Aplicação de Redes Neurais para Previsão dos Custos de Refino e de Extração da Petrobras

José Ronaldo Bezerra Gabriel
IBMEC
Av. Pres. Wilson, 118 - Centro, Rio de Janeiro - RJ
ronaldogabriel@petrobras.com.br

Marcos dos Santos
IBMEC
Av. Pres. Wilson, 118 - Centro, Rio de Janeiro - RJ
marcosdossantos@ime.eb.br

RESUMO

As incertezas estão presentes em qualquer instituição. No setor petrolífero, considerado de intenso capital e alto risco, elas são ainda mais potencializadas. Nessa conjuntura, modelos preditivos podem ser empregados para reduzir essas incertezas e melhorar as decisões empresariais. Diante disso, este artigo propõe um método de previsão para os custos de refino e de extração da Petrobras, baseado em séries temporais, por meio das redes neurais recorrentes. Para atender a esse propósito, foi treinado e testado o modelo numa amostra contendo 72 observações trimestrais de cada custo. Em seguida, foi avaliada a precisão do modelo por meio de gráficos e com o auxílio dos indicadores do erro percentual absoluto médio [MAPE] e da raiz quadrada média do erro [RMSE]. Os resultados mostraram que o modelo proposto foi capaz de oferecer estimativas razoáveis para os custos estudados, indicando que ele pode servir para previsões nos processos decisórios da empresa.

PALAVRAS CHAVE. Custos de Refino e de Extração, Previsão de Séries Temporais, Redes Neurais Recorrentes.

Tópicos: PO na Área de Energia, Petróleo e Gás; Inteligência Computacional.

ABSTRACT

Uncertainties are present in any institution. In the oil and gas industry, considered capital-intensive and high-risk, they are even more potentiated. In this context, predictive models can be used to reduce these uncertainties and improve business decisions. Therefore, this article proposes a forecasting method for Petrobras' refining costs and lifting cost, based on time series, through recurrent neural networks. To meet this purpose, the model was trained and tested on a sample containing 72 quarterly observations of each cost. Then, the accuracy of the model was evaluated using graphs and with the aid of the indicators of the mean absolute percentage error [MAPE] and the root mean square error [RMSE]. The results showed that the proposed model was able to provide reasonable estimates for the costs studied, indicating that it can serve as forecasts in the company's decision-making processes.

KEYWORDS. Refining Costs and Lifting Cost. Time Series Forecasting, Recurrent Neural Networks.

Paper topics: OR in the Energy, Oil and Gas Area; Computational Intelligence.

DOI: 10.59254/sbpo-2023-174772
http://live-proceedings.pantheonsite.io/p/174772

1. Introdução

No Brasil, o cenário petrolífero tem como destaque a Petrobras. Segundo os dados da Agência Nacional do Petróleo, Gás Natural e Biocombustíveis (ANP) [ANP 2022], a estatal brasileira respondeu no exercício de 2021 como operadora por mais de 93% da produção de petróleo e mais de 91% da produção de gás natural no território brasileiro. Ainda de acordo com a ANP, em 2021, o parque de refino brasileiro contava com 18 refinarias de petróleo, com capacidade para processar 2,4 milhões de barris/ dia. Doze dessas refinarias pertencem à Petrobras e respondem por 82,3% da capacidade total.

Conforme a Lei nº 9.478/1997 [BRASIL 1997], a indústria de petróleo pode ser definida como um conjunto de atividades econômicas relacionadas com exploração, desenvolvimento, produção, refino, processamento, transporte, importação e exportação de petróleo, gás natural, outros hidrocarbonetos fluidos e seus derivados. Esse setor, segundo Dos Santos et al. [2011], é considerado de intenso capital, altos riscos e longos prazos de maturação de seus investimentos.

Nesse contexto, escolhas erradas podem majorar ainda mais as suas perdas. Essas falhas, na maioria das vezes, estão relacionadas a incertezas. Bugshan [2020] aponta que a volatilidade expõe as indústrias relacionadas ao petróleo a uma incerteza maior do que outras indústrias. Ilyas [2021] ressalta que a incerteza influencia o valor de um projeto de investimento específico e o valor intrínseco da empresa, e por consequência, a riqueza dos acionistas. Segundo o autor, é amplamente discutido na literatura as incertezas que afetam as decisões de investimentos, o autor cita, por exemplo, as incertezas, política, econômica, macro, de custo de entrada e de custo de produção.

Basicamente, as empresas de óleo e gás possuem dois tipos de custos relacionados a produção. O custo de extração - são todos os custos operacionais após a fase de exploração e perfuração de poços, em inglês, é chamado de *lifting cost* -, e o custo de refino - que são todos os custos associados à transformação do petróleo em combustíveis e outros derivados [CBIE 2019], [Petrobras 2023].

Martinez [2001] afirma que é imprescindível ter um modelo de previsão de custos para o planejamento estratégico e representa uma opção de desenvolvimento para a indústria de petróleo e gás. Assim, estudos que possam ajudar a prever esses custos podem ser valiosas ferramentas para uma maior previsibilidade dos seus resultados; para contribuir na confecção de orçamentos; para apoiar o crescimento empresarial e fundamentalmente auxiliar os gestores a serem mais assertivos no processo decisório.

A relevância desses custos no planejamento, controle e tomadas de decisões da Petrobras pode ser notado no seu Plano Estratégico 2023-2027 [Petrobras 2022]. Nesse documento, a empresa projeta que nesse período seu custo de extração vai se manter competitivo e ficará no patamar de 5,5 dólares por barril de óleo equivalente [US$/boe]. No que diz respeito ao custo de refino, o plano não projeta um valor para o quinquênio 2023-2027, mas destaca investimentos para estar entre os melhores refinadores do mundo em eficiência e desempenho operacional. A análises desses dois custos também faz parte do relatório sobre o desempenho financeiro da empresa divulgado ao mercado ao final de cada trimestre.

