

Hybrid System for Forecasting Meteorological Time Series in Penedo, Alagoas

Elmo Araújo Filho¹, José Lucas Bispo dos Santos¹, Marília G. F. de M. Oliveira²
Augusto C. F. de M. Oliveira³, Gustavo H. F. de M. Oliveira¹

¹ Sistemas de Informação, Universidade Federal de Alagoas (UFAL)¹.

²Universidade Federal Rural de Pernambuco (UFRPE)².

³Engenharia de Software, Universidade de Pernambuco (UPE)³.

{elmo.filho, jose.santos7}@arapiraca.ufal.br, marilia.gfmo@gmail.com
augusto.oliveira@upe.br, gustavo.oliveira@penedo.ufal.br

Abstract. *Weather forecasts are important for various sectors of society, impacting the adoption of strategies to improve agriculture and the management of natural disasters. This study focuses on this context, specifically on forecasting climate time series in the city of Penedo, Alagoas, due to the complex characteristics of its climate. As a method, a hybrid system was proposed that combines linear and non-linear forecasting models. The results demonstrated that the system achieved the best performance in six out of nine databases tested, standing out as a robust solution for climate prediction.*

Resumo. *As previsões meteorológicas são importantes para vários setores da sociedade, impactando na adoção de estratégias para melhorar a agricultura e na gestão de desastres naturais. Este estudo se concentra nesse contexto, especificamente na previsão de séries temporais climáticas na cidade de Penedo, Alagoas, devido às características complexas do seu clima. Como método, foi proposto um Sistema Híbrido que combina modelos lineares e não-lineares de previsão. Os resultados demonstraram que o sistema obteve o melhor desempenho em seis de nove bases de dados testadas, destacando-se como uma solução robusta para a previsão climática.*

1. Introdução

Uma série temporal climática, segundo [Morettin and Toloí 1985], pode ser definida como “qualquer conjunto de observações de variáveis ordenadas no tempo”, incluindo temperatura, precipitação, pressão atmosférica, umidade, velocidade e direção do vento, entre outras. Na literatura, diversos estudos utilizam essas variáveis para prever ocorrências climáticas, como secas severas, fortes chuvas ou alterações na normalidade do clima [BARROS et al. 2012].

A análise dessas variáveis permite previsões meteorológicas extremamente benéficas para múltiplos setores da sociedade, incluindo o planejamento agrícola, a previsão de possíveis inundações ou deslizamentos de terra [Soares et al. 2022]. À medida que essas previsões se tornam mais precisas, sua influência nas atividades diárias e no planejamento estratégico para o futuro se torna cada vez mais importante.

No entanto, prever séries temporais climáticas ainda é um desafio, especialmente em regiões com características climáticas complexas. O Brasil, com sua vasta extensão territorial e diversidade ambiental, agrícola e climática, exemplifica bem essas dificuldades [BARROS et al. 2012].

O estado de Alagoas, na região Nordeste do país, apresenta desafios específicos devido a “irregularidades na precipitação pluviométrica e pouca variação sazonal na radiação solar, fotoperíodo e temperatura do ar”, o que complica a previsão meteorológica pelo poder público [BARROS et al. 2012]. Além disso, a escassez de dados climáticos disponíveis, especialmente nas cidades do interior de Alagoas, agrava esses desafios, pois muitas localidades não possuem estações meteorológicas para coleta de dados.

Uma exceção é a cidade de Penedo, onde há uma estação meteorológica localizada no Instituto Federal de Alagoas (IFAL) e gerida em colaboração com a Secretaria de Meio Ambiente (SEMARH). Nessa estação, dados meteorológicos são continuamente observados, registrados e organizados, formando conjuntos de dados que podem ser utilizados para gerar informações valiosas [Freitas 2023].

Com esses dados, torna-se possível explorar a utilização de sistemas de previsão meteorológica baseados em técnicas de aprendizado de máquina. No entanto, é importante ressaltar que não existe um modelo individual de aprendizado de máquina que garanta desempenho ótimo em todos os conjuntos de dados [de Mattos Neto et al. 2020]. Por esse motivo, autores como [de Oliveira and Ludermir 2014], [Faruk 2010] e [de Mattos Neto et al. 2020] recomendam o uso de sistemas híbridos de previsão, que combinam diferentes modelos individuais para capturar um maior número de padrões nos dados.

Baseado nisso, este artigo relata os resultados de um projeto de Iniciação Científica cujo objetivo foi desenvolver um Sistema Híbrido (SH) para previsão de séries temporais climáticas em Penedo, Alagoas. O SH combina modelos lineares e não-lineares de forma integrada, onde os modelos lineares buscam identificar relações lineares entre os dados, enquanto os modelos não-lineares capturam relações mais complexas. A hipótese é que essa combinação possa fornecer previsões mais precisas do que modelos individuais, gerando contribuições significativas para a previsão de dados climáticos.

Para alcançar esse objetivo, este trabalho descreve, na seção 2, a fundamentação teórica sobre a previsão de séries temporais e a definição de sistemas híbridos. Na seção 3, apresentamos os métodos utilizados. Na seção 4, discutimos os experimentos e resultados obtidos. Finalmente, na seção 5, enumeramos as principais conclusões deste estudo.

