

Avaliação da Qualidade Preditiva de Modelos Baseado em Intervalo de Previsão

Carlos Luis Euribe Solorzano
Faculdade de Tecnologia
Universidade Estadual de Campinas
São Paulo, Brasil
c246894@dac.unicamp.br

Taylon Luan Congio Martins
Faculdade de Engenharia Elétrica e Computação
Universidade Estadual de Campinas
São Paulo, Brasil
t177379@dac.unicamp.br

Resumo—De forma geral, os modelos preditivos apresentam uma porcentagem de erro em suas previsões que podem ser detectadas com indicadores globais como o Erro Médio Quadrático (MSE) e a Raiz do Erro Médio Quadrático (RMSE), mas não permitem analisar cada uma das previsões e conhecer exatamente quais ficam fora dos limites de aceitabilidade e determinar em que nível de confiança as previsões do modelo são aceitáveis, portanto, é importante ter alternativas complementares que possam dar uma descrição pontual delas. Como solução se propõe uma análise utilizando Intervalos de Predição (IP) de 90%, 95% e 98%. Como metodologia se utilizará métodos de Estatística Descritiva (análise de normalidade), e Inferência Estatística (análise de normalidade, amostragem, intervalos de confiança, predição e teste de hipóteses). O estimador utilizado é baseado na diferença pontual entre o valor real e a predição obtida. Os dados de previsões são gerados por dois modelos, um de regressão linear e um de rede neural sobre volumes de produção de petróleo utilizando uma série temporal. Os resultados foram para o modelo de regressão linear: 90% (4.7%), 95% (3.6%), 98% (2.5%), e para a rede neural: 90% (8.6%), 95% (6.4%), 98% (5.1%). Conclui-se que baseado nos resultados estatísticos o modelo de regressão linear com intervalos de predição de 90% e 95% são os mais precisos na predição pontual e que a metodologia é eficaz na detecção das previsões que ficam fora dos limites dos respectivos IP.

Palavras-chave—*Bootstrap, intervalo de predição, intervalo de confiança, série temporal, produção de petróleo.*

I. INTRODUÇÃO

Se por um lado, realmente é pouco provável que um modelo preditivo apresente 100% de precisão nas previsões, sempre tem uma margem de erro. Por outro, um modelo que não apresenta uma precisão de 100% não significa que não seja candidato para a tomada de decisões.

Dentro dessa margem de erro dos modelos há previsões específicas que devem ser identificadas e conhecidas para que se possa ter algum tipo de consideração sobre eles, por exemplo: estabelecer um nível de risco, quão críticos podem ser ou qual poderia ser seu nível de aceitabilidade.

Em outras palavras, a margem de erro num modelo pode apresentar pontualmente 10 previsões, mas com a aplicação e avaliação utilizando o intervalo de predição 5 das 10 podem ser consideráveis aceitáveis de acordo com um intervalo de previsão estabelecido.

Por outra abordagem, os indicadores globais utilizados para medir a precisão como o MSE e RMSE nos dão um indicador resumido sobre todo o modelo. Entretanto, eles não podem fornecer informação específica sobre cada uma das previsões.

Para a solução utilizamos o Intervalo de Previsões, que nos permitirá identificar e conhecer quais são aquelas que ficam dentro e fora dos limites de aceitabilidade estabelecidos por os IP. O valor da contribuição da análise reside em analisar as previsões de toda a série temporal em conjunto e determinar em que IP as previsões do modelo são aceitáveis através de teste de hipóteses para cada IP, onde é avaliada a porcentagem de previsões que ficam fora do respectivo IP contra o nível de significância.

Para construir estes intervalos de previsão, o ponto de partida será utilizar a diferença entre o valor real e o previsto como indicador ou estatístico. Se a distribuição dessas diferenças não for normal, deve-se usar Bootstrap.

Os dados para a análise são obtidos de dois modelos preditivos, um modelo de regressão linear e um baseado em uma rede neural. Os modelos são simplesmente referências para fazer a demonstração da solução proposta.