Este artigo está organizado da seguinte forma, inicia-se por esta introdução e, em seguida, a seção 2 apresenta a descrição do problema, em que são expostos a importância de um modelo de previsão, bem como o porquê da utilização das redes neurais artificiais para tal finalidade. Depois, na seção 3, é apresentada a fundamentação teórica, na qual é realizada uma breve revisão dos modelos de previsão e a lógica das redes neurais. Na seção 4 é apresentada a metodologia usada no trabalho, seguida pela seção 5, apresentação e discussão dos resultados, na qual são mostrados e discutidos os resultados das previsões realizadas assim como os seus erros. E por fim, na seção 5, são feitas as considerações finais, seguidas pelas referências bibliográficas.

DOI: 10.59254/sbpo-2023-174772
http://live-proceedings.pantheonsite.io/p/174772

1. Descrição do Problema

Conforme mencionado, a indústria petrolífera é rodeada de grandes incertezas. Nesse cenário, a avaliação de seus custos é fundamental para o sucesso dessas companhias, uma vez que o seu principal produto – o petróleo - é uma *commoditie* e seu preço é determinado pelo mercado e influenciado por diversos fatores que, na maioria das situações, essas empresas não conseguem controlar.

De acordo com o Instituto Brasileiro de Petróleo e Gás (IPB) [PBP 2023], os custos das atividades da indústria petrolífera variaram ao longo do tempo. Ainda de acordo com o Instituto, os custos de exploração e produção, que em paralelo à elevação dos preços do petróleo, se elevaram no início da última década, mas devido a ganhos operacionais vêm caindo nos últimos anos.

Diante desse contexto, o principal desafio desse estudo é buscar capturar essas mudanças na movimentação dos dados e buscar predizer da melhor forma possível os seus valores futuros. Para isso, é imprescindível um modelo adequado de previsão. Martínez [2001] aponta que métricas incorretas de estimativas podem levar os gestores a tomarem decisões erradas.

Nesse sentido, este trabalho se propõe a criar um modelo para estimar os valores futuros dos custos de extração e refino da Petrobras, baseado em séries temporais, por meio da utilização de redes neurais artificiais. A aplicação das redes neurais surge da necessidade de criar um modelo com os menores erros possíveis.

Com o desenvolvimento da inteligência artificial e tecnologia de computador, os algoritmos de aprendizado foram gradualmente introduzidos na previsão de dados e que com o tempo as Redes Neurais Recorrentes e suas variantes foram as técnicas mais aplicadas para estimativas de séries temporais. Os resultados mostraram que com o advindo dessas novas tecnologias, a precisão da previsão foi significativamente melhorada [Sagheer e Kotb 2019], [Lim at al. 2021], [Wu at al. 2022].

3. Fundamentação Teórica

Séries temporais ou *time series*, como o termo em inglês é conhecido, pode ser definido como um conjunto de observações sequenciadas ao logo do tempo que possuem relações entre si. Nielsen [2021] salienta que os dados de séries temporais abrangem um grande leque de áreas de conhecimento e de variados casos de uso. Segundo o autor, qualquer dado que tenha um eixo ordenado pode ser analisado por meio de método de séries temporais.

Existem na literatura uma variada gama de modelos para análises de séries temporais. A previsão de séries temporais é a tarefa de prever valores futuros de uma determinada sequência usando dados históricos [Sagheer e Kotb 2019]. Nielsen [2021] relata que apesar de grande importância do tema, somente na década de 70 é que surgiu o método Box-Jenkins que apresentou o Modelo Autorregressivo, Integrado e de Média Móvel (ARIMA), considerado como uma contribuição fundamental para a análise de séries temporais, esse método foi apresentado pela primeira vez num livro de estatística intitulado *Time series Analysis: Forecast na control.*

O modelo ARIMA é um dos métodos mais amplamente aplicados na previsão de séries temporais [Munin at al. 2023]. Além do ARIMA, outros modelos também ganharam importância, a exemplo do SARIMA que é uma extensão do modelo ARIMA original com componentes sazonais, o modelo de suavização, o modelo de suavização exponencial de Holt-Winters, entre outros.

Ao longo dos últimos anos, o aumento da capacidade computacional possibilitou, entre outras coisas, o desenvolvimento cada vez mais de técnicas sofisticadas de previsão [Wu at al. 2022]. Munin at al. [2023] argumenta que os modelos baseados em Inteligência Artificial (IA) têm atraído cada vez mais interesse por melhor capturar as características não lineares das séries temporais.

Surgida nesse contexto, as redes neurais artificiais (RNAs) é um modelo matemático baseado em neurônios biológicos. As contribuições pioneiras para a área de RNA foram as de McCulloch e Pitts [1943], que introduziram a ideia de redes neurais como máquinas computacionais. Em 1958, Frank Rosenblatt, propôs uma estrutura em RNA usadas atualmente [Kim at al. 2023], [Pasquoto 2010].