2. Fundamentação Teórica

2.1. Previsão de Séries Temporais

Com o avanço das tecnologias e com o desenvolvimento de novos sensores, as possibilidades de monitoração do clima têm melhorado consideravelmente ao longo dos anos. Autores como [Bochenek and Ustrnul 2022] salientam que esses avanços têm possibilitado o estudo e investigação de abordagens importantes, como o uso de aprendizado de máquina. Ainda, destacaram que nos últimos anos o número de trabalhos que utilizaram aprendizado de máquina para análises de mudanças climáticas aumentaram significativamente, gerando sistemas de previsões cada vez mais precisos.

A eficiência desses algoritmos de aprendizagem de máquina depende do ajuste correto dos dados para que eles sejam capazes de capturar as correlações e padrões existentes. Por exemplo, uma série temporal, segundo [de Oliveira and Ludermir 2014], é formalmente representada por um vetor de dados $X_t = \{x_1, x_2, x_3, \dots, x_t\}$ que precisa ser transformado em um par de dados (\mathbf{X}, y) . Nesse par, \mathbf{X} representa a matriz dos padrões de entrada e y os alvos que os modelos precisam aprender a prever. Esse processamento pode ser feito utilizando a técnica de janela de tempo ou janela deslizante.

A janela de tempo divide a série temporal em vários segmentos menores através da expressão $Z_t = \{X_{t-w+1}, X_{t-w+2}, \dots, X_t\}$. Em que Z_t representa uma janela de entradas para o instante de tempo t e w o tamanho dessa janela. A exemplo, padrões de entrada como $X_1 = \{x_1, x_2, \dots, x_w\}$ são gerados e podem ser usados para prever o alvo x_{w+1} . Essa metodologia quando aplicada ao longo de uma série temporal, cria uma matriz dos padrões de entrada \mathbf{X} e o vetor de saída y , permitindo a aplicação de qualquer algoritmo de previsão. Com isso, as previsões geradas, podem embasar tomadas de decisões nos mais diversos setores de uma sociedade.

2.2. Sistemas Híbridos de Previsão

No mundo real, reconhecer os padrões de uma série temporal pode ser uma tarefa complexa, uma vez que os dados são estocásticos (ou seja, possuem um comportamento aleatório e imprevisível) e podem se alterar ao longo do tempo [Figueiredo Filho and Silva Júnior 2009]. Alguns autores, como [Zhang 2003], discutem que os padrões das séries temporais podem ser compostos por dois tipos: lineares e não-lineares. Esse tipo de formação dificulta a previsão para modelos de aprendizado de máquina individuais, pois eles podem não ser capazes de capturar todas as variações dos padrões existentes de uma só vez.

Por esse motivo, os sistemas híbridos são mais adequados, visto que eles fazem uso da combinação de mais de um modelo de previsão, comumente, um modelo linear e outro não-linear [de Oliveira and Ludermir 2014]. Exemplos de modelos lineares são os métodos estatísticos como AR, ARMA e ARIMA, que encontram a relação entre os dados de entrada e saída de forma linear. Já os modelos não-lineares são aqueles que conseguem modelar padrões mais complexos, como curvaturas, a exemplo algoritmos de aprendizagem de máquina como as redes neurais artificiais, SVR, árvore de decisão, entre outros.

Esta característica motivou diversas pesquisas sobre o uso de sistemas híbridos na literatura, como os trabalhos de [Aly 2020], [de Oliveira et al. 2021] e [Han et al. 2022]. Em [Aly 2020] foi proposto a combinação de redes neurais artificiais e filtro de Kalman para previsão da velocidade de vento. [de Oliveira et al. 2021] investigou a possibilidade de selecionar o modelo mais indicado para cada padrão da série temporal, utilizando os algoritmos MLP, SVR, RBF, LSTM e CNN. Por fim, em [Han et al. 2022] analisou o uso da regressão linear e da rede neural BP para prever chuvas de curta duração.

Embora os estudos na literatura destaquem o potencial dos sistemas híbridos para previsão, são raros os trabalhos focados na previsão climática de regiões específicas, como o que está sendo investigado neste estudo. Dessa forma, o desenvolvimento de um sistema de previsão para a cidade de Penedo, Alagoas, não apenas pode oferecer segurança no planejamento de atividades como o plantio, mas também na prevenção de impactos de

chuvas severas e outros eventos climáticos.

3. Metodologia

Nesta seção, será apresentado o Sistema Híbrido de previsão de séries temporais climáticas para a cidade de Penedo, Alagoas. O método proposto visa combinar modelos lineares e não-lineares para aprimorar a precisão das previsões climáticas. As etapas do método proposto são ilustradas em duas etapas apresentadas na Figura 1.

