Previsão de séries temporais de consumo diário de gás natural no Brasil

Vinicius Claudino Ferraz * **Patrícia de Oliveira e Lucas** *,** **Frederico Gadelha Guimarães** *** **Walmir Matos Caminhas** ****

* *Programa de Pós-Graduação em Engenharia Elétrica, Universidade Federal de Minas Gerais (UFMG) Av. Antônio Carlos 6627, 31270-901, Belo Horizonte, Brasil*
** *Instituto Federal do Norte de Minas Gerais (IFNMG), Campus Salinas, Brasil*
*** *Departamento de Ciência da Computação, Universidade Federal de Minas Gerais (UFMG), Belo Horizonte, Brasil*
**** *Departamento de Engenharia Eletrônica, Universidade Federal de Minas Gerais (UFMG), Belo Horizonte, Brasil*

Abstract: The forecast of natural gas consumption assumes crucial importance in the short and long-term planning of distributing companies. Anticipating the volume of gas to be distributed not only enables future investments in infrastructure and resource management policies, but also avoids potential economic losses. Despite the relevance of natural gas in the Brazilian energy matrix, particularly for the industrial sector, this significance is not adequately reflected in research in the field, creating a gap in published studies, especially regarding short-term forecasts. In this context, the objective of this study was to conduct a daily forecast study, with horizons of 1, 4, and 7 steps ahead, for three specific natural gas Receiving Stations. To this end, the treatment applied to the time series of each delivery point was presented, followed by the evaluation of classical, machine learning, and deep learning models. The results obtained show that forecasting natural gas consumption by receiving stations is not a trivial task and that machine learning models did not outperform classical forecasting methods.

Resumo: A previsão do consumo de gás natural assume uma importância crucial no planejamento de curto e longo prazo das empresas distribuidoras. Antecipar o volume de gás a ser distribuído não apenas viabiliza investimentos futuros em infraestrutura e políticas de gestão desse recurso, mas também evita possíveis perdas econômicas. Apesar da relevância do gás natural na matriz energética brasileira, sobretudo para o setor industrial, essa importância não se reflete devidamente nas pesquisas da área, gerando uma lacuna nos estudos publicados, especialmente no que diz respeito às previsões de curto prazo. Nesse contexto, o objetivo deste trabalho foi realizar um estudo de previsão diária, com horizontes de 1, 4 e 7 passos à frente, para três Estações de Recebimento de gás natural específicas. Para tanto, foi apresentado o tratamento aplicado nas séries temporais de cada ponto de entrega, seguido da avaliação de modelos clássicos, de aprendizado de máquina e de aprendizado profundo. Os resultados obtidos evidenciam que a previsão do consumo de gás natural por Estações de Recebimento não é uma tarefa trivial e que os modelos de aprendizado de máquina não superaram os métodos clássicos de previsão.

Keywords: Natural gas, forecasting, industrial, time series, consumption.
Palavras-chaves: Gás natural, previsão, industrial, séries temporais, consumo.

1. INTRODUÇÃO

O gás natural é uma fonte de energia fóssil formada a partir da decomposição da matéria orgânica de animais e plantas pré-históricos. Ele é extraído por meio da perfuração de depósitos subterrâneos e pode ser encontrado tanto associado quanto não associado a depósitos de petróleo. Em comparação com outros combustíveis fósseis, como o carvão e o petróleo, o gás natural é considerado uma fonte de energia mais limpa, de acordo com o (EIA, 2022).

A indústria do gás natural é dividida em produção, transporte e distribuição. Para a distribuição, o gás natural é transportado por dutos até empresas distribuidoras, que fornecem para consumidores residenciais, comerciais e industriais. A regulação da distribuição de gás natural é feita pelo governo; e os contratos de fornecimento geralmente incluem penalidades para previsões de consumo com erros fora de um intervalo de tolerância (Soldo et al., 2014).

⋆ O presente trabalho foi realizado com o apoio financeiro das agências de fomento CAPES e CNPq, processo no. 312991/2020-7 e 309832/2022-5; e da Companhia de Gás de Minas Gerais – GASMIG por meio de Projeto de PD&I.

Para evitar perdas econômicas, previsões precisas do consumo de gás natural são cruciais para empresas distribuidoras no planejamento de curto prazo. Além das questões regulatórias, a previsão de gás natural auxilia no planejamento de investimentos futuros em infraestrutura e políticas de gestão desse recurso (Shaikh and Ji, 2016).

Segundo Khan (2015), a natureza da demanda de energia e o conhecimento de seus determinantes são cruciais para previsões precisas das necessidades de energia. Por isso, a previsão do consumo de gás natural tem sido investigada em diferentes níveis (mundial, nacional, regional, por sistema de distribuição, por cliente individual, etc), diferentes horizontes de previsão (curto, médio e longo prazo) e diferentes fatores de influência (parâmetros econômicos, dados meteorológicos, dados históricos de consumo de gás natural, demográficos e de calendário) (Liu et al., 2021).