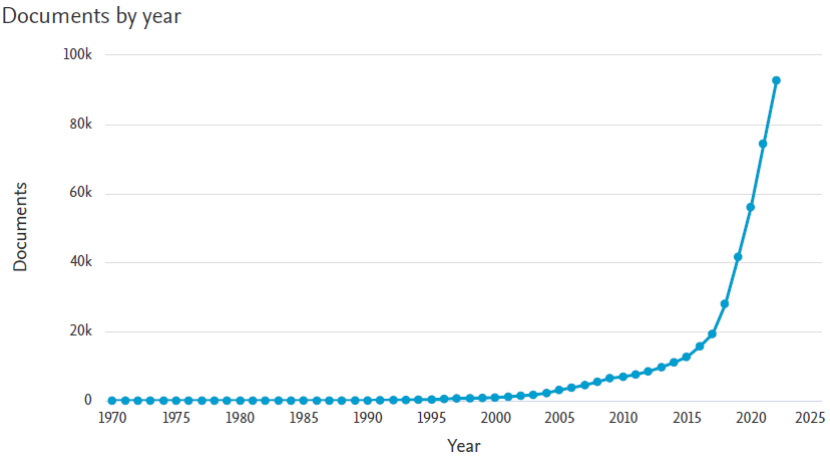
DOI: 10.59254/sbpo-2023-174772
http://live-proceedings.pantheonsite.io/p/174772

As RNAs são, portanto, modelos de aprendizado de máquina não lineares que consistem em camada de entrada, camada de saída e camadas intermediárias/ocultas. As camadas são compostas por neurônios interconectados que por intermédio de funções matemáticas calculam os valores das saídas [Demir at al. 2021].

Duas pesquisas bibliográficas foram realizadas neste trabalho para verificar o uso das redes neurais na previsão de séries temporais. A primeira foi feita na base de dados *ScienceDirect* com a busca nos últimos 5 anos nas áreas de *Energy, Economics, Econometrics and Finance e Business, Management and Accounting* da frase "time series forecasting" no título de artigos revisados e considerados de livre acesso na plataforma. Os resultados mostraram que dos 14 documentos encontrados, 8 utilizaram de alguma forma redes neurais artificiais na previsão de seus dados, descartando 01 documento que não possuía característica formal de artigo e 01 trabalho de avaliação qualitativa, chegou ao percentual de 66,67% do uso de redes neurais na previsão de séries temporais.

A segunda pesquisa, realizada na base *Scopus* para as strings ("time" AND "series" AND "neural" AND "network" AND "artificial") resultou em 418.994 publicações, onde é possível verificar o crescimento do uso de redes neurais em pesquisas relacionadas a séries temporais ao longo do tempo, conforme pode ser visualizado na Figura 1.

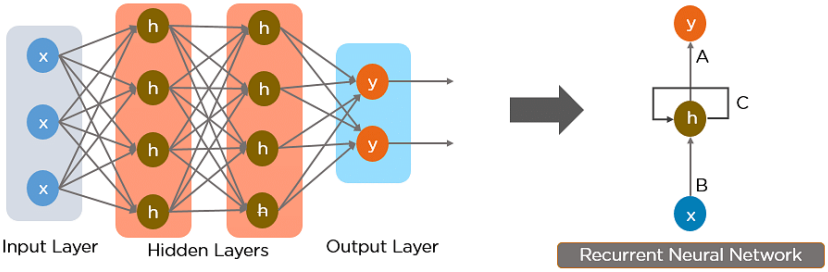
Figura 1: Série temporal de publicações por ano no Scopus



Fonte: biblioteca digital *Scopus*

Em modelos de previsão de séries temporais, a rede neural recorrente ou conforme seu termo em inglês *recurrent neural Network* (RNN) é a variante das redes neurais artificias mais amplamente utilizada [Guo at al. 2021]. A RNN consiste em neurônios artificiais interconectados que aprendem a relação entre a entrada e a saída dos dados ajustando os pesos entre os neurônios através de um algoritmo de retropropagação [Pasquoto 2010], [Amalou at al. 2022], [Kim at al. 2023]. Um exemplo desse tipo de arquitetura é mostrado da Figura 2 abaixo.

Figura 2: Arquitetura de uma rede neural recorrente



DOI: 10.59254/sbpo-2023-174772
http://live-proceedings.pantheonsite.io/p/174772

Fonte: https://www.simplilearn.com/tutorials/deep-learning-tutorial/rnn

Nesse modelo, “x” são os valores da camada de entrada, “h” é cada neurônio dentro das camadas ocultas e “y” são os valores da camada de saída. A, B e C são os parâmetros de rede usados para melhorar a saída do modelo. Em qualquer momento t, a entrada atual é uma combinação da entrada em x(t) e x(t-1). A saída a qualquer momento é trazida de volta para a rede para melhorar a saída.

Dentro das RNNs, cada passo em direção a um resultado é chamado de época, *epoch* em inglês. Assim, o *epoch error* é o erro do modelo em cada passagem dos dados. A taxa de aprendizado (*learning rate*) indica a que ritmo os pesos são atualizados.

Uma variada gama de estudos anteriores se dedicaram a fazerem previsões usando as redes neurais artificiais na área de energia. Sagheer e Kotb [2019], por exemplo, propuseram um modelo de previsão de séries temporais utilizando uma arquitetura profunda de rede neural recorrente do tipo *Long-Short Term Memory* (LSTM), a métrica foi aplicada na previsão de produção de petróleo e se mostrou mais assertiva do que os outros modelos convencionais. Ruiz at al. [2016] por sua vez, se propuseram a fornecer uma metodologia para prever o consumo futuro de energia usando métodos das RNNs. Os resultados do trabalho revelaram que as redes neurais são adequadas para realização de previsão de consumo de energia.

4. Metodologia

Esse trabalho pode ser classificado como uma pesquisa aplicada que possui uma abordagem predominantemente quantitativa, de cunho descritivo, que se utiliza das ferramentas da análise de dados com algoritmos computacionais para analisar e apresentar os resultados encontrados.