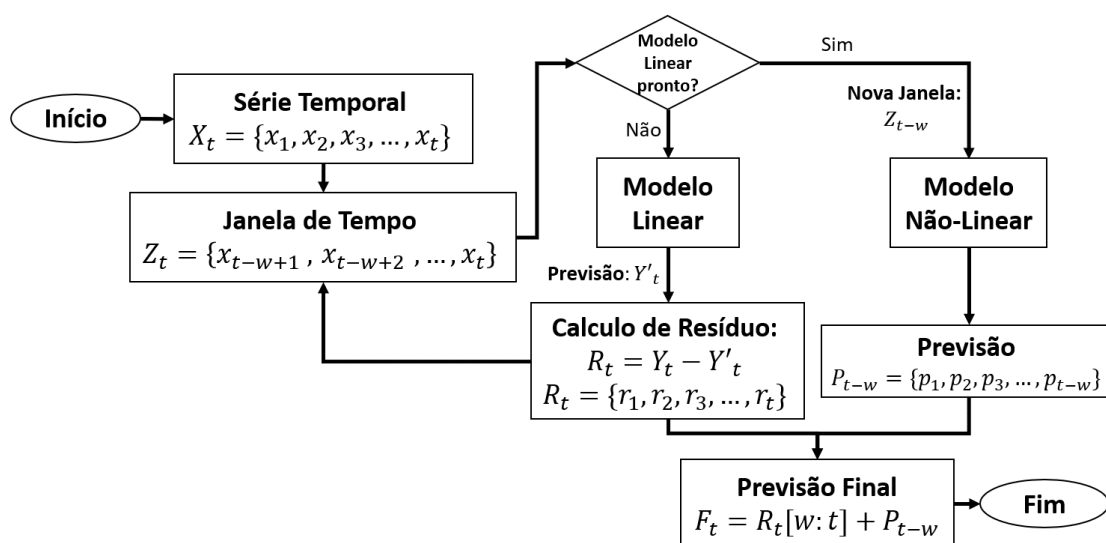


Figura 1. Estrutura geral do Sistema Híbrido

O processo se inicia com o recebimento da série temporal com os dados climáticos, representado pela variável $X_t = \{x_1, x_2, x_3, \dots, x_t\}$. Em seguida, a série temporal é submetida à aplicação da janela de tempo Z_t , conforme discutido na seção 2.1. Essa etapa é crucial para transformar a série temporal no formato (\mathbf{X}, y) , adequado para que o modelo de previsão consiga aprender as correlações dos dados.

Após isso, o Sistema Híbrido inicia o treinamento de um modelo linear para fazer a previsão inicial. [Figueiredo Filho and Silva Júnior 2009] afirmam que um modelo linear é um tipo de modelo matemático ou estatístico que pressupõe uma relação linear entre a variável de entrada \mathbf{X} e a variável de saída y . Exemplos de modelos lineares são a regressão linear, lasso, lars-lasso e outros. Estes modelos são rápidos e eficazes na captura de padrões simples nos dados, isso se torna uma escolha ideal para a primeira fase de treinamento. A previsão Y'_t gerada pelo modelo linear é utilizada para realizar o cálculo dos resíduos da próxima etapa.

Os resíduos, denotados por R_t são componentes essenciais na construção e melhoria do Sistema Híbrido, eles representam a diferença entre os valores reais observados Y_t e os valores previstos Y'_t gerados pelo modelo linear. Matematicamente definidos como: $R_t = Y_t - Y'_t$. Esses resíduos são os padrões que o modelo linear não conseguiu capturar. Como os resíduos são um vetor numérico, eles precisam ser novamente submetidos a técnica de janela de tempo $Z_{t-w} = \{r_{t-w+1}, r_{t-w+2}, \dots, r_t\}$ para que o próximo modelo consiga interpretar adequadamente os dados.

Após a transformação, o modelo não-linear é treinado para capturar os padrões restantes nos resíduos. Os modelos não-lineares são mais flexíveis e podem modelar uma gama mais complexa de padrões nos dados. Diversos modelos podem ser usados nesta etapa, como a Arvore de decisão, MLP e SVR. Então, logo após o seu treinamento a previsão do modelo não-linear pode ser calculada e representada por: $P_{(t-w)} = \{p_1, p_2, p_3, \dots, p_{(t-w)}\}$

Com ambos os modelos treinados, a previsão final F_t é obtida somando a previsão não-linear $P_{(t-w)}$ e a previsão do modelo linear $R_t[w : t]$. Devido a aplicação da janela de tempo nos resíduos, a previsão do modelo não-linear tem tamanho $t - w$, em que t é o tamanho original da série temporal e w é o tamanho da janela de tempo aplicada. Dessa forma, para garantir que a soma seja feita nos instantes de tempos adequados, usa-se $R_t[w : t]$, que significa que a soma é feita em R_t a partir da posição w até a posição final t .

4. Experimentos

Nesta seção serão descritos as bases utilizadas e o tratamento aplicado, métricas de avaliações e configuração dos experimentos.

4.1. Bases Utilizadas

As séries temporais climáticas utilizadas nos experimentos foram obtidas na estação meteorológica da cidade de Penedo, localizada no IFAL (Campus Penedo). O conjunto de dados climáticos possui 34 séries temporais abrangendo um quantitativo de 201.076 registros, compreendendo o período de 10 de maio de 2023 a 29 de janeiro de 2024. Neste estudo, apenas 9 séries temporais foram selecionadas para a experimentação, como demonstrado na Figura 2. Cada série temporal apresenta, aproximadamente, 5.914 registros, proporcionando um conjunto de dados robusto para a análise climática.

