models from WEKA software*

Bruno da Rosa Pinto<sup>1</sup>, José Airton Azevedo dos Santos<sup>2</sup>

### RESUMO

Este trabalho tem como objetivo comparar o desempenho de previsão de três modelos de séries temporais: Regressão Linear (LR), Máquinas de Vetores de Suporte (SVM) e Redes Neurais Artificiais (MLP). Para realizar esta comparação utilizou-se uma base de dados, de velocidade do vento, obtida através do Instituto Nacional de Meteorologia (INMET). Esta base apresenta uma série histórica de velocidade do vento, na estação meteorológica de Palmeira dos Índios, no período entre 2008 e 2020. Modelos de previsão, baseados em redes neurais MLP, Máquinas de Vetores de Suporte e Regressão Linear, foram implementados no software WEKA. Resultados obtidos, dos três modelos, foram comparados por meio das métricas RSME (*Root Mean Squared Error*), MAPE (*Mean Absolute Percent Error*) e MAE (*Mean Absolute Error*). Verificou-se, para um horizonte de curto prazo (14 instâncias), que a Rede Neural *Multilayer Perceptron* apresentou o melhor desempenho.

**Palavras-chave:** Modelagem; Previsão; Series temporais; WEKA.

### ABSTRACT

This work aims to compare the forecasting performance of three time series models: Linear Regression (LR), Support Vector Machines (SVM) and Artificial Neural Networks (MLP). To carry out this comparison, a wind speed database was used, obtained from the National Institute of Meteorology (INMET). This database presents a historical series of wind speed, at the Palmeira dos Índios meteorological station, in the period between 2008 and 2020. Forecast models, based on MLP neural networks, Support Vector Machines and Linear Regression, were implemented in the WEKA software. Results obtained from the three models were compared using the metrics RSME (*Root Mean Squared Error*), MAPE (*Mean Absolute Percent Error*) and MAE (*Mean Absolute Error*). It was found, for a short-term horizon (14 instances), that the Multilayer Perceptron Neural Network presented the best performance.

**Keywords:** Modeling; Forecasts; Time series; WEKA.

<sup>1</sup> Discente. Departamento de Engenharia Elétrica (DAELE - UTFPR). E-mail: bruno.r\_pinto@hotmail.com. ORCID: <https://orcid.org/0009-0007-9674-7019>

<sup>2</sup> Doutor em Engenharia Elétrica. Docente do Programa de Pós-Graduação em Tecnologias Computacionais para o Agronegócio (PPGTCA). Universidade Tecnológica Federal do Paraná (UTFPR). E-mail: airton@utfpr.edu.br. ORCID: <https://orcid.org/0000-0002-2568-5734>

## 1. INTRODUÇÃO

As séries temporais, amplamente utilizadas por todas as áreas do conhecimento, são ferramentas úteis para fazer inferências sobre o futuro e apoiar decisões. São utilizadas, por exemplo, na caracterização de processos evolutivos, na propagação de doenças, no comportamento de índices de preços, etc (SICSÙ, 2012). Séries temporais podem ser definidas como um conjunto de dados coletados ao longo de um período de tempo, geralmente em intervalos iguais. Segundo Cunha (2019), é possível, por meio de modelagem matemática, extrair informações destes dados, supondo a continuidade do comportamento apresentado anteriormente.

Dentre as técnicas de previsão de séries temporais tem-se a Regressão Linear, as Máquinas de Vetores de Suporte e as Redes Neurais artificiais. Regressão linear multivariada é uma técnica que gera uma equação que descreve uma relação estatística entre as variáveis preditoras e a variável de resposta. Máquina de vetores de suporte é uma técnica de previsão, de séries temporais, baseada na teoria do aprendizado estatístico e a Rede Neural Artificial é uma técnica baseada na estrutura biológica do cérebro humano (ANTONIOLLI et al., 2022; HAYKIN, 2001).