O consumo também varia de acordo com o tipo de consumidor. Consumidores industriais consomem a mesma quantidade de gás natural independente da estação do ano e do dia. Consumidores residenciais, que usam o gás para aquecimento, são caracterizados por ter um consumo significativamente menor no verão do que no inverno. Já os consumidores residenciais, que usam o gás apenas para cozinhar, são caracterizados pela variação diurna do consumo (Szoplik, 2015). Gao and Shao (2021) ainda acrescentam que as diferentes condições e estágios de desenvolvimento de um país também influenciam no seu consumo de gás natural.

No Brasil, o gás natural ocupa a terceira posição na matriz energética nacional, ficando atrás apenas do petróleo e da cana-de-açúcar. Nos últimos dez anos, sua participação evoluiu de 5,4% para 13,3%, graças à sua utilização em termoelétricas e à extensão da malha duto-viária, que permitiu seu uso tanto pela indústria quanto em edificações residenciais, comerciais e públicas. O setor industrial é o maior consumidor de gás natural no país, respondendo por cerca de 34% do consumo total, e esse perfil tem se mantido relativamente constante desde 2000. Já no setor residencial, o consumo é de 11,5%, e sua utilização está relacionada à cocção de alimentos e aquecimento de água (EPE, 2018).

A previsão de consumo de gás natural é uma tarefa desafiadora devido à sua complexidade e à influência dos vários fatores citados. Ao longo dos anos, foram testados muitos modelos em diferentes cenários. Segundo Liu et al. (2021), a partir de 2006, com a conscientização sobre conservação de energia e proteção ambiental, o número de trabalhos na área aumentou e envolve modelos como: modelos de aprendizagem de máquina gerais e de aprendizado profundo (*deep learning*), modelos combinados (*ensemble learning*) e modelos estatísticos tradicionais.

Apesar da importância do gás natural na matriz energética brasileira, especialmente para a indústria, há poucos estudos de previsão de consumo de gás natural no Brasil. Nenhum trabalho de previsão de curto prazo foi encontrado na literatura e apenas um com previsão de médio prazo. Nesse contexto, o objetivo desse trabalho é fazer um estudo de previsão diária de 1, 4 e 7 passos à frente de três diferentes estações de recebimento de gás natural da Companhia de Gás de Minas Gerais (GASMIG). Sete métodos de previsão diferentes foram comparados: NAIVE, Auto-regressivo integrado de médias móveis (ARIMA), *Weighted High Order Fuzzy Times Series* (WHOFTS), *Evolving neo-fuzzy neural network* (NFN) (Silva et al., 2014), Perceptron Multicamadas (MLP), *Random Forest* (RF) e *Long Short Term Memory* (LSTM).

O restante deste trabalho está organizado da seguinte forma. A Seção 2 fornece uma breve revisão do estado da arte em previsão de vazão de gás natural. Na Seção 3, são descritas a metodologia de pré-processamento das séries temporais e a comparação dos métodos de previsão. A Seção 4 descreve os experimentos, resultados e discussões. Por fim, a Seção 5 contém algumas conclusões e possibilidades para trabalhos futuros.

2. TRABALHOS RELACIONADOS

Pesquisas que fazem previsões de longo e médio prazo trabalham com séries de dados anuais, trimestrais ou mensais. Normalmente, essas previsões utilizam uma variação maior de variáveis de entrada, como fatores econômicos, população, dados climáticos e dados históricos de consumo. Em Gao and Shao (2021), o produto interno bruto (PIB) e a população foram utilizados para treinar um modelo ANFIS (*Adaptive Neuro-fuzzy Inferece System*) para previsão anual de gás natural em quatro países do Oriente Médio. Sen et al. (2019) utilizam dados históricos de população, PIB, inflação, preço do gás natural e a temperatura média do inverno para fazer previsões anuais na Turquia. Para isso, utilizaram os modelos de regressão logarítmica, linear e não linear e análise de correlação para selecionar os valores defasados mais relevantes.

Behrouznia et al. (2010) usam 5 modelos ANFIS para fazer previsões anuais para 5 países: Argentina, Brasil, Colômbia, Venezuela e Cuba. Os modelos foram treinados usando dados históricos de população, PIB e consumo de gás natural. Costa et al. (2019) também usam as mesmas variáveis para fazer a previsão anual no Brasil com o modelo de estimativa dos mínimos quadrados robusto. O estudo apresenta um modelo para previsão do consumo de gás natural de todos os setores (industrial, geração de energia elétrica, comercial e público, residencial e transporte) e dois modelos específicos para indústria e geração de energia elétrica, respectivamente.

Já em Gracias et al. (2012), foram avaliados modelos utilizados no estudo da dinâmica populacional para prever o consumo, a produção e a importação anual de gás natural no Brasil. Os autores concluíram que o modelo logístico é o mais indicado para previsão de consumo, exponencial para previsão de produção e *Bertalanffy* para previsão de importação.