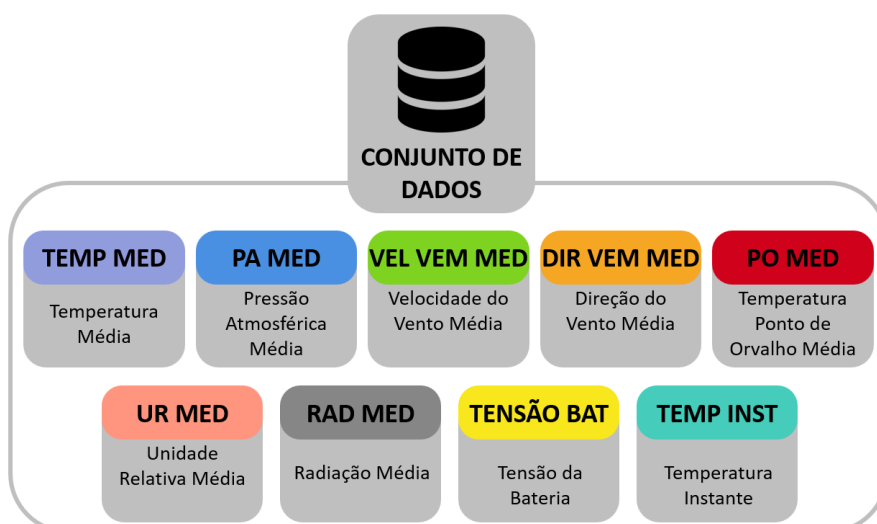


Figura 2. Variáveis coletadas e armazenadas na base da estação meteorológica

4.2. Tratamento das Bases

As séries temporais foram separadas em conjuntos de treinamento e teste para a realização dos experimentos. A separação foi dividindo os dados em duas partes temporais. Foram

atribuídos 80% do período inicial para treinar o modelo, e os 20% restantes para testar a capacidade de generalização dos modelos em dados diferentes dos treinados. Essa estratégia é chamada de holdout, que é amplamente utilizada em séries temporais, uma vez que mantém a ordem temporal dos dados e evita o vazamento de informações do futuro para o passado.

Para garantir que todas as variáveis fossem comparáveis e que os modelos avaliados conseguissem convergir durante o treinamento, aplicou-se a normalização Min-Max aos dados. A normalização, indicada por [SILVA 2018] redimensiona a série temporal para um intervalo específico [0, 1], sem alterar o formato original da série. O cálculo é representado pela seguinte expressão: $X_{norm} = (X - X_{min}) / (X_{max} - X_{min})$. Onde X_{norm} representa o valor resultante da série, X representa o primeiro valor da série, enquanto X_{min} seria o valor mínimo e X_{max} o valor máximo.

4.3. Métricas de Avaliação

Para avaliar a qualidade das previsões, a métrica utilizada foi o Erro Quadrático Médio (MSE) por ser uma medida simples e que mensura a distância entre as previsões e os valores reais. Isso significa que quanto mais próximo de zero melhor se torna a precisão do algoritmo escolhido. O MSE é representado pela expressão abaixo:

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (1)$$

Onde y_i representa os valores reais, \hat{y}_i representa os valores previstos e n o número total de observações.

Para comparar os resultados entre os algoritmos, foi utilizada a biblioteca AutoRank proposta por [Herbold 2020]. O AutoRank segue o protocolo estabelecido por [Demšar 2006], que primeiro avalia se as amostras seguem uma distribuição normal antes de aplicar o teste estatístico adequado. Se os dados não forem normais, correções de Cohen são aplicadas. Em seguida, o teste não paramétrico de Friedman é aplicado para afirmar a hipótese nula de que não há diferença significativa entre as abordagens. Se a hipótese nula for rejeitada, o teste post-hoc de Nemenyi é aplicado para identificar quais abordagens têm desempenhos diferentes. Ambos os testes são realizados com um nível de significância de $\alpha = 0,05$.

4.4. Configuração dos Experimentos

Para a composição do Sistema Híbrido, foram selecionados dois algoritmos: a Regressão Linear e a Árvore de Decisão. A Regressão Linear foi escolhida por ser um modelo com treinamento rápido e capaz de capturar principalmente correlações lineares, enquanto a Árvore de Decisão foi escolhida por sua natureza não-linear e boa capacidade de generalização em diferentes problemas. O Sistema Híbrido é comparado a seis algoritmos individuais: três lineares (Regressão Linear, Lasso e Lars Lasso) e três não lineares (Árvore de Decisão, MLP Regressor e SVR), obtidos da plataforma Scikit-Learn¹.

¹<https://scikit-learn.org/stable/index.html>

Os seis algoritmos foram submetidos a técnica Grid Search que avalia todas as combinações possíveis de um conjunto predefinido de parâmetros. Os parâmetros que geraram o melhor desempenho de previsão para o conjunto de treinamento estão destacados na Tabela 1 e foram utilizados nos experimentos.

Tabela 1. Gridsearch e melhores parâmetros por modelo

Algoritmos	Grid Search	Melhores Parâmetros
LinearRegression	fit_intercept: [True, False], copy_X: [True, False]	copy_X: True, fit_intercept': True
Lasso	alpha: [0.001, 0.1, 1, 10, 100], max_iter:[1000,5000, 10000]	alpha: 0.001, max_iter': 1000
LassoLars	—	—
DecisionTreeRegressor	max_depth:[10,20,30], min_samples_split: [1,2,4], min_samples_leaf: [7, 8, 9, 10, 11,15,20]	max_depth: 10, min_samples_leaf: 4, min_samples_split: 2
MLPRegressor	hidden_layer_sizes: [(5,), (10,), (15,), (20,), (30,)], activation: ['relu', 'tanh'], alpha: [0.0001, 0.001], max_iter': [50, 80, 100, 150]	activation: tanh, alpha: 0.0001, hidden_layer_sizes: (30,), max_iter: 80
SVR	kernel: ['rbf', 'sigmoid'], C: [0.1, 1, 10, 100]	C: 0.1, kernel: linear

Os experimentos foram realizados usando a linguagem de programação Python e executados na plataforma do Google Colab. Para os experimentos com as séries meteorológicas, cada algoritmo foi executado 50 vezes por ser uma amostra significativamente grande e que permite entender a variação de diferentes configurações de inicialização dos algoritmos. Esse processo fornece uma amostra segura para os testes de Friedman com o Pós-teste de Nemenyi.