Diversos trabalhos de previsão, presentes na literatura, utilizaram modelos de séries temporais do software WEKA. Dentre eles podem-se citar os trabalhos de: Santos e Spancerski (2021) que utilizaram Máquinas de Vetores de Suporte e Redes Neurais Artificiais para previsão do preço do açúcar. Rustick et al. (2023) que utilizaram Redes Neurais MLP para previsão da produção, importação e consumo de fertilizantes no Brasil. Antonioli et al. (2022) que utilizaram Regressão Linear, Máquinas de Vetores de Suporte e Redes Neurais Artificiais para previsão do consumo de energia elétrica em um frigorífico, localizado no estado do Paraná. Já Pimentel et al. (2023) utilizaram Regressão Linear e Redes Neurais Artificiais para previsão da geração de energia hidráulica no Brasil.

Neste contexto, este trabalho tem como objetivo comparar o desempenho de previsão de três modelos de séries temporais: a Regressão Linear (LR), as Máquinas de Vetores de Suporte (SVM) e as Redes Neurais Artificiais (MLP). De maneira a verificar qual, dos três modelos, apresenta melhor adequação aos dados e, portanto, apresenta previsões mais precisa sobre a situação futura.

Este trabalho está organizado da seguinte maneira. Na seção Materiais e Métodos são descritas as metodologias empregadas na previsão da velocidade do vento. Na Seção

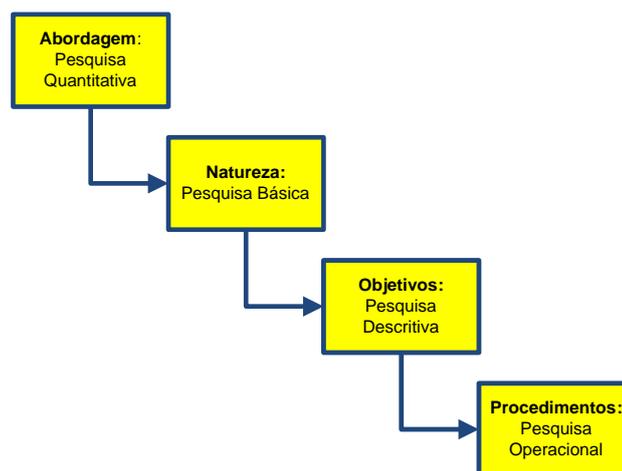
Resultados e Discussão são apresentados os resultados obtidos da aplicação dos três modelos do software WEKA. Comentários finais e conclusões finalizam o trabalho.

## 2. MATERIAIS E MÉTODOS

Nesta seção aborda-se a metodologia utilizada para realizar essa pesquisa, detalhando os métodos propostos para realização dos experimentos, bem como as técnicas utilizadas.

Tipo de Pesquisa:

Os procedimentos metodológicos, desenvolvidos nesta pesquisa, são apresentados no fluxograma da Figura 1 (GIL, 2010).



**Figura 1.** Procedimentos Metodológicos

Onde:

- ✓ Pesquisa Quantitativa: caracteriza-se pelo uso de dados numéricos. Nesta abordagem aplicam-se também técnicas estatísticas como a média, desvio padrão, coeficiente de variação, etc;
- ✓ Pesquisa Básica: tem como propósito analisar conhecimentos úteis para o avanço da ciência e da tecnologia;
- ✓ Pesquisa Descritiva: pode ser definida como aquela que descreve uma realidade de forma imparcial, sem interferências de quem está pesquisando.
- ✓ Pesquisa Operacional: busca solução para um determinado problema por meio de modelos matemáticos ou estatístico.

## Base de Dados:

Inicialmente, para previsão da velocidade do vento, obteve-se da estação meteorológica A327 (Figura 2), localizada no município de Palmeira dos Índios no estado do Alagoas, uma base de dados com 156 instâncias (Jan/2008 - Dez/2020) (INMET, 2022).



**Figura 2.** Estação Meteorológica de Palmeira dos Índios  
**Fonte:** Pereira et al. (2017).

A estação de Palmeira dos Índios possui uma torre com 100m de altura, com anemômetros em 03 níveis (30m, 70m e 100m) e sensores de direção de vento (30m e 70m). As medidas utilizadas, na composição da base de dados, são geradas a partir da média das medidas dos sensores anemométricos no período de um mês.