Beyca et al. (2019) apresentam um estudo sobre previsão mensal de gás natural em Istambul na Turquia. Como variáveis de entrada, usou-se consumo de gás natural, temperatura, preço do gás natural e população. Os modelos de regressão linear multivariada, SVM (*Support Vector Machine*) e RNA (Redes Neurais Artificiais) foram testados e comparados. O SVM apresentou melhores resultados. Meira et al. (2022) propõem prever o consumo mensal de gás natural de 18 economias europeias usando dados históricos de consumo. A abordagem seguida é um ensemble do tipo Bagging composto pelos modelos RNA, suavização

exponencial e ARIMA (*Autoregressive Integrated Moving Average*).

Santos (2019) aplica regressão linear para prever o consumo mensal de gás natural para 4 segmentos no Brasil: (1) residencial e comercial, (2) automotivo, (3) industrial e (4) geração de energia elétrica. Todos os modelos usaram como entrada o crescimento populacional, PIB e dados históricos de consumo. Além disso, o modelo (1) usou taxa de câmbio, (3) indicador de produção industrial e (4) energia natural afluente.

Previsões de curto prazo envolvem séries de dados semanais, diárias ou horárias e permitem que empresas de distribuição consigam equilibrar a oferta e a demanda de seus consumidores. Para essas previsões, usa-se principalmente dados climáticos, o tipo de data e dados históricos de consumo. Ravnik and Hriberšek (2019) apresentam um estudo que cria perfis de consumo para realizar a previsão diária de consumo de gás natural de 1 passo à frente na Eslovênia. As previsões foram feitas levando-se em consideração dados históricos de consumo e temperatura usando regressão logística. Em Soldo et al. (2014), investigou-se a influência da radiação solar na previsão de dois sistemas de consumo: casa individual e distribuidora local. Foram avaliados os modelos RNA, SVM, passeio aleatório, modelo de correlação de temperatura, regressão *stepwise* e um modelo linear *Autoregressive-exogenous* (ARX). Os autores concluíram que o uso da radiação solar melhorou a precisão das previsões e que o modelo ARX apresentou melhor desempenho.

Em Anagnostis et al. (2020), é feito um estudo com RNA para previsão de consumo diário de gás natural em 15 cidades da Grécia. Avaliou-se uma RNA treinada com dados históricos de consumo e temperatura, uma LSTM (*Long Short Term Memory*) treinada apenas com dados históricos de consumo e uma DNN com dados históricos de consumo, temperatura e indicadores de dia (mês, dia, dia da semana e feriado). Resultados apontaram melhor desempenho para o modelo DNN.

Andelkovic and Bajatovic (2020) apresentam uma abordagem híbrida de modelos *deep learning* para previsão de consumo diário de gás natural de várias regiões da China. A estrutura funciona da seguinte forma: um *autoencoder* é usado para extrair informações estáticas e dinâmicas da série histórica de consumo, uma LSTM é aplicada também à série de consumo para capturar o padrão de dependência de longo prazo e, por fim, uma camada de rede personalizada faz o acoplamento dessas informações para realizar a previsão. A abordagem proposta superou outros modelos como LSTM, um modelo híbrido de CNN (*Convolutional Neural Network*) e LSTM, SVM e árvore de decisão.

O estudo de Rehman et al. (2022) apresenta uma comparação de modelos *deep learning* e métodos estatísticos na previsão horária de consumo de gás natural em 5 locais na Espanha. As variáveis de entrada usadas foram: dados históricos de consumo, indicador de feriados ao longo do ano e condições meteorológicas.

Em Andelkovic and Bajatovic (2020), as variáveis meteorológicas são utilizadas juntamente com os históricos de consumo para previsão horária de gás natural na cidade de Novi Sad na Sérvia. Os autores perceberam que as variáveis temperatura, radiação e humidade relativa tinham grande influência no consumo de gás natural, já que grande parcela é usada para aquecimento das residências na cidade. O modelo ANFIS foi usado para fazer as previsões.

Outros três trabalhos de previsão horária fazem uso de técnicas de decomposição de séries temporais como forma de extrair mais informações dos dados originais. Em Svoboda et al. (2021), é feita uma comparação dos algoritmos *Random Forest* e Gradiente *Boosting*. Os autores utilizaram várias variáveis para treinar os modelos, são elas: componente sazonal, de tendência e residual da série temporal de consumo de gás natural, indicador de feriados e dias que os antecedem, estações do ano, variáveis climáticas, preço do gás natural, indicador de dia, dia da semana e mês.

Su et al. (2019) apresenta uma estrutura híbrida que usa a transformada de Wavelet, uma rede LSTM e um Algoritmo Genético (AG). A transformada de Wavelet é usada para decompor a série histórica de consumo de gás natural em componentes de alta e baixa frequência. Essas componentes são dadas como entrada para a LSTM, cuja arquitetura é selecionada usando um AG. Qiao et al. (2019) decompõe a série original de consumo de gás natural em várias IMF's (*Intrinsic Mode Function*) e um termo de tendência residual para treinar três tipos de RNA e um SVM.