5. Resultados

A média e o desvio padrão para o MSE dos algoritmos para as nove bases de dados estão apresentados na Tabela 2. Os algoritmos com melhores desempenhos para cada base estão destacados em negrito e para afirmar esses resultados estatisticamente, pode-se observar a Figura 3. As discussões dos resultados serão organizadas em duas subseções, a análise do desempenho dos modelos (5.1) lineares, e (5.2) não-lineares.

Tabela 2. Resultados dos modelos para o MSE.

BASES	MLP REGRESSOR	ARVORE DE DECISÃO	SVM	LASSO	LASSO LARS	REGRESSÃO LINEAR	SH
TEMP MED	0.0021 (0.000)	0.0017 (1.492)	0.0026 (0.000)	0.0016 (2.168)	0.0021 (0.000)	0.0016 (0.000)	0.0016 (1.365)
PA MED	0.0006 (0.000)	0.0003 (2.134)	0.0017 (2.168)	0.0002 (0.000)	0.0007 (1.084)	0.0002 (5.421)	0.0002 (1.387)
VEL VEN MED	0.0069 (0.000)	0.0067 (1.613)	0.0085 (1.734)	0.0068 (1.734)	0.0068 (0.000)	0.0068 (0.000)	0.0068 (2.038)
DIR VEN MED	0.0219 (0.000)	0.0198 (1.298)	0.0243 (0.000)	0.0216 (3.469)	0.0216 (0.000)	0.0216 (3.469)	0.0198 (3.469)
PO MED	0.0043 (0.000)	0.0037 (5.018)	0.0035 (0.000)	0.0036 (4.336)	0.0041 (8.673)	0.0035 (0.000)	0.0035 (0.000)
UR MED	0.0049 (0.000)	0.0031 (1.357)	0.0067 (0.000)	0.0038 (8.673)	0.0045 (8.673)	0.0038 (8.673)	0.0031 (1.256)
RAD MED	0.0072 (0.000)	0.0063 (9.090)	0.0090 (0.000)	0.0091 (0.000)	0.0093 (0.000)	0.0090 (0.000)	0.0063 (4.544)
TENSAO BAT	0.0134 (0.001)	0.0006 (1.036)	0.0089 (0.000)	0.0169 (3.469)	0.0178 (0.000)	0.0169 (0.000)	0.0189 (0.000)
TEMP INST	0.0031 (0.000)	0.0027 (6.433)	0.0028 (4.336)	0.0026 (4.336)	0.0029 (4.336)	0.0026 (4.336)	0.0026 (4.336)

5.1. Análise do Desempenho dos Modelos Lineares

Nesta primeira análise, ao observar a Tabela 2, percebe-se que o Sistema Híbrido obteve os melhores resultados para cinco das nove bases, enquanto que a Regressão Linear obteve o melhor desempenho em uma e empatou com o Sistema Híbrido em duas. No ranking

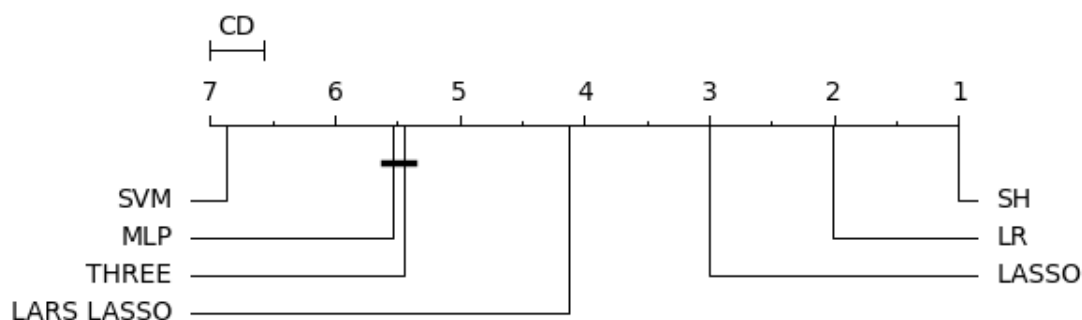


Figura 3. Teste de Friedman com o pós-teste de Nemenyi para o MSE

de Friedman (Fig. 3) observa-se que os modelos lineares figuraram as melhores posições, na ordem do melhor para o pior: Sistema Híbrido, Regressão Linear, LASSO e LARS Lasso.

A distância crítica de Nemenyi (CD) ilustra quando dois algoritmos possuem desempenho estatisticamente iguais, especificamente quando uma linha se sobrepõe a dois algoritmos. A partir dos resultados observados, nota-se que todos os algoritmos melhores ranqueados são estatisticamente diferentes dos demais, com destaque para o primeiro colocado: o Sistema Híbrido (SH). Estes resultados apontam para a predominância de padrões lineares nas séries temporais climáticas experimentadas.