Os dados a serem utilizados apresentam os volumes de produção de petróleo diário do poço 11 de UNISIM-CEPETRO, cujo conjunto de dados originalmente está composto por 3023 registros e não contém valores nulos, faltantes nem duplicados. Para a padronização dos dados foi utilizada a normalização MinMax (valores entre 0 e 1). As sequências foram criadas com uma janela de tamanho 100 para prever o dia seguinte. Os dados foram divididos em treinamento e teste. A abordagem com a rede neural apresenta um desenvolvimento baseado em aprendizagem profunda (quatro camadas ocultas), [1], [2], no qual o pré-processamento não é necessário [3] sendo o objetivo modelar os dados tal quais são obtidos da produção, no mesmo sentido e para fornecer previsões que possam ser comparadas a mesma abordagem foi aplicado ao modelo de regressão linear. É importante considerar que a rede neural usa uma função de ativação não linear.

II. PRINCIPAIS CONCEITOS

MSE. Erro médio quadrático. Pequenos valores indicam que as previsões foram muito próximas que os valores verdadeiros.

RMSE. Raiz do erro quadrático médio. Define as inconsistências entre os valores reais e os valores previstos. Um valor próximo a zero indica melhor desempenho do modelo.

Teste de Shapiro-Wilk. [4] É um teste para saber se uma distribuição de dados é normal, onde se o p-valor é maior

que o nível de significância os dados apresentam uma distribuição normal. É ideal para amostras de dados pequenas (menores de 50).

Teste de Kolmogorov-Smirnov. [5], [6] É um teste para determinar se uma distribuição de dados é normal, onde se o p-valor é menor que o nível de significância os dados apresentam uma distribuição normal. É ideal para amostras maiores de 50.

Bootstrap. É uma técnica de reamostragem, utilizada para inferir o valor de uma estatística para dados de qualquer distribuição (neste caso, as diferenças entre os valores reais e previstos) e após poder construir intervalos de confiança e predição. [7], [8]

Média amostral. É uma medida estatística de tendência central, é o valor médio de todo o conjunto de dados.

$$\text{media} = \frac{\sum_{i=1}^n x_i}{n}$$

Variância amostral. É uma medida de dispersão dos valores com relação à média amostral.

$$s^2 = \frac{\sum_{i=1}^n (x_i - \text{media})^2}{n - 1}$$

Desvio padrão amostral. É a raiz quadrada da variância amostral.

$$\text{desvio}_{\text{padrao}} = S = \sqrt{s^2}$$

Intervalo de confiança (IC). Para obter os IC foi utilizado o método de percentis. [9], [10].

$$IC_{\text{inf}} = \text{np.percentile}(\text{media}_{\text{amostral}}, \text{confiança})$$

$$IC_{\text{sup}} = \text{np.percentile}(\text{media}_{\text{amostral}}, \text{confiança})$$

Intervalo de predição (IP). Estima um intervalo em que se pode produzir um dado ou estimativa futura em base a dados já obtidos. [11], [12], [13], [14].

O valor crítico de T bilateral se calcula:

n = tamanho da amostra

$$\text{valor}_{\text{critical}} = \text{stats.t.ppf}\left(\left(\frac{1 + \text{confiança}}{2}\right), n - 1\right)$$

$$\text{margem}_{\text{erro}} = \text{valor}_{\text{critical}} \cdot \text{desvio}_{\text{padrao}} \cdot \sqrt{1 + \frac{1}{n}}$$

$$IP_{\text{inf}} = \text{media}_{\text{amostral}} - \text{margem}_{\text{erro}}$$

$$IP_{\text{sup}} = \text{media}_{\text{amostral}} + \text{margem}_{\text{erro}}$$

III. METODOLOGIA

Após do treinamento e teste os indicadores dos modelos são apresentados na Tabela 1, onde pode-se ver que os indicadores da rede neural são ligeiramente melhores.