3. METODOLOGIA

Nesta Seção é apresentada a metodologia usada no pré-processamento dos dados e na configuração dos modelos usados para prever as séries temporais diárias de consumo de gás natural de 1, 4 e 7 passos à frente. O estudo foi realizado em três bases de dados referentes a três estações de recebimento de gás natural diferentes da Companhia de Gás de Minas Gerais (GASMIG).

3.1 Pré-processamento das bases de dados

As bases de dados utilizadas neste estudo abrangem um período de 6 anos (2018 a 2023) de consumo diário de três estações de recebimento de gás natural distintas, denominadas A, B e C. Os dados até junho de 2023 (B1) foram registrados diretamente pelos sensores, porém apresentam algumas inconsistências, como o registro de vazões negativas e *outliers*. Já os registros de julho a dezembro de 2023 (B2) foram realizados manualmente e não apresentam tais inconsistências, o que os torna uma referência confiável para a correção dos dados B1.

Para corrigir as inconsistências de B1, calculou-se a média (μ_{B2}) e o desvio padrão (σ_{B2}) para cada dia da semana usando B2 e em seguida substituiu-se os valores negativos e *outliers* de B1 pela média μ_{B2}. Foram considerados *outliers* os valores maiores ou menores que $\mu_{B2} \pm \sigma_{B2}$.

A Figura 1 apresenta uma amostra normalizada da série temporal de consumo diário referente ao ano 2023 para cada estação. Percebe-se que, por serem estações que atendem consumidores industriais, nas três séries temporais, o consumo não varia independente da estação do ano e do dia. Esse padrão é visível também nos outros anos (2018 a 2022).

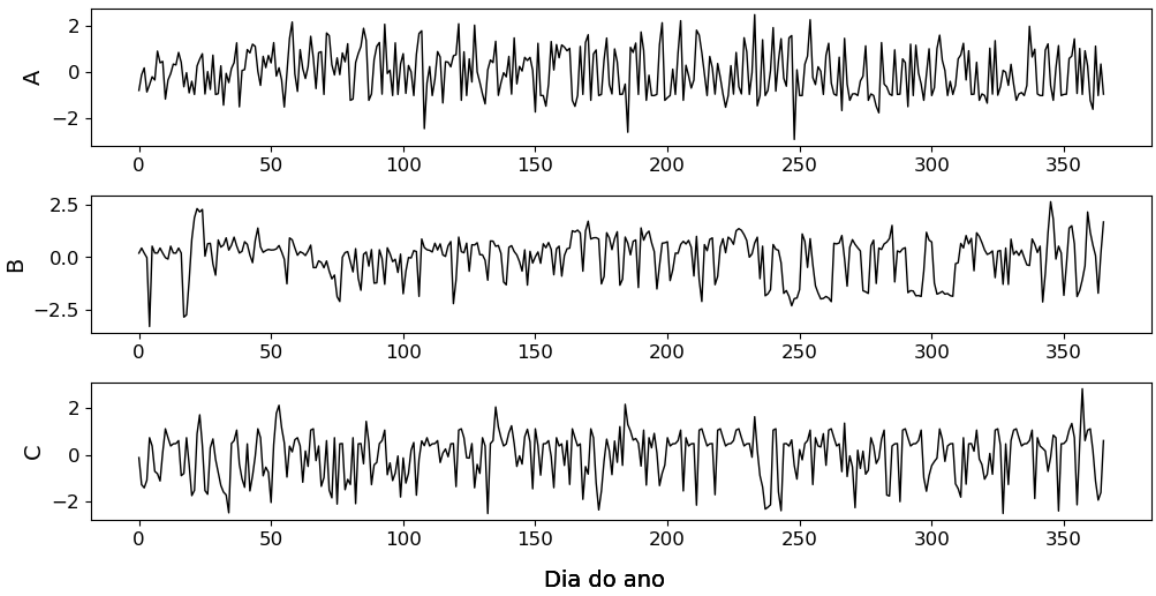


Figura 1. Série temporal referente ao ano 2023 de consumo diário de gás natural das estações de recebimento A, B e C.

3.2 Configuração dos modelos

Os modelos comparados foram o NAIVE, Auto-regressivo integrado de médias móveis (ARIMA), *Weighted High Order Fuzzy Times Series* (WHOFTS), *Evolving neo-fuzzy neural network* (NFN), Perceptron Multicamadas (MLP), *Random Forest* (RF) e *Long Short Term Memory* (LSTM). Para o NAIVE, ARIMA e WHOFTS, foi empregada a abordagem de previsão recursiva. Para o NFN, RF e MLP, adotou-se um modelo especializado para cada horizonte de previsão, enquanto para o LSTM foi utilizada a técnica de previsão direta, ou seja, o modelo possui uma saída para cada horizonte de previsão.

Os hiperparâmetros dos modelos foram otimizados utilizando um ano dos primeiros registros das bases de dados. Para a otimização dos hiperparâmetros do ARIMA utilizou-se a função `auto_arima` da biblioteca Pyramid [1]. Os resultados obtidos foram os seguintes: para A, $p = 1$, $d = 1$ e $q = 2$; para B, $p = 3$, $d = 1$ e $q = 1$; e para C, $p = 1$, $d = 1$ e $q = 1$. Aqui, p representa a ordem do modelo, d é o grau de diferenciação e q é a ordem do modelo de média móvel.