Os algoritmos de previsão têm um melhor desempenho em um conjunto de dados com um número grande de instâncias. Segundo Silva (2014) métodos numéricos como a interpolação de dados, os ajustes de dados e a transformada de Fourier podem ser utilizados para incrementar o número de instâncias de uma base de dados existente. Neste trabalho, para aumentar o número de instâncias, expandiu-se o conjunto de dados para valores diários, por meio de interpolação linear. Inicialmente, o conjunto de dados possuía 156 instâncias, sendo que após a interpolação passou para 4749 instâncias (Conjunto de Dados INMET\_INT). Observa-se que esta operação empresta a este conjunto de dados caráter teórico, mesmo advindo de dados reais.

## Algoritmos:

Os algoritmos, do software WEKA, utilizados neste trabalho, são:

- ✓ Regressão Linear (LR): A técnica de Regressão Linear é utilizada em várias áreas como economia, finanças, engenharia, entre outras. É um método preditivo que tem

como objetivo linearizar o comportamento de um conjunto de dados. Segundo Favero (2015) o objetivo da análise de regressão é propiciar ao pesquisador condições de avaliar como se comporta uma variável  $Y$  com base no comportamento de uma ou mais variáveis  $X$ .

- ✓ Rede Neural *Multilayer Perceptron* (MLP): As redes neurais MLPs são sistemas que possuem a capacidade de adquirir, armazenar e utilizar conhecimento experimental. São formadas por unidades de processamento simples denominadas de neurônios. Os neurônios, da rede MLP, são dispostos em camadas e interligados por conexões. As conexões são associadas a pesos. As redes MLPs são constituídas por uma camada de entrada, uma ou mais camadas ocultas e uma camada de saída. O treinamento da rede é realizado pelo algoritmo *Backpropagation*. Este algoritmo utiliza pares de entrada e saída para ajustar, por meio de correção de erros, os pesos da rede (HAYKIN, 2001; ANTONIOLLI et al., 2022).
- ✓ Máquina de Vetor de Suporte (SVM): As SVMs, assim como as redes MLPs, podem ser utilizadas em diferentes áreas, tais como reconhecimentos de padrão, processamento de imagens, bioinformática, entre outras. As SVMs pode ser aplicadas em problemas de classificação e regressão. Para problemas de regressão são denominadas de SVR (*Support Vector Regression*). O algoritmo SMOReg, do software WEKA, implementa a máquina de vetor de suporte para regressão (SVR) (SPANCERSKI; SANTOS, 2021).

#### Etapas do Trabalho:

Três etapas foram realizadas, neste trabalho, para formar os modelos de Regressão Linear (LR), Redes Neurais *Multilayer Perceptron* (MLP) e de Máquina de Vetor de Suporte Otimizada para Regressão (SMOReg):

1. Análise de dados: Inicialmente, na primeira etapa, realizou-se um pré-processamento dos dados obtidos do INMET, visando adequá-los aos algoritmos. Isso se fez por meio da eliminação da incompletude dos dados e da aplicação da interpolação. Na sequência realizou-se uma análise exploratória destes dados;
2. Treinamento e Validação: Na segunda etapa, treinamento e validação, foi selecionado, por meio de métricas, os melhores modelos LR, MLP e SMOReg;
3. Teste: Na última etapa, etapa de teste, os modelos foram testados para dados que não participaram da etapa de treinamento e validação.

## Recurso de Software:

Neste trabalho, como recurso de software, foi utilizado o WEKA. Este software é gratuito e muito utilizado no meio acadêmico. O WEKA trabalha com arquivos no formato ARFF (*Attribute-Relation File Format*). O ARFF é um arquivo texto contendo um conjunto de observações, precedido por um pequeno cabeçalho (Figura 3). O cabeçalho é utilizado para fornecer informações a respeito dos campos que compõem o conjunto de observações. Neste cabeçalho os atributos são declarados, podem ser *numeric*, *date* ou *string*. Na segunda parte do arquivo se localizam os dados, que são estruturados de acordo com as declarações da primeira seção e são separados por vírgula (SOUZA, 2021).

```
@relation 'Vento'  
@attribute Data date "yyyy-MM-dd"  
@attribute Velocidade numeric  
  
@data  
  
2008-01-01,3.446476  
2008-01-02,3.438435  
2008-01-03,3.430395  
2008-01-04,3.422354  
2008-01-05,3.414313  
2008-01-06,3.406272  
2008-01-07,3.398232  
2008-01-08,3.390191  
2008-01-09,3.38215  
2008-01-10,3.374109
```

**Figura 3.** Arquivo ARFF

## Métricas:

Neste trabalho, os modelos obtidos foram avaliados pelas métricas Erro Médio Absoluto Percentual (MAPE), Raiz Quadrada do Erro Médio Quadrático (RMSE) e Erro Médio Absoluto (MAE). O RMSE é a raiz do erro médio quadrático da diferença entre a predição e o valor real, o MAE representa o desvio médio entre os valores observados e preditos pelo modelo. Já o MAPE mede o tamanho do erro em termos percentuais (CANKURT; SUBASI, 2015; PINHEIRO, 2020).