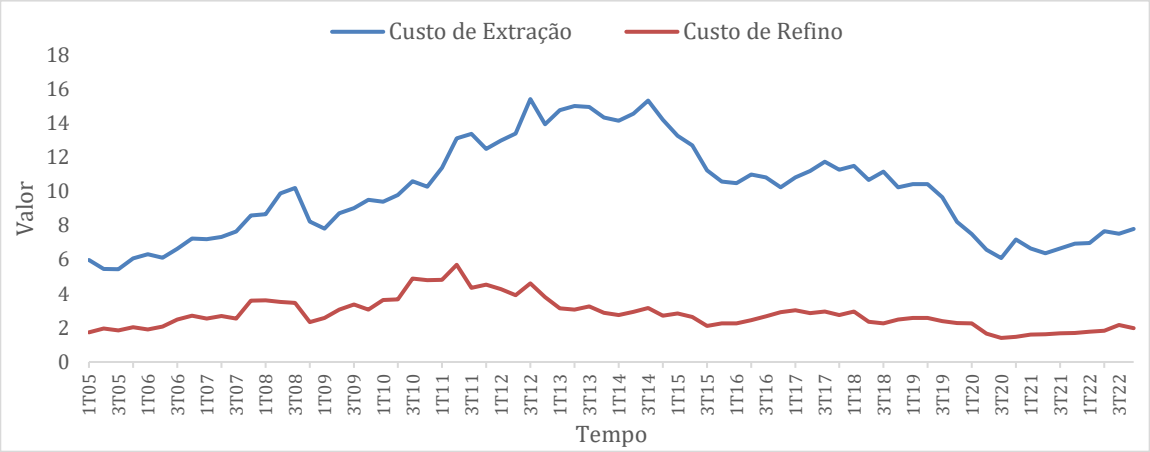
A população da pesquisa são todos os valores dos custos de refino e de extração de petróleo da Petrobras ao longo do período de sua existência. Os valores da amostra selecionados foram todos aqueles constantes nos documentos denominados “Release de Resultados em R$” disponíveis no site de relação com investidores da Petrobras, cujo o endereço eletrônico é: https://www.investidorpetrobras.com.br/resultados-e-comunicados/central-de-resultados/. Nesse site, os custos de refino e os de extração constam a partir do ano de 2006, sendo publicados trimestralmente.

Os documentos publicados contêm a comparação dos custos correntes com os períodos anteriores, em vista disso foi possível coletar os dados desde o primeiro trimestre de 2005. Desse modo, a composição da amostra foi de 72 observações trimestrais para cada custo referente ao período do primeiro trimestre de 2005 (1T05) até o quarto trimestre de 2022 (4T22). Foram escolhidos para a amostra os custos de refino no brasil e os custos de extração de petróleo sem as participações governamentais. O custo de extração tem seu valor expresso em dólar por barril de óleo equivalente (US$/boe), já o custo de refino tem seu valor expresso em dólar por barril de petróleo (US$/barril).

Os dados foram tabulados em planilha eletrônica e sua evolução pode ser visualizada na Figura 3.

Figura 3: Evolução de custos de refino e extração da Petrobras por trimestre.

DOI: 10.59254/sbpo-2023-174772
http://live-proceedings.pantheonsite.io/p/174772

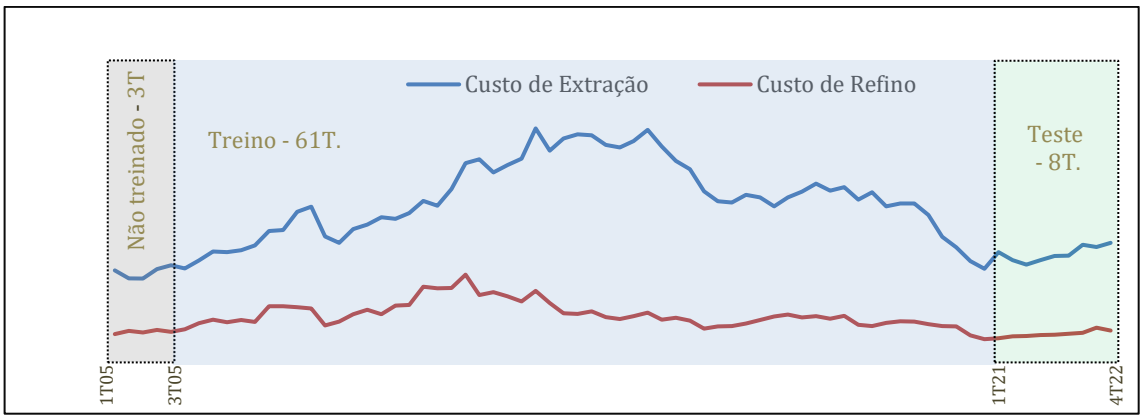


Fonte: elaborado pelos autores a partir dos dados da pesquisa

O tratamento, a análise dos dados, o treinamento e o teste do modelo foram efetuados utilizando o software R v. 4.1.2 [R Core Team 2021], e as bibliotecas "rnn" [Quast 2016], "dplyr" [Wickham at. al 2023], "tidyverse" [Wickham at. al 2019], "readxl" [Wickham e Bryan 2023], "nnfor" [Kourentzes 2022] e "MLmetrics" [Yan 2016]. A técnica de previsão usada foi a de redes neurais recorrentes por meio das funções “trainr”, para o treinamento da RNN, e “predictr” para realizar as saídas de previsões do modelo. Ambas as funções são da biblioteca “rnn”.

A proposta do modelo é utilizar os três últimos períodos como parâmetro para estimar os custos do trimestre seguinte. Diante disso, os dados foram treinados e testados a partir do 3T05, haja vista que as 3 primeiras observações coletadas não possuíam 3 observações anteriores. Porém elas foram utilizadas como base na previsão do 3T05. Após isso, partiu-se para a separação da amostra em treinamento e teste. Optou-se por usar 61 observações (61T) na fase de treinamento e 8 observações (8T) para a fase de teste para as duas modalidades de custos, esse processo pode ser mais bem compreendido na Figura 4:

Figura 4: Separação dos dados de treinamento e de teste da amostra.