Para otimizar os hiperparâmetros do WHOFTS, foi utilizado o Algoritmo Genético (AG) desenvolvido por Lucas et al. (2019), implementado na biblioteca PyFTS [2]. Os hiperparâmetros otimizados incluíram o número de partições (np), método de particionamento (mp), função de pertinência (fp), ordem (o), α-cut (ac) e lags (l). Os resultados obtidos foram os seguintes:

- Para A: $np = 25$, $mp =$ grid, $fp =$ triangular, $o = 1$, $ac = 0.2$ e $l = 1$.
- Para B: $np = 12$, $mp =$ grid, $fp =$ triangular, $o = 1$, $ac = 0.19$ e $l = 1$.
- Para C: $np = 35$, $mp =$ grid, $fp =$ triangular, $o = 1$, $ac = 0$ e $l = 1$.

[1] https://github.com/alkaline-ml/pmdarima
[2] https://pyfts.github.io/pyFTS/

Os hiperparâmetros do modelo LSTM foram selecionados de maneira empírica, sendo eles: porcentagem de *dropout* = 0.02, normalização em lote, número de unidades = 67, número de camadas = 4 e número de entradas = 7.

Para o MLP, NFN e RF, foi empregado o método de busca em grade (*grid search*). Em ambos os métodos, o número de entradas foi definido como 7. Os hiperparâmetros otimizados para o MLP foram o número de camadas intermediárias, resultando nos seguintes conjuntos para os casos A, B e C: (5, 20, 5), (10, 50, 10) e (10, 50, 10), respectivamente.

Para o NFN, foi utilizada uma entrada com 5 funções de pertinência para todas as bases de dados.

Os hiperparâmetros otimizados do *Random Forest* (RF) foram os seguintes: o número de árvores de decisão (na), o número máximo de características consideradas ao procurar a melhor divisão (mf) e o número mínimo de amostras necessárias para estar em um nó folha (ms). Os resultados obtidos foram os seguintes:

- Para A: $na = 10$, $mf = \log_2 T$ e $ms = 5$.
- Para B: $na = 5$, $mf = \sqrt{T}$ e $ms = 5$.
- Para C: $na = 10$, $mf = \log_2 T$ e $ms = 5$.

Onde T é o número total de características.

3.3 Métricas

As métricas de avaliação usadas foram o Erro Absoluto Percentual (APE) e o APE Médio (MAPE) (Equação 1 e 2) medidos para cada horizonte de previsão. Nessas equações, n é o número de amostras de teste, y são os valores reais e $\hat{y}$ são os valores previstos pelos modelos.

Além disso, foi calculada a acurácia dos modelos quando a previsão tem APE individual menor que 5%. Essa medida foi empregada para verificar se os modelos avaliados geram previsões dentro da faixa aceitável de erro.

Tabela 1. MAPE das previsões de 1, 4 e 7 passos à frente dos modelos NAIVE, ARIMA, WHOFTS, NFN, MLP, RF e LSTM para A.

	1	4	7
NAIVE	30.49 ± 17.7	27.21 ± 12.10	24.89 ± 16.31
ARIMA	25.12 ± 11.6	23.47 ± 11.28	22.72 ± 7.9
WHOFTS	31.74 ± 22.85	27.27 ± 13.9	22.7 ± 9.9
NFN	28.18 ± 14.35	28.54 ± 17.12	27.82 ± 13.57
MLP	26.52 ± 9.66	26.83 ± 9.52	28.07 ± 12.97
RF	33.87 ± 31.95	33.09 ± 28.97	35.81 ± 36.55
LSTM	59.41 ± 17.72	59.21 ± 17.12	60.01 ± 15.72

Tabela 2. MAPE das previsões de 1, 4 e 7 passos à frente dos modelos NAIVE, ARIMA, WHOFTS, NFN, MLP, RF e LSTM para B.

	1	4	7
NAIVE	17.98 ± 17.27	18.83 ± 19.68	17.68 ± 18.35
ARIMA	18.50 ± 16.64	21.16 ± 16.78	17.05 ± 16.24
WHOFTS	19.97 ± 14.53	22.22 ± 15.26	19.61 ± 16.53
NFN	18.30 ± 8.80	18.80 ± 8.81	19.34 ± 7.06
MLP	20.33 ± 8.64	23.58 ± 10.7	23.47 ± 10.89
RF	21.39 ± 9.60	25.37 ± 12.66	24.43 ± 12.10
LSTM	96.76 ± 1.38	96.43 ± 1.58	96.22 ± 1.60

$$\text{APE} = \frac{|y - \hat{y}|}{y} \cdot 100\% \tag{1}$$

$$\text{MAPE} = \frac{1}{n} \sum_{i=1}^{n} APE_i \tag{2}$$

4. EXPERIMENTOS

O objetivo dos experimentos foi avaliar o desempenho de diferentes modelos na previsão diária de consumo de gás natural. Após o pré-processamento, as bases de dados foram divididas em 10 janelas sobrepostas com deslocamento de 6 meses. Foi usado 1 ano de dados para treino e 6 meses para teste. Os modelos foram treinados e avaliados em cada janela, observando os horizontes de previsão de 1, 4 e 7 dias à frente. A comparação dos modelos é feita usando a média do MAPE das 10 janelas de teste.