Para visivelmente entender os resultados, as Figuras 4a, 4b, 4c, 5a, 5b, 5c ilustram o desempenho de previsão de todos os algoritmos para o conjunto de testes. As representações visuais foram geradas apenas para as últimas 50 observações de cada conjunto de teste, por conta do seu tamanho elevado (1.322 observações), o que dificultaria a interpretação das previsões.

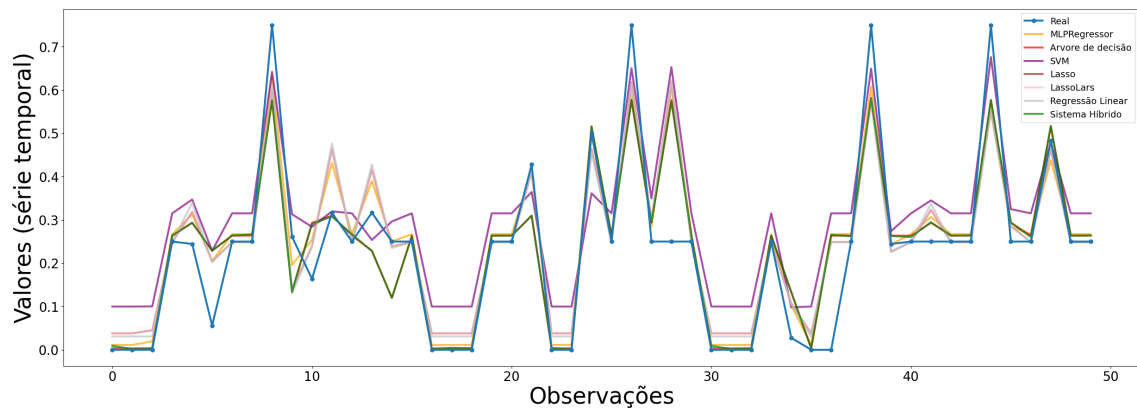
Em destaque, nota-se que nas Figuras 4c, 5b os modelos lineares foram os que mais se aproximaram dos dados reais. Em contrapartida, nas Figuras 4a, 5c os seus desempenhos foram inferiores aos dos modelos não-lineares, como da Árvore de Decisão.

5.2. Análise do Desempenho dos Modelos Não-lineares

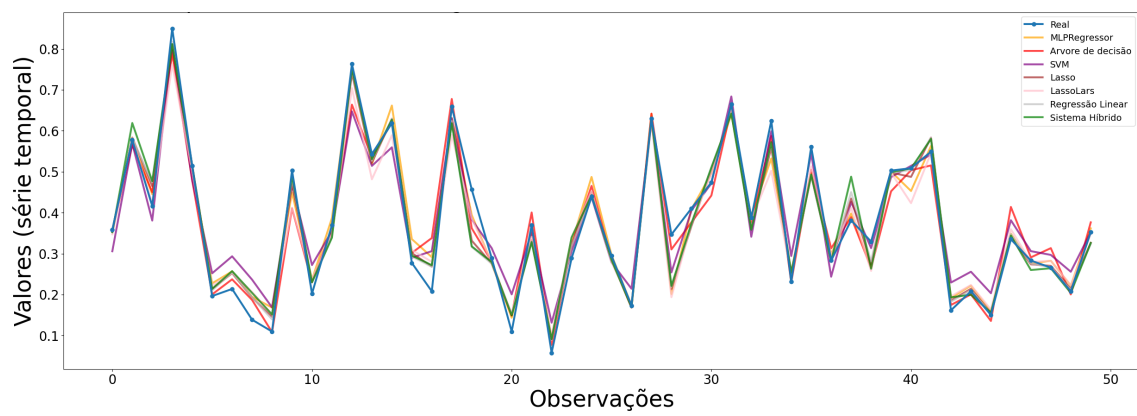
Nesta segunda análise, ao observar a Tabela 2, percebe-se que o modelo não-linear Árvore de Decisão (THREE) apresentou os melhores desempenhos de previsão para 3 das 9 bases. No entanto, mostrou desempenho estatisticamente igual ao MLP (Fig. 3).

No geral, nota-se que os modelos não-lineares estiveram entre as piores colocações, apesar de terem previsões bastante próximas dos valores reais nas Figuras 4b, 5a e 5c. Entende-se que a maior variabilidade dos resultados ocorre porque os modelos são mais suscetíveis a certas características dos dados que não se repetem em todos as bases avaliadas. Por esse motivo, os modelos não-lineares não foram suficientemente robustos para garantir uma boa previsão nos diferentes tipos de séries temporais climáticas.