Fonte: elaborado pelos autores a partir dos dados da pesquisa

O modelo de treinamento para previsão de séries temporais da rede neural recorrente empregado nesse trabalho pode ser descrito seguindo Ruiz et al. [2016], como se segue na Eq. 1.

$$y(t) = f(y(t-1), y(t-2), y(t-3)) + \varepsilon(t) \quad (1)$$

DOI: 10.59254/sbpo-2023-174772
http://live-proceedings.pantheonsite.io/p/174772

Onde: y (t) = valor previsto; y (t − 1) = valor do primeiro trimestre anterior ao valor previsto; y (t − 2) = valor do segundo trimestre anterior ao valor previsto; y (t − 3)) = valor do terceiro trimestre anterior ao valor previsto e ε (t) = termo de erro.

Souza at al. [2022] destaca que a normalização dos dados em redes neurais decorre fundamentalmente de razões empíricas, pois garante uma estabilidade e convergência mais rápida e com uma maior probabilidade de obter resultados mais precisos. Os dados foram normalizados fazendo uso da função "linscale" da biblioteca "nnfor". Essa função normaliza as observações utilizando a metodologia denominada min-max, que mapeia cada elemento no intervalo (0, 1) e é dada de acordo com a Eq. 2.

$$Xscaled = \frac{X - Xmin}{Xmax - Xmin} \tag{2}$$

Onde: Xscaled= valor normalizado; X = valor do dado da amostra; Xmin = valor mínimo dos dados da amostra e Xmax = valor máximo dos dados da amostra.

Uma das características das funções "trainr" e "predictr" é que os dados estejam em estrutura de "array". No software R, um *array* é definido utilizando a função array(). No treinamento do modelo, utilizou-se uma taxa de aprendizagem ou (learningrate) de 0,5 e o número de épocas ou (numepochs) igual a 100, o número de neurônios ou (hidden_dim) utilizados na foi o padrão da própria função "trainr" que é igual a 10. A função de erro de cada época (epoch error) do algorítmico foi usada para o aperfeiçoamento dos modelos. Outros parâmetros foram testados, mas sem nenhuma melhora significativa nos resultados.

Seguindo Guo at al. [2021], para avaliação e comparação das previsões realizadas nos testes dos dois custos, foram usados como métricas de erro o MAPE e o RMSE, as fórmulas dos cálculos dos dois indicadores são mostradas as nas Eqs. 3 e 4.

$$MAPE = \frac{1}{n} \sum_{i=1}^{n} \frac{|y_i - \hat{y}_i|}{y_i} \text{ X } 100\% \tag{3}$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2} \tag{4}$$

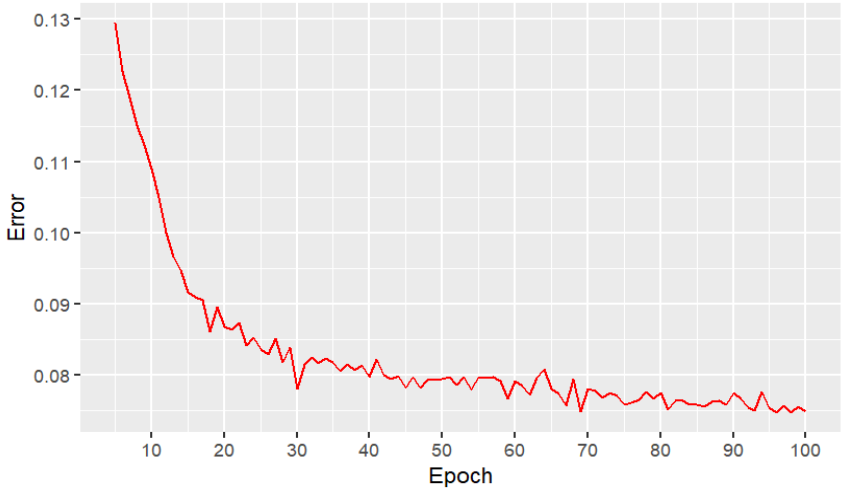
Onde: y_i = valor real de cada dado da amostra; $\hat{y}_i$ = valor previsto de cada dado da amostra e n = é o número de observações da amostra.

5. Apresentação e Discussão dos Resultados

O gráfico apresentado na Figura 5 abaixo mostra a queda no termo de erro à medida que vão sendo processadas as épocas da rede neural para a previsão dos custos de refino. Nele, é possível observar que não há muito ganho na redução do erro a partir da época 69. O valor final do erro na época 100 encontrado pelo algoritmo foi de aproximadamente 0.0748.