4.1 Resultados

Nesta Seção são apresentados os resultados dos experimentos que comparam os modelos NAIVE, ARIMA, WHOFTS, NFN, MLP, RF e LSTM.

As Tabelas 1, 2 e 3 apresentam a média e o desvio padrão do MAPE de 1, 4 e 7 passos à frente das 10 janelas de teste dos pontos A, B e C. Para verificar se existe diferença significativa nas médias dos resultados, os testes estatísticos de Friedman e o teste post-hoc de Nemenyi foram aplicados para cada horizonte de previsão. Os testes foram escolhidos após verificação, por meio do teste de Shapiro-Wilk, que existem amostras que não provêm de uma distribuição normal.

Os resultados dos testes são apresentados em diagramas de diferenças críticas nas Figuras 2, 3 e 4, para A, B e C, respectivamente. Observando os três horizontes de previsão e as três bases de dados, os modelos NAIVE, ARIMA, WHOFTS, NFN, MLP e RF apresentam médias estatisticamente iguais, enquanto que o LSTM obteve desempenho abaixo dos demais.

Tabela 3. MAPE das previsões de 1, 4 e 7 passos à frente dos modelos NAIVE, ARIMA, WHOFTS, NFN, MLP, RF e LSTM para C.

	1	4	7
NAIVE	6.09 ± 4.20	6.68 ± 2.92	6.54 ± 2.89
ARIMA	4.75 ± 2.30	5.40 ± 2.57	5.24 ± 1.57
WHOFTS	5.73 ± 1.42	5.73 ± 1.93	6.12 ± 1.29
NFN	5.77 ± 1.52	7.43 ± 4.75	6.51 ± 0.86
MLP	6.17 ± 0.96	6.54 ± 1.43	6.58 ± 1.14
RF	5.71 ± 0.85	6.44 ± 0.99	6.42 ± 1.05
LSTM	99.46 ± 0.18	99.47 ± 0.18	99.46 ± 0.19

Por fim, a Figura 5 apresenta a acurácia dos modelos de previsão para estimativas com MAPE menores que 5% para as estações de recebimento A, B e C. Percebe-se que, para as bases de dados A e B, os modelos não conseguiram atingir 50% de estimativas dentro da faixa de MAPE menor que 5%. Em C, com exceção do LSTM, todos os modelos atingem acurácia de 50%. O modelo LSTM apresentou acurácia 0 em B e C e muito próxima de 0 em A. Também é possível observar que os modelos ARIMA e NAIVE se sobressaem em relação aos demais em B e C.

4.2 Discussão

O objetivo dos experimentos foi avaliar diferentes modelos estatísticos e de aprendizado de máquina na previsão diária de volume de gás natural em três horizontes de previsão. Os resultados obtidos mostraram que modelos clássicos como o NAIVE e ARIMA apresentaram resultados estatisticamente iguais aos modelos WHOFTS, NFN, MLP e RF. Esse resultado também foi percebido em termos de acurácia, em que o NAIVE e ARIMA se sobressaíram em duas das três bases de dados. O LSTM apresentou resultados bem inferiores aos demais modelos. Isso pode ser explicado por ser um modelo de aprendizado profundo, que depende de um conjunto grande de dados para treinamento.

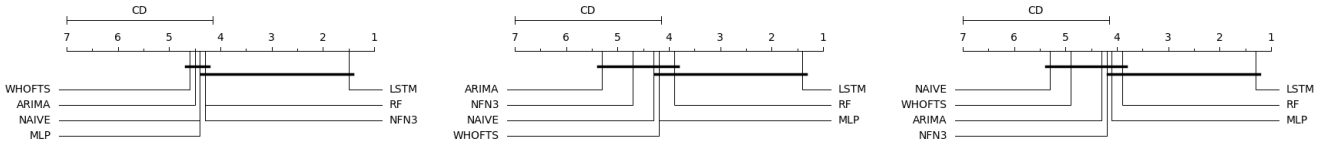
Por fim, além dos experimentos mostrarem uma comparação entre modelos, é importante perceber que o problema de previsão de volume de gás natural não é um problema trivial, pois, no geral, os modelos apresentaram baixa acurácia dentro da faixa aceitável de erro.

5. CONCLUSÃO

O objetivo deste trabalho foi conduzir um estudo de previsão diária de 1, 4 e 7 passos à frente para três diferentes estações de recebimento de gás natural da Companhia de Gás de Minas Gerais (GASMIG). Para isso, foram avaliados modelos clássicos, de aprendizado de máquina e de aprendizado profundo. Os resultados indicaram que os modelos de aprendizado de máquina apresentaram desempenho semelhante aos modelos clássicos de previsão, como NAIVE e ARIMA.