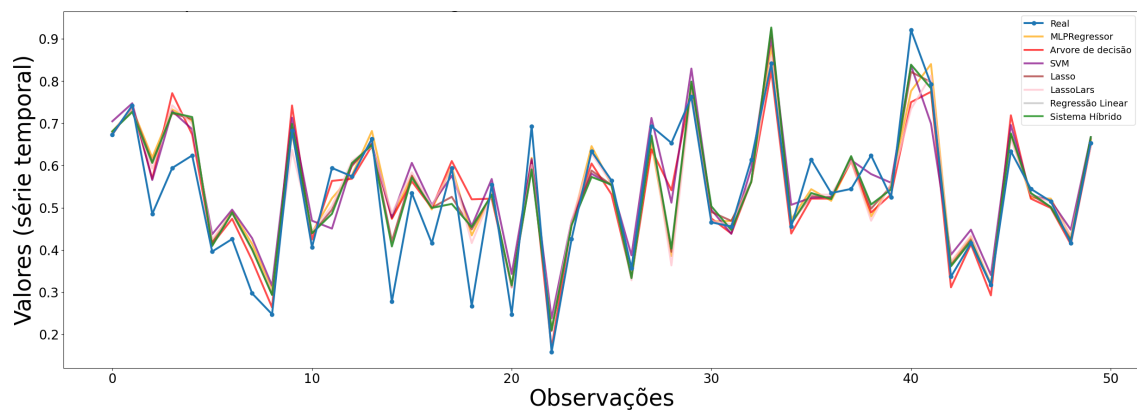
Em contrapartida, o Sistema Híbrido, na maioria dos gráficos, soube lidar com as diferentes características dos conjuntos de dados, demonstrando uma previsão superior em relação aos demais modelos. Esses resultados confirmam a nossa hipótese de pesquisa de que a combinação de modelos distintos de previsão, especialmente um linear e um não-linear, pode oferecer previsões climáticas mais precisas e robustas em comparação aos modelos individuais de previsão.