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

**Tabela 1.** Equações MAE, RMSE e MAPE.

Sigla	Equação
-------	---------

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

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

### 3. RESULTADOS E DISCUSSÃO

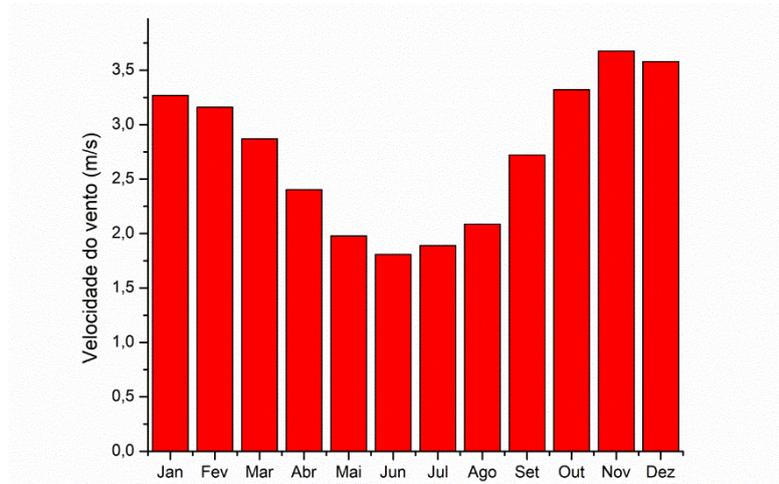
Inicialmente, neste trabalho, realizou-se uma análise descritiva dos dados do conjunto INMET\_INT (Tabela 2).

**Tabela 2.** Análise descritiva.

Parâmetro analisado	Resultado
Quantidade	4749
Média (m/s)	2,392
Mínimo (m/s)	0,844
Máximo (m/s)	4,269
Desvio Padrão (m/s)	0,788
Coefficiente de Variação (%)	32,9

Pode-se observar, dos dados apresentados na Tabela 2, que a velocidade ficou, para o período em estudo, em média de 2,392 m/s. Apresentando, neste período, velocidade mínima e máxima de 0,844 m/s e 4,269 m/s. Nota-se também, da tabela, um coeficiente de variação dos dados de 32,9% considerado alto de acordo com Pimentel (2009). Observa-se que quanto maior o coeficiente de variação maior a variação nos dados.

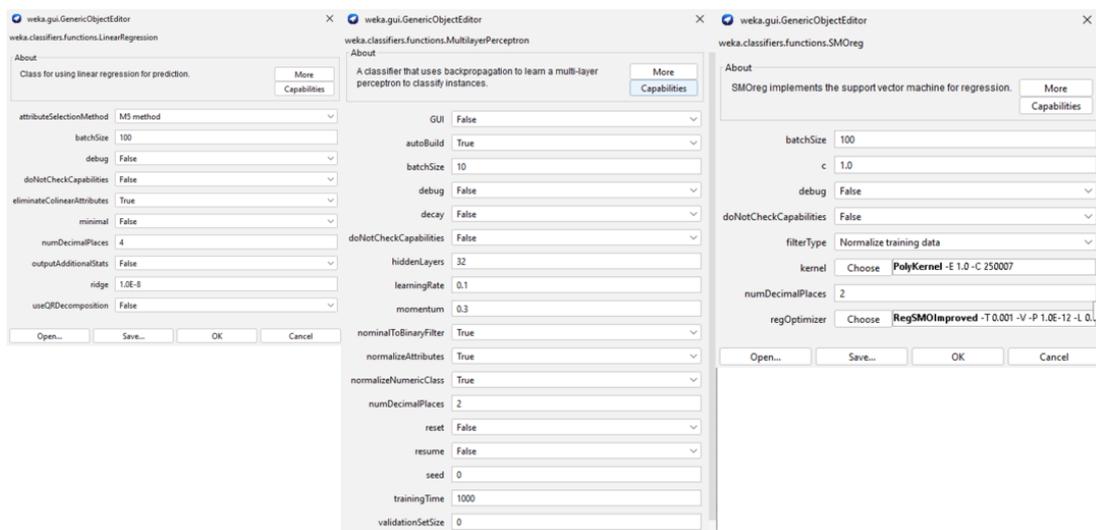
A média mensal da velocidade do vento, durante o período estudado, é apresentada na Figura 4. Pode-se notar, por meio desta figura, que os meses, com menor velocidade, ocorrem entre abril e setembro. Observou-se também uma velocidade média mínima no mês junho e uma velocidade média máxima para o mês de novembro.