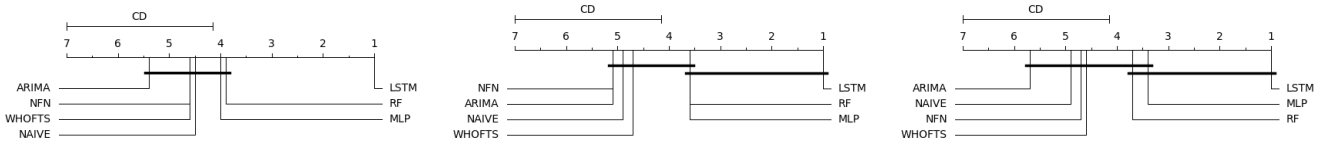
AGRADECIMENTOS

Este trabalho foi parcialmente financiado por (i) Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), processos no. 312991/2020-7 e 309832/2022-5; (ii) Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) por meio do Programa de Excelência



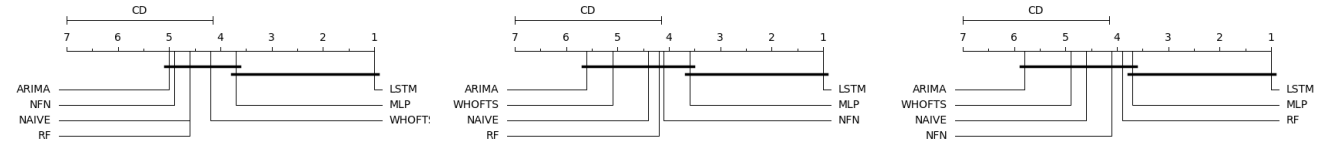
(a) Horizonte 1 (b) Horizonte 4 (c) Horizonte 7

Figura 2. Diagrama de diferenças críticas dos resultados dos modelos NAIVE, ARIMA, WHOFTS, NFN, MLP, RF e LSTM para A.



(a) Horizonte 1 (b) Horizonte 4 (c) Horizonte 7

Figura 3. Diagrama de diferenças críticas dos resultados dos modelos NAIVE, ARIMA, WHOFTS, NFN, MLP, RF e LSTM para B.



(a) Horizonte 1 (b) Horizonte 4 (c) Horizonte 7

Figura 4. Diagrama de diferenças críticas dos resultados dos modelos NAIVE, ARIMA, WHOFTS, NFN, MLP, RF e LSTM para C.

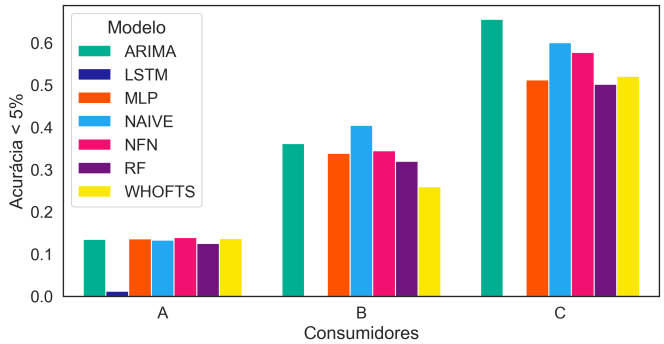


Figura 5. Acurácia dos modelos de previsão para estimativas com APEs individuais menores que 5% para as estações de recebimento A, B e C.

Acadêmica PROEX; (iii) Convênio de PD&I entre a Companhia de Gás de Minas Gerais – GASMIG e a UFMG.

REFERÊNCIAS

Anagnostis, A., Papageorgiou, E., and Bochtis, D. (2020). Application of Artificial Neural Networks for Natural Gas Consumption Forecasting. *Sustainability*, 12(16), 6409. doi:10.3390/su12166409. URL `https://doi.org/10.3390/su12166409`.

Andelkovic, A.S. and Bajatovic, D. (2020). Integration of weather forecast and artificial intelligence for a short-term city-scale natural gas consumption prediction. *Journal of Cleaner Production*, 266, 122096. doi:10.1016/J.JCLEPRO.2020.122096.

Behrouznia, A., Saberi, M., Azadeh, A., Asadzadeh, S.M., and Pazhoheshfar, P. (2010). An adaptive network based fuzzy inference system-fuzzy data envelopment analysis for gas consumption forecasting and analysis: The case of South America. *2010 International Conference on Intelligent and Advanced Systems*. doi:10.1109/icias.2010.5716160. URL `https://doi.org/10.1109/icias.2010.5716160`.

Beyca, O.F., Ervural, B.C., Tatoglu, E., Ozuyar, P.G., and Zaim, S. (2019). Using machine learning tools for forecasting natural gas consumption in the province of Istanbul. *Energy Economics*, 80, 937–949. doi:10.1016/J.ENECO.2019.03.006.

Costa, O.L.V., de Oliveira Ribeiro, C., Ho, L.L., Rego, E.E., Parente, V., and Toro, J. (2019). A robust least square approach for forecasting models: an application to Brazil natural gas demand. *Energy Systems*, 11(4), 1111–1135. doi:10.1007/s12667-019-00351-1. URL `https://doi.org/10.1007/s12667-019-00351-1`.