Figura 5: Curva de erros para a previsão dos custos de refino da Petrobras

DOI: 10.59254/sbpo-2023-174772
http://live-proceedings.pantheonsite.io/p/174772

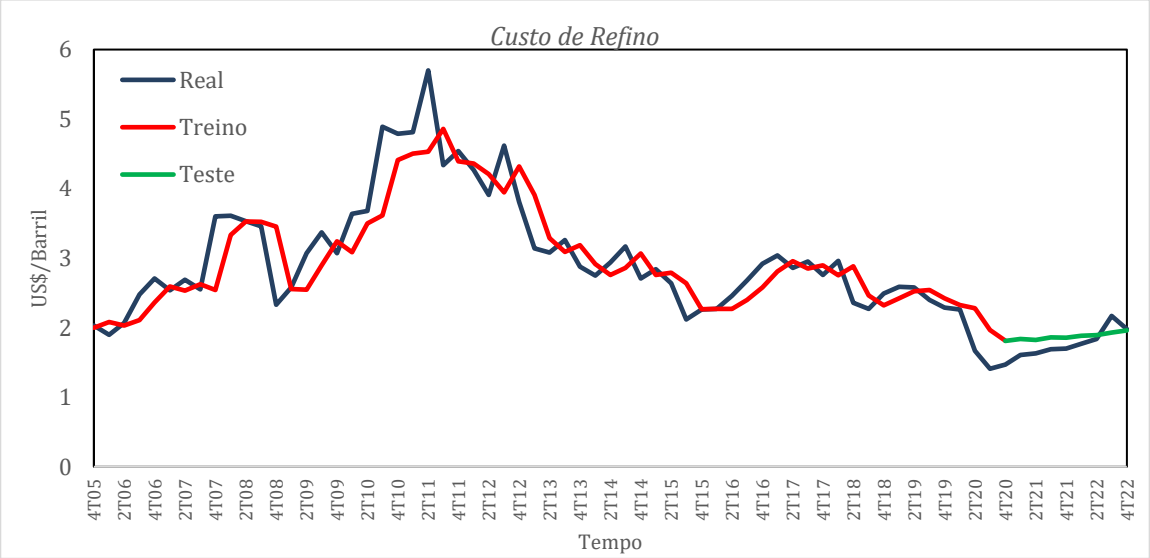


Fonte: elaborado pelos autores a partir dos dados da pesquisa

A Figura 6 a seguir apresenta os resultados reais e os estimados referentes às amostras de treino e de teste para o custo de refino. Inicialmente é possível visualizar os movimentos de elevação e posteriores quedas nos custos de refino da Petrobras, o IPB [2023] aponta que essa queda é decorrente dos ganhos na eficiência operacional.

No que se refere aos valores previstos, tanto na fase de treinamento como na fase de teste, eles são próximos dos custos reais incorridos no refino da Petrobras. Esses achados foram condizentes com os resultados encontrados por Sagheer e Kotb [2019] e Ruiz at al. [2016] nos quais indicam que as redes neurais são instrumentos eficazes em previsões de séries temporais.

Figura 6 – Previsão do custo de refino da Petrobras

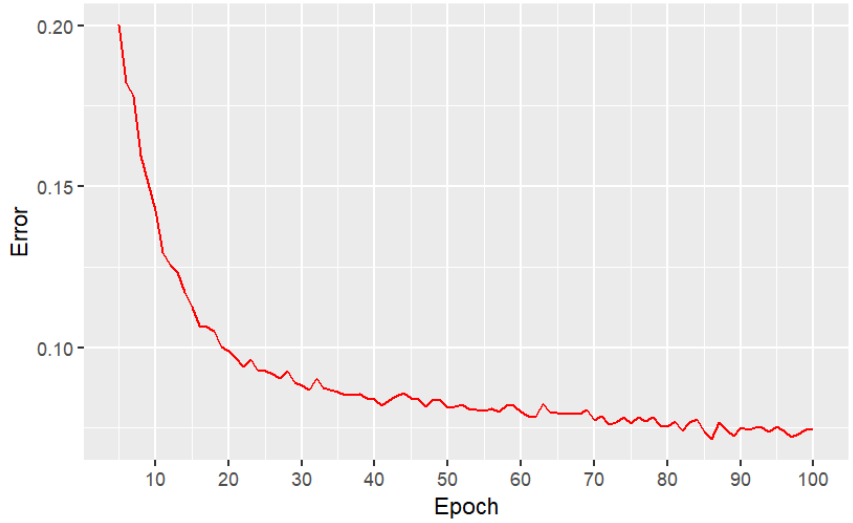


Fonte: elaborado pelos autores a partir dos dados da pesquisa

A Figura 7 demonstra a redução do erro ao passo que vão aumentando o número de épocas da fase de treinamento do custo de extração da Petrobras. Observa-se que a partir da época 86 o valor do erro já não apresenta queda significativa. O erro na época 100 foi de aproximadamente 0,074.