**Figura 4.** Média mensal da velocidade do vento

### Treinamento e Validação:

Vários modelos, dos três algoritmos, foram treinados por meio do software WEKA. Os melhores parâmetros, encontrados para os modelos LR, MLP e SMOReg, são apresentados na Figura 5. Os modelos foram treinados com 3551 amostras (75%) e validados com 1184 amostras (25%).



**Figura 5.** Parâmetros de configuração dos modelos

Os resultados das métricas, para os parâmetros apresentados na Figura 5, são apontados na Tabela 3.

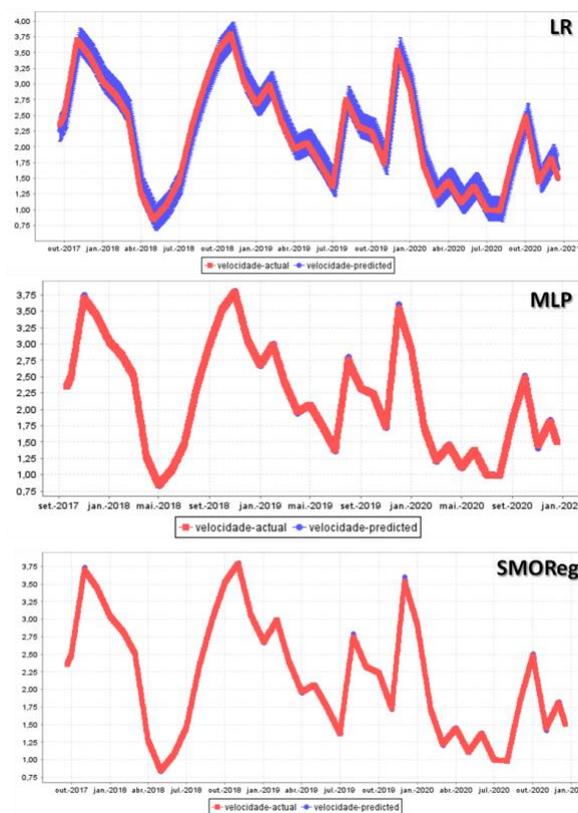
**Tabela 3.** Resultados das Métricas – Conjunto de Validação.

MÉTRICAS	LR	MLP	SMOReg
----------	----	-----	--------

<b>MAE (m/s)</b>	0,1209	0,0053	0,0064
<b>MAPE (%)</b>	5,981	0,2628	0,1549
<b>RMSE (m/s)</b>	0,1481	0,0094	0,069

Observa-se, dos resultados, apresentados na Tabela 3, que o algoritmo SMOReg obteve o menor erro para a métrica MAPE e o algoritmo MLP obteve os menores erros para as métricas MAE e RMSE.

A Figura 6 apresenta os resultados obtidos pelos três modelos, para os conjuntos de validação, considerando um intervalo de confiança de 95%.