EIA (2022). Natural gas explained - U.S. Energy Information Administration (EIA). URL `https://www.eia.gov/energyexplained/natural-gas/`.

EPE (2018). Informe: Mercado Internacional de GNL — Ciclo 2018/2027. Technical report, Empresa de Pesquisa Energética - Ministério Minas Energia, Rio de Janeiro. URL `https://www.epe.gov.br/pt/publicacoes-dados-abertos/publicacoes/informe-mercado-internacional-de-gnl-2018-2027`.

Gao, F. and Shao, X. (2021). Forecasting annual natural gas consumption via the application of a novel hybrid model. *Environmen-*

tal Science and Pollution Research, 28(17), 21411–21424. doi:10.1007/S11356-020-12275-W/TABLES/9. URL https://link.springer.com/article/10.1007/s11356-020-12275-w.

Gracias, A.C., Lourenço, S.R., and Rafikov, M. (2012). Estimation of natural gas production, import and consumption in Brazil based on three mathematical models. *Natural Resources*, 3(2), 42–47.

Khan, M.A. (2015). Modelling and forecasting the demand for natural gas in Pakistan. *Renewable and Sustainable Energy Reviews*, 49, 1145–1159. doi:10.1016/J.RSER.2015.04.154.

Liu, J., Wang, S., Wei, N., Chen, X., Xie, H., and Wang, J. (2021). Natural gas consumption forecasting: A discussion on forecasting history and future challenges. *Journal of Natural Gas Science and Engineering*, 90, 103930. doi:10.1016/J.JNGSE.2021.103930.

Lucas, P.d.O.e., Silva, P.C.d.L., and Guimarães, F.G. (2019). Otimização Evolutiva de Hiperparâmetros para Modelos de Séries Temporais Nebulosas. In *Anais do 14º Simpósio Brasileiro de Automação Inteligente*, 1 – 7. Ouro Preto.

Meira, E., Cyrino Oliveira, F.L., and de Menezes, L.M. (2022). Forecasting natural gas consumption using Bagging and modified regularization techniques. *Energy Economics*, 106. doi:10.1016/j.eneco.2021.105760.

Qiao, W., Huang, K., Azimi, M., and Han, S. (2019). A Novel Hybrid Prediction Model for Hourly Gas Consumption in Supply Side Based on Improved Whale Optimization Algorithm and Relevance Vector Machine. *IEEE Access*, 7, 88218–88230. doi:10.1109/ACCESS.2019.2918156.

Ravnik, J. and Hriberšek, M. (2019). A method for natural gas forecasting and preliminary allocation based on unique standard natural gas consumption profiles. *Energy*, 180, 149–162. doi:10.1016/J.ENERGY.2019.05.084.

Rehman, A., Zhu, J.J., Segovia, J., and Anderson, P.R. (2022). Assessment of deep learning and classical statistical methods on forecasting hourly natural gas demand at multiple sites in Spain. *Energy*, 244, 122562. doi:10.1016/J.ENERGY.2021.122562.

Santos, B.M.d. (2019). Projeção de demanda de gás natural no mercado brasileiro utilizando método estat\'\istico.

Sen, D., Günay, M.E., and Tunç, K.M. (2019). Forecasting annual natural gas consumption using socio-economic indicators for making future policies. *Energy*, 173, 1106–1118. doi:10.1016/J.ENERGY.2019.02.130.

Shaikh, F. and Ji, Q. (2016). Forecasting natural gas demand in China: Logistic modelling analysis. *International Journal of Electrical Power & Energy Systems*, 77, 25–32. doi:10.1016/J.IJEPES.2015.11.013.

Silva, A.M., Caminhas, W., Lemos, A., and Gomide, F. (2014). A fast learning algorithm for evolving neo-fuzzy neuron. *Applied Soft Computing*, 14, 194–209. doi:https://doi.org/10.1016/j.asoc.2013.03.022. URL https://www.sciencedirect.com/science/article/pii/S1568494613001373.

Soldo, B., Potocnik, P., Simunovic, G., Saric, T., and Govekar, E. (2014). Improving the residential natural gas consumption forecasting models by using solar radiation. *Energy and Buildings*, 69, 498–506. doi:10.1016/J.ENBUILD.2013.11.032.

Su, H., Zio, E., Zhang, J., Xu, M., Li, X., and Zhang, Z. (2019). A hybrid hourly natural gas demand forecasting method based on the integration of wavelet transform and enhanced Deep-RNN model. *Energy*, 178, 585–597. doi:10.1016/J.ENERGY.2019.04.167.

Svoboda, R., Kotik, V., and Platos, J. (2021). Short-term natural gas consumption forecasting from long-term data collection. *Energy*, 218, 119430. doi:10.1016/J.ENERGY.2020.119430.

Szoplik, J. (2015). Forecasting of natural gas consumption with artificial neural networks. *Energy*, 85, 208–220. doi:10.1016/J.ENERGY.2015.03.084.