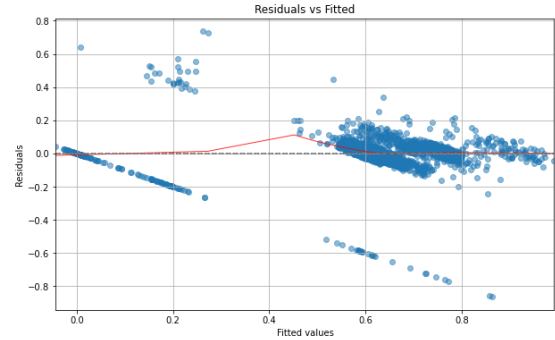
TABELA 1. INDICADORES DOS MODELOS

Modelo	R ²	MAE	MSE	RMSE
Regressão Linear	0.89	0.046	0.009	0.099
Rede Neural MLP	0.97	0.021	0.003	0.050

Para avaliar o modelo de regressão linear utilizamos alguns gráficos [15], [16] como a Figura 1, onde pode-se

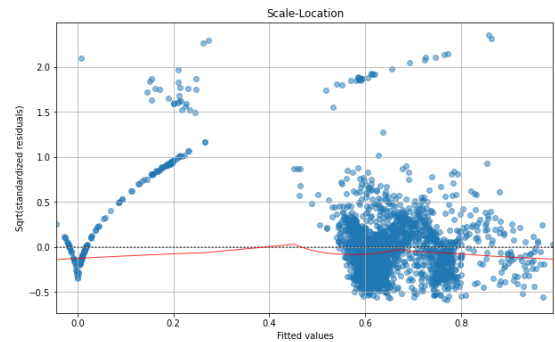
observar que o modelo não consegue capturar bem as não linearidades de acordo com a irregularidade da linha vermelha e alguns padrões em forma de retas verticais inclinadas.

Figura 1. Gráfico Residuals vs Fitted



No caso, da Figura 2, percebe-se que a linha vermelha apresenta algumas irregularidades, mas ainda é horizontal, portanto, pode-se assumir que a homocedasticidade é parcialmente cumprida [17]. As irregularidades da linha vermelha são geradas porque o modelo não captura bem as não linearidades. Quanto ao acúmulo de pontos entre 0.6 e 0.8, mostra que há forte heterocedasticidade.

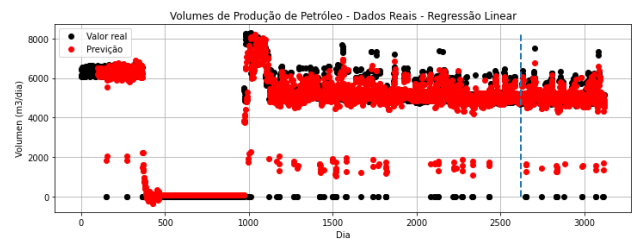
Figura 2. Gráfico Scale-Location



As duas Figuras anteriores amostram que existem algumas não linearidades que não foram capturadas pelo modelo de regressão linear.

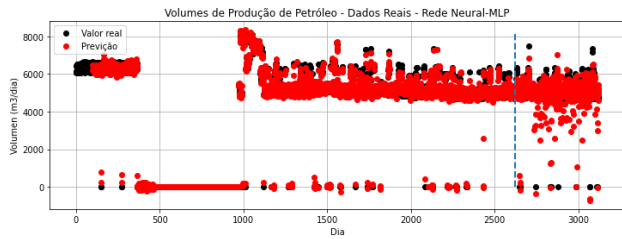
No caso da Figura 3, o modelo de regressão apresenta algumas previsões (pontos vermelhos) não precisas na parte superior e inferior com relação aos dados reais (pontos pretos).

Figura 3. Previsões com Regressão Linear



De acordo com a Figura 4, o modelo com a rede neural apresenta melhor precisão com relação à Figura 3, tanto em cima como embaixo dos dados reais.

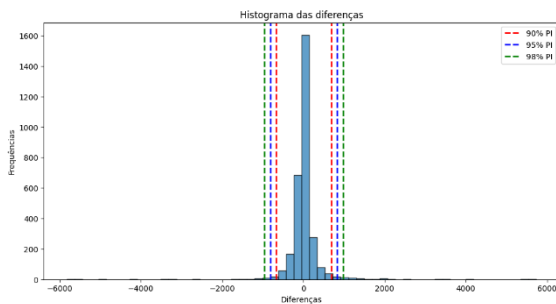
Figura 4. Previsões com Rede Neural MLP



A segunda parte focada na análise estatística dos dados para fazer a avaliação da qualidade das previsões dos modelos. Nosso indicador estatístico para avaliar as previsões são as diferenças entre os valores reais e as previsões.