Figura 7: Curva de erros para a previsão dos custos de extração da Petrobras

DOI: 10.59254/sbpo-2023-174772
http://live-proceedings.pantheonsite.io/p/174772

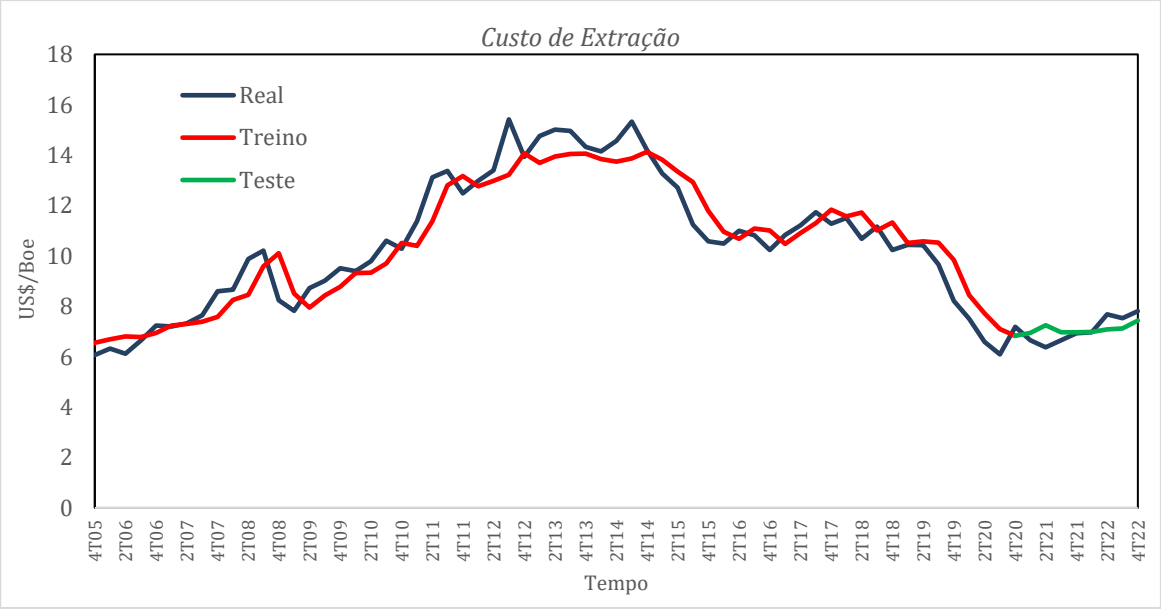


Fonte: elaborado pelos autores a partir dos dados da pesquisa

Na Figura 8 é apresentado o gráfico que compara os resultados estimados com os dados do treino e do teste com os dados reais do custo de extração da Petrobras. Semelhante aos valores dos custos de refino, é possível perceber que inicialmente os custos de extração cresceram e posteriormente tiveram um movimento de queda. Esse recuo, segundo o Centro Brasileiro de Infraestrutura (CBIE) [2020], se deu devido aos investimentos em tecnologia de ponta para abaixar o custo de extração em águas profundas, principalmente nos enormes campos do Pré-sal, que são altamente produtivos em petróleo de boa qualidade.

Em relação às previsões, verifica-se que o modelo foi capaz de estimar com uma certa margem de segurança os custos associados à extração de petróleo da Petrobras. Novamente verifica-se que os resultados foram ao encontro com as conclusões dos estudos de Sagheer e Kotb [2019] e Ruiz at al. [2016] no que se refere adequabilidade do uso de redes neurais para predições. E mais uma vez reforça a hipótese de que essa técnica pode ajudar a realizar razoáveis previsões desses custos.

Figura 8 – Previsão do custo de extração da Petrobras



DOI: 10.59254/sbpo-2023-174772
http://live-proceedings.pantheonsite.io/p/174772

Fonte: elaborado pelos autores a partir dos dados da pesquisa

A Tabela 1, apresentada a seguir, é um resumo do MAPE e RMSE da fase de treinamento e da fase de teste dos dois custos modelados no algoritmo. Pode-se verificar que tanto para o custo de refino quanto para o custo de extração, o MAPE e o RMSE são menores para os dados de teste do que para os dados de treino. O MAPE teve uma redução de 20,63% para o custo de refino e de 22,16% para o custo de extração, já no RMSE, a redução foi de 61,90% para o custo de refino e de 46,43% para o custo de extração.

Uma possível explicação para essas quedas nos indicadores de erro pode estar relacionada a maior estabilidade nesses custos nos períodos utilizados para a validação dos modelos, conforme pode ser visualizado nas Figuras 6 e 8.

Tabela 1. MAPE e RMSE das previsões

Custo	Custo de Refino		Custo de Extração	
Fase	MAPE (%)	RMSE (US$/Boe)	MAPE (%)	RMSE (US$/Boe)
Treino	10,52	0,42	6,68	0,84
Teste	8,35	0,16	5,20	0,45

Fonte: elaborado pelos autores a partir dos dados da pesquisa

6. Considerações Finais

Este estudo demonstrou que o uso de redes neurais recorrentes pode ser útil na previsão dos custos de refino e de extração da Petrobras. Por intermédio dos resultados obtidos na pesquisa, foi possível verificar que o modelo proposto foi capaz de oferecer estimativas razoáveis paras os custos futuros da empresa.

O modelo conseguiu capturar satisfatoriamente os movimentos de subida e queda nos valores da amostra. Além disso, na fase de teste, que estava num momento de razoável estabilidade, o método realizou previsões com menores erros.

Apesar de perceber os benefícios da técnica, é importante ressaltar que ela não deve ser usada de forma isolada, mas sim de forma complementar. Dessa maneira, entende-se que os resultados obtidos terão uma maior confiabilidade.

Espera-se que esse trabalho estimule outros pesquisadores a desenvolverem melhores modelos de previsão de custos, não apenas no setor de petróleo/energia, como em outros setores da economia.

Recomenda-se que trabalhos futuros possam estudar e desenvolver modelos melhores com a inclusão de variáveis exógenas aos dados pesquisados nesse artigo. É primordial também estudos que possam fazer comparações com outras técnicas de previsão de séries temporais e outras variantes das redes neurais.

Referências

Amalou, I., Mouhni, N., e Abdali, A. (2022). Multivariate time series prediction by RNN architectures for energy consumption forecasting. *Energy Reports*, *8*, 1084-1091.

ANP (2022). Anuário Estatístico Brasileiro do Petróleo, Gás Natural e Biocombustíveis 2022. Agência Nacional do Petróleo, Gás Natural e Biocombustíveis. Web page: https://www.gov.br/anp/pt-br/centrais-de-conteudo/publicacoes/anuario-estatistico/. Acessado: 2023-03-23.

Box, G. E., e Jenkins, G. M. (1976). Time series analysis. Forecasting and control. Holden- Day Series in Time Series Analysis. Revised ed., San Francisco: Holden-Day.

DOI: 10.59254/sbpo-2023-174772
http://live-proceedings.pantheonsite.io/p/174772

BRASIL (1997). Lei nº 9.478, 06 de agosto de 1997. Dispõe sobre a política energética nacional, as atividades relativas ao monopólio do petróleo, institui o Conselho Nacional de Política Energética e a Agência Nacional do Petróleo e dá outras providências. Diário Oficial da União. Brasília, DF, de 07 de agosto de 1997.