**Figura 6.** Velocidade (m/s) – Conjunto de validação (SMOReg)

Teste:

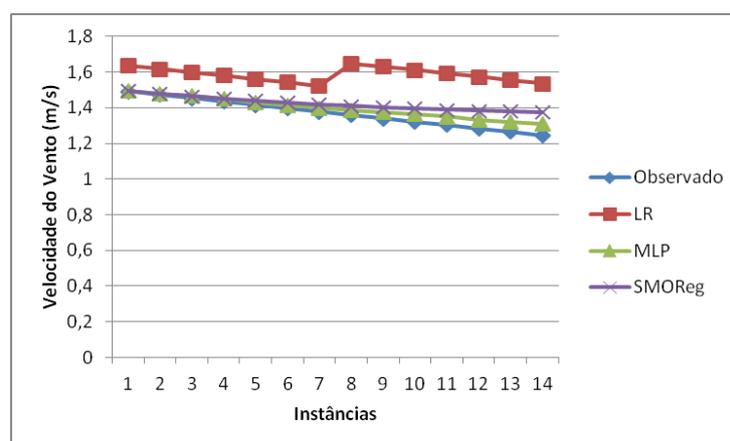
Na Tabela 4 apresentam-se os dados, observados, preditos, os Erros Relativos Percentuais (ERP) e o MAPE, para as 14 instâncias que não participaram da etapa de treino e validação.

**Tabela 4.** Previsões da velocidade (m/s), ERP (%) e MAPE (%).

Observado	LR	ERP_LR	MLP	ERP_MLP	SMOReg	ERP_SVM
1,491819	1,6367	9,7117	1,4954	0,24	1,4945	0,1797
1,472881	1,6179	9,8459	1,4802	0,4969	1,4789	0,4087
1,453942	1,5991	9,9838	1,466	0,8293	1,4648	0,7468
1,435003	1,5803	10,1252	1,4511	1,1217	1,4521	1,1914
1,416065	1,5614	10,2633	1,4296	0,9558	1,4405	1,7256
1,397126	1,5425	10,4052	1,4149	1,2722	1,4295	2,3172
1,378188	1,5236	10,551	1,4001	1,5899	1,4194	2,9903
1,359249	1,6491	21,3243	1,387	2,0416	1,4104	3,7632
1,34031	1,6304	21,6435	1,3745	2,5509	1,4022	4,6176
1,321372	1,6116	21,9641	1,3626	3,1201	1,395	5,5721
1,302433	1,5929	22,3019	1,3505	3,6906	1,3889	6,6389
1,283495	1,5741	22,6417	1,3329	3,8493	1,3837	7,8072
1,264556	1,5554	22,9997	1,3209	4,4556	1,3788	9,0343
1,245618	1,5364	23,3444	1,309	5,0884	1,3746	10,3549
<b>MAPE = 16,2218%</b>		<b>MAPE = 2,2359%</b>		<b>MAPE = 4,0963%</b>		

Por meio dos resultados apresentados, na Tabela 4, conclui-se que os resultados das previsões, dos três modelos, estão próximos aos fornecidos pelo INMET\_INT. Contudo, o modelo MLP apresenta um MAPE menor que os modelos LR e SMOReg (MAPE\_LR= 16,22%, MAPE\_MLP= 2,24% e MAPE\_SMOReg=4,1%).

Os resultados das previsões, do conjunto de teste, em termos gráficos, são apresentados na Figura 7.



**Figura 7.** Velocidade (m/s) – Conjunto de teste

Pode-se observar, do gráfico apresentado na Figura 7, a diferença de comportamento dos três modelos, em relação aos dados do conjunto INMET\_INT (Observado), em especial a discrepância da regressão linear. Destaca-se que o modelo SMOReg esteve levemente

mais próximo, do comportamento dos dados observados, nas 4 primeiras previsões, contudo, nas previsões seguintes o erro aumentou significativamente quando comparado ao modelo MLP.

#### 4. CONSIDERAÇÕES FINAIS

Neste trabalho, resultados de modelos de séries temporais, de arquitetura LR, MLP e SMOReg, foram comparados. Os modelos foram construídos com base na interpolação dos dados, de velocidade do vento, fornecidos pelo Instituto Nacional de Meteorologia (INMET) no período de análise de 2008 à 2020, totalizando um conjunto de 4735 dados.

A comparação de modelos de previsão para velocidade do vento é complexa pois os resultados são influenciados por muitos fatores, como o volume de dados para o teste, conjunto de entrada disponível, amplitude da variação da velocidade do vento, configuração exata do algoritmo, etc. Portanto, assumir que um modelo é melhor que o outro é um processo desafiador, devido ao número de situações a serem levadas em considerações.