(a) DIR VEM MED

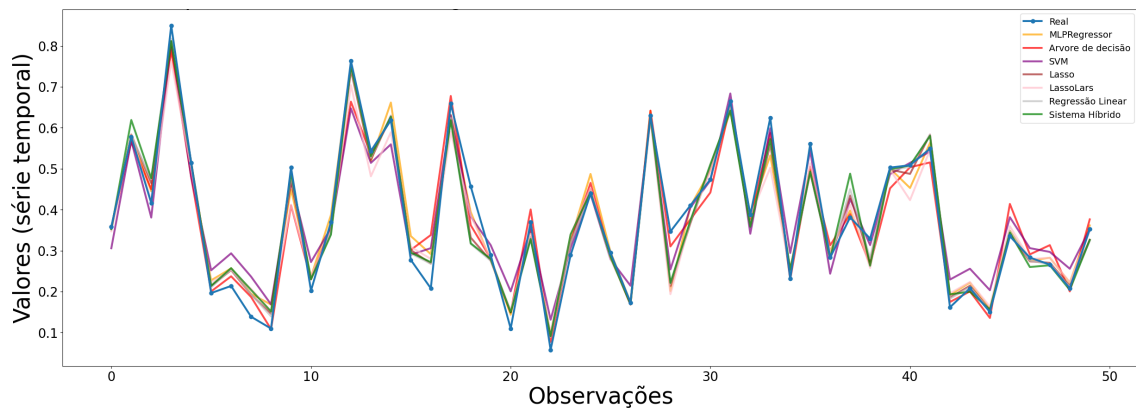


(b) PA MED

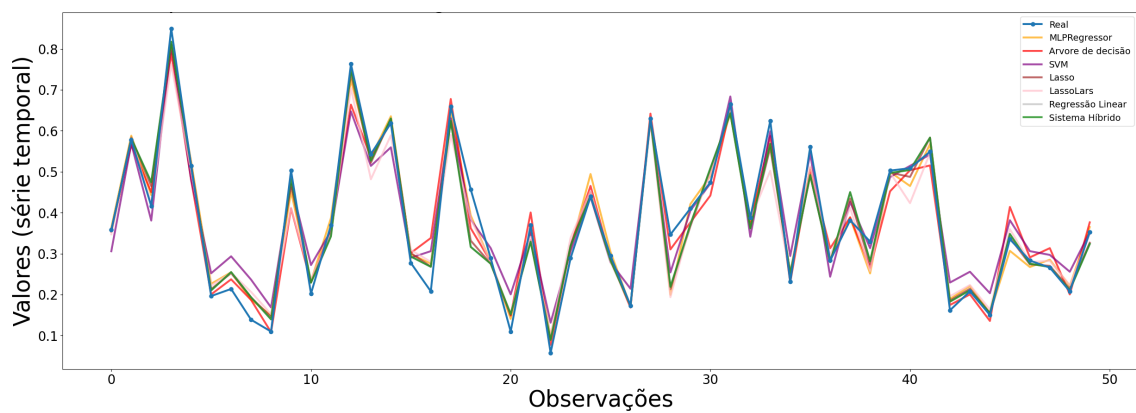


(c) PO MED

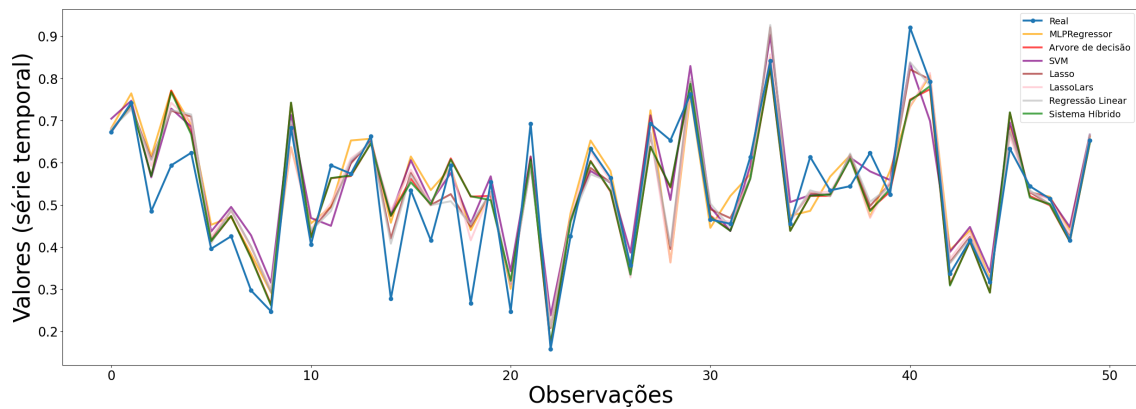
Figura 4. Previsões para as séries meteorológicas de direção do vento, pressão atmosférica e ponto de orvalho



(a) RAD MED



(b) TEMP MED



(c) UR MED

Figura 5. Previsões para as séries meteorológicas de radiação, temperatura e unidade relativa média

6. Conclusão

Este trabalho relatou o resultado de um projeto de iniciação científica que teve como objetivo o desenvolvimento de um Sistema Híbrido para previsão de séries temporais climáticas na cidade de Penedo, Alagoas. As séries temporais climáticas experimentadas foram obtidas na estação meteorológica do IFAL em Penedo.

O Sistema Híbrido desenvolvido obteve os melhores resultados em cinco das nove bases avaliadas, destacando-se como o algoritmo com melhor desempenho geral, conforme indicado pelo ranking de Friedman e pela distância crítica de Nemenyi. Os resultados também revelam que os modelos não-lineares avaliados estiveram entre as piores colocações no Ranking de Friedman, mostrando que não conseguiram capturar adequadamente os padrões das séries climáticas. Esses indícios confirmaram a hipótese de pesquisa deste estudo que: a combinação de modelos distintos, como um linear e um não-linear, poderia oferecer robustez na previsão de diferentes séries temporais climáticas.

Para trabalhos futuros, sugere-se a investigação da seleção dinâmica de regressores, onde o modelo mais adequado para cada padrão de entrada é utilizado. Também é motivada a integração de técnicas avançadas de aprendizado profundo, como redes LSTM. Para aprofundar as discussões, é recomendado incluir mais variáveis climáticas de outras regiões com diferentes características e implementar uma plataforma web para disponibilizar as previsões ao público.

Referências

- Aly, H. H. (2020). An intelligent hybrid model of neuro wavelet, time series and recurrent kalman filter for wind speed forecasting. *Sustainable Energy Technologies and Assessments*, 41:100802.
- BARROS, A. H. C., de Araújo Filho, J., Da Silva, A., and SANTIAGO, G. A. C. (2012). Climatologia do estado de alagoas. .
- Bochenek, B. and Ustrnul, Z. (2022). Machine learning in weather prediction and climate analyses—applications and perspectives. *Atmosphere*, 13(2):180.
- de Mattos Neto, P. S., de Oliveira, J. F. L., Júnior, D. S. d. O. S., Siqueira, H. V., Marinho, M. H. D. N., and Madeiro, F. (2020). A hybrid nonlinear combination system for monthly wind speed forecasting. *IEEE Access*, 8:191365–191377.
- de Oliveira, J. F., Silva, E. G., and de Mattos Neto, P. S. (2021). A hybrid system based on dynamic selection for time series forecasting. *IEEE Transactions on Neural Networks and Learning Systems*, 33(8):3251–3263.
- de Oliveira, J. F. L. and Ludermir, T. B. (2014). A hybrid evolutionary system for parameter optimization and lag selection in time series forecasting. In *2014 Brazilian Conference on Intelligent Systems*, pages 73–78. IEEE.
- Demšar, J. (2006). Statistical comparisons of classifiers over multiple data sets. *Journal of Machine learning research*, 7(Jan):1–30.
- Faruk, D. Ö. (2010). A hybrid neural network and arima model for water quality time series prediction. *Engineering applications of artificial intelligence*, 23(4):586–594.
- Figueiredo Filho, D. B. and Silva Júnior, J. A. (2009). Desvendando os mistérios do coeficiente de correlação de pearson (r). *Revista Política Hoje*, 18(1):115–146.
- Freitas, C. (2023). Dataset e corpus. *Processamento de linguagem natural: conceitos, técnicas e aplicações em português*.
- Han, S., Xu, J., Yan, M., and Liu, Z. (2022). Using multiple linear regression and bp neural network to predict critical meteorological conditions of expressway bridge pavement icing. *Plos one*, 17(2):e0263539.

- Herbold, S. (2020). Autorank: A python package for automated ranking of classifiers. *Journal of Open Source Software*, 5(48):2173.
- Morettin, P. A. and Toloí, C. M. d. C. (1985). Previsão de séries temporais. .
- SILVA, H. B. d. (2018). Técnicas para redução de dimensionalidade de séries temporais e detecção de velocidades extremas do vento para geração eólica. Master's thesis, Universidade Federal de Pernambuco.
- Soares, A. C. et al. (2022). Análise de séries temporais com uso de redes neurais artificiais em dados meteorológicos para previsão de chuva e eventos climáticos severos. .
- Zhang, G. P. (2003). Time series forecasting using a hybrid arima and neural network model. *Neurocomputing*, 50:159–175.