Bugshan, A. (2022). Oil price volatility and corporate cash holding. *Journal of Commodity Markets*, *28*, 100237.

CBIE (2019). O que são os custos de extração do Petróleo? Web page: https://cbie.com.br/o-que-sao-os-custos-de-extracao-do-petroleo/. Acessado: 2023-04-02.

CBIE (2020). Quais as diferenças entre os tipos de Produção de Petróleo? Web page: https://cbie.com.br/quais-as-diferencas-entre-os-tipos-de-producao-de-petroleo/. Acessado: 2023-04-01.

Demir, S., Mincev, K., Kok, K., e Paterakis, N. G. (2021). Data augmentation for time series regression: Applying transformations, autoencoders and adversarial networks to electricity price forecasting. *Applied Energy*, *304*, 117695.

Dos Santos, O. M., dos Santos, A., e da Silva, P. D. A. (2011). Reconhecimento de perdas para redução ao valor recuperável de ativos: impairment em ativos de exploração e produção de petróleo. *BBR-Brazilian Business Review*, *8*(2), 68-95.

Guo, X., Gao, Y., Li, Y., Zheng, D., e Shan, D. (2021). Short-term household load forecasting based on Long-and Short-term Time-series network. *Energy Reports*, *7*, 58-64.

Ilyas, M., Khan, A., Nadeem, M., e Suleman, M. T. (2021). Economic policy uncertainty, oil price shocks and corporate investment: evidence from the oil industry. *Energy Economics*, *97*, 105193.

Instituto Brasileiro de Petróleo e Gás (IPB) (2023). Evolução dos custos médios de extração e refino da Petrobras. Web page: https://www.ibp.org.br/observatorio-do-setor/snapshots/evolucao-dos-custos-medios-de-extracao-e-refino-da-petrobras/ Acessado: 2023-04-08.

Kim, H., Park, S., e Kim, S. (2023). Time-series clustering and forecasting household electricity demand using smart meter data. *Energy Reports*, *9*, 4111-4121.

Kourentzes, N. (2022). nnfor: Time Series Forecasting with Neural Networks. R package version 0.9.8. https://CRAN.R-project.org/package=nnfor

Lim, B., Arık, S. Ö., Loeff, N., e Pfister, T. (2021). Temporal fusion transformers for interpretable multi-horizon time series forecasting. *International Journal of Forecasting*, *37*(4), 1748-1764.

Martinez, R. E. (2001). Forecast techniques for lifting cost in gas and oil onshore fields. In *SPE Latin American and Caribbean Petroleum Engineering Conference*. OnePetro.

McCulloch, W.S., Pitts, W., (1943). A logical calculus of the ideas immanent in nervous activity. Bull. Math. Biophys. 5 (4), 115–133.

Munim, Z. H., Fiskin, C. S., Nepal, B., e Chowdhury, M. M. H. (2023). Forecasting container throughput of major Asian ports using the Prophet and hybrid time series models. *The Asian Journal of Shipping and Logistics*.

DOI: 10.59254/sbpo-2023-174772
http://live-proceedings.pantheonsite.io/p/174772

Nielsen, A. (2019). Practical time series analysis: Prediction with statistics and machine learning. [S.l.]: O'Reilly Media.

Pasquotto, J. L. D. (2010). *Previsão de séries temporais no varejo brasileiro: uma investigação comparativa da aplicação de redes neurais recorrentes de Elman* (Doctoral dissertation, Universidade de São Paulo).

Petrobras (2022). Plano Estratégico (2023-2027). Web page: https://petrobras.com.br/pt/quem-somos/plano-estrategico/. Acessado: 2023-03-23.

Quast, B. A. (2016). rnn: a Recurrent Neural Network in R

R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Rosenblatt, F., (1958). The perceptron: a probabilistic model for information storage and organization in the brain. Psychol. Rev. 65 (6), 386.

Ruiz, L. G. B., Cuéllar, M. P., Calvo-Flores, M. D., e Jiménez, M. D. C. P. (2016). An application of non-linear autoregressive neural networks to predict energy consumption in public buildings. *Energies*, *9*(9), 684.

Sagheer, A., e Kotb, M. (2019). Time series forecasting of petroleum production using deep LSTM recurrent networks. Neurocomputing, 323, 203-213.

Sousa, M., Tomé, A. M., e Moreira, J. (2022). Long-term forecasting of hourly retail customer flow on intermittent time series with multiple seasonality. *Data Science and Management*, *5*(3), 137-148.

Wickham, H., Bryan, J. (2023). readxl: Read Excel Files. R package version 1.4.2. https://CRAN.R-project.org/package=readxl

Wickham, H., François, R., Henry, L., Müller, K., Vaughan, D. (2023). dplyr: A Grammar of Data Manipulation. R package version 1.1.0. https://CRAN.R-project.org/package=dplyr

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T. L., Miller, E., Bache, S. M., Müller, K., Ooms, J., Robinson, D., Seidel, D. P., Spinu, V., Takahashi, K., Vaughan, D., Wilke, C., Woo, K., Yutani, H. (2019). "Welcome to the tidyverse." _Journal of Open Source Software_, *4*(43), 1686. doi: 10.21105/joss.01686 (URL: https://doi.org/10.21105/joss.01686).

Wu, Z., Mu, Y., Deng, S., e Li, Y. (2022). Spatial–temporal short-term load forecasting framework via K-shape time series clustering method and graph convolutional networks. *Energy Reports*, *8*, 8752-8766.

Yan, Y. (2016). MLmetrics: Machine Learning Evaluation Metrics. R package version 1.1.1. https://CRAN.R-project.org/package=MLmetrics

DOI: 10.59254/sbpo-2023-174772
http://live-proceedings.pantheonsite.io/p/174772