No entanto, observou-se, neste trabalho, que os modelos MLP e SMOReg apresentaram resultados relevantes para um horizonte de curto prazo (14 instâncias). Porém, o modelo MLP apresentou menor MAPE que o modelo SMOReg. As previsões do modelo MLP, para o conjunto de teste, foram bem precisas e as diferenças entre valores observados e preditos foram pequenas. Portanto, a proximidade entre valores preditos e observados demonstram a boa capacidade de generalização, para um horizonte de 14 instâncias, do modelo MLP.

Embora, os algoritmos MLP e SMOReg tenham apresentado bons resultados, sugere-se, para trabalhos futuros: uma comparação com redes neurais híbridas, a busca por bancos de dados maiores e com maior precisão na leitura, uma vez que a falta e a imprecisão dos dados dificultam a análise e levam a resultados equivocados e incoerentes, e, por último, a utilização de outros softwares e linguagens como MATLAB e Python.

#### REFERÊNCIAS

ANTONIOLLI, E. A. **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, 2021.

CANKURT, S.; SUBASI, A. Comparasion of linear regression and neural network models forecasting tourist arrivals to turkey. **Eurasian Journal of Science & Engineering**, 2015.

CUNHA, M. M. **Previsão de séries temporais por meio de métodos estatísticos e neurais: aplicação a uma indústria de bebidas.** Trabalho de Conclusão de Curso. Universidade Tecnológica Federal do Paraná, Medianeira, 2019.

FÁVERO, L. P. **Análise de Dados.** Rio de Janeiro: Elsevier, 2015. Disponível em: <https://integrada.minhabiblioteca.com.br/#/books/9788595153226/>. Acesso em: 22 mai. 2022.

GIL, A. C. **Como Elaborar Projetos de Pesquisa.** São Paulo: Atlas, 2010.

HAYKIN, S. **Neural networks: a comprehensive foundation.** New Delhi: Pearson Prentice Hall, 2001.

INMET. **Estação: Palmeira dos Índios A327.** Disponível em: <https://tempo.inmet.gov.br/TabelaEstacoes/A327>. Acesso em: 20 mai. 2022.

PEREIRA, L. T.; PERCEUS, M.; PEREIRA, E. T. Modelagem matemática e previsão da velocidade do vento em Palmeira dos índios/AL com a utilização do modelo Boz & Jenkins. *Revista de Extensão da Uneal*, v. 2, 2017.

PIMENTEL, C. C. et al. Previsão de geração de energia hidráulica no Brasil: um estudo de caso usando redes neurais artificiais e regressão linear. **Caderno Pedagógico**, 2023.

PIMENTEL, F. **Curso de estatística experimental.** 15. ed. Piracicaba: Fealq, 2009.

PINHEIRO, T. C., SANTOS, J. A. A., PASA, L. A. Gestão da produção de frangos de corte por meio de redes neurais artificiais. **Revista Holos**, 2020.

RUSTICK, A.; SANTOS, J. A. A.; SANTOS, C. O. Aplicação de modelos de redes neurais artificiais na previsão da produção, importação e consumo de fertilizantes no Brasil. *Peer Review*, v. 5, 2023.

SANTOS, J. A. A.; ANTONIOLI, E. A.; GUEDES, A. L.; PASA, L. A. Previsão do consumo de energia elétrica em um frigorífico: um estudo de caso usando regressão linear, redes neurais e máquinas de vetor de suporte. **Tekne e logos**, v. 13, 2022.

SICSÚ, A. L.; DANA, S. **Estatística Aplicada: Análise Exploratória De Dados.** São Paulo: Editora Saraiva, 2012. Disponível em: <https://integrada.minhabiblioteca.com.br/#/books/9788502177574/>. Acesso em: 19 mai. 2022.

SILVA, S. R. **Modelagem de Oxigênio Dissolvido utilizando Redes Neurais Artificiais.** Dissertação de Mestrado. Instituto Federal de Goiás. Goiânia, 2014.

SOUZA, B. C. S. **Mineração de dados aplicada a previsão do preço de ações utilizando o WEKA.** Trabalho de Conclusão de Curso. Pontifícia Universidade Católica de Goiás. Goiânia, 2021.

SPANCERSKI, J. S.; SANTOS, J. A. A. Previsão do preço do açúcar: um estudo de caso utilizando o software WEKA. **Revista Engenharia e Tecnologia**, v. 13, 2021.