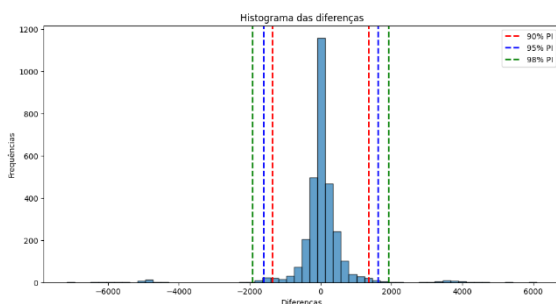
De acordo com a Figura 5, para o modelo com a rede neural, pode-se observar que as maiores quantidades dos dados ficam próxima da média.

Figura 5. Distribuição de Diferenças – Rede Neural



No caso do modelo de regressão linear, os dados também estão próximos da média, mas com uma maior distribuição em volta da média, ou seja, apresenta menor precisão.

Figura 6. Distribuição de Diferenças – Regressão Linear



Primeiro, fazendo uso de Estatística Descritiva, avaliamos a normalidade das diferenças, utilizando gráficos Q-Q Plot.

Figura 7. Gráfico Q-Q Plot de Diferenças – Rede Neural

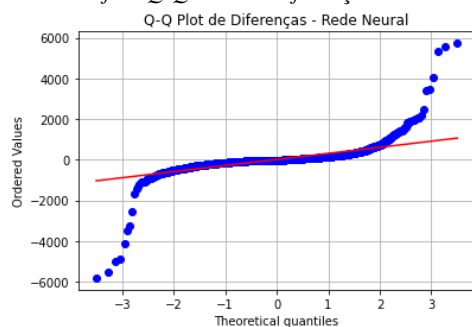
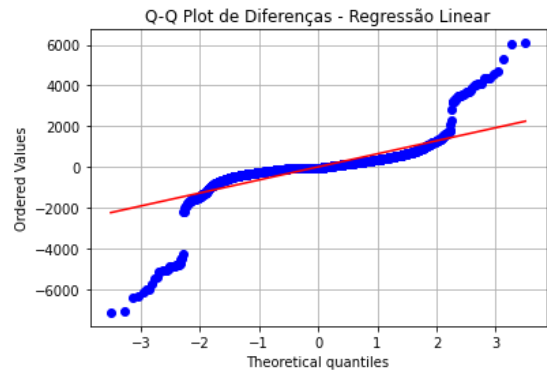


Figura 8. Gráfico Q-Q Plot de Diferenças – Regressão Linear



Segundo, utilizando Inferência Estatística e testes de normalidade [18], aplicamos o teste de Shapiro-Wilk [19] para avaliar a normalidade da distribuição das Diferenças de acordo as seguintes hipóteses:

H0: Os dados de diferenças estão normalmente distribuídos.

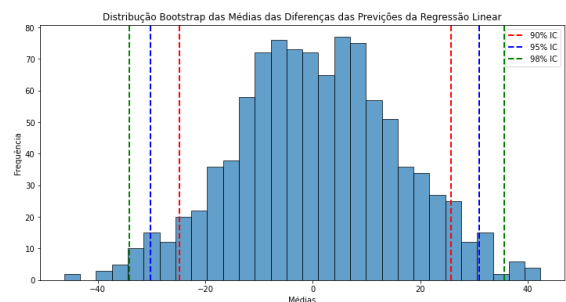
H1: Os dados de diferenças não estão normalmente distribuídos.

Para o modelo de regressão o resultado para o valor estatístico (W) foi de 0.6054. No caso do modelo com rede neural, o valor para o estatístico foi 0.5172.

Com o nível de significância estabelecido em 0.05, obteve-se $p\text{-valor} < 0.05$ para ambos os modelos, portanto não se pode afirmar que os dados apresentam uma distribuição normal, mas existem indícios de que apresentam uma distribuição não normal. No caso do teste de Kolmogorov-Smirnov [20] sobre as mesmas hipóteses de normalidade, os resultados para o modelo de regressão linear foi 0.5494. No caso do modelo com rede neural se obteve 0.5708 para o estatístico. Com $p\text{-valor} < 0.05$ para ambos os modelos, portanto, os dados não apresentam uma distribuição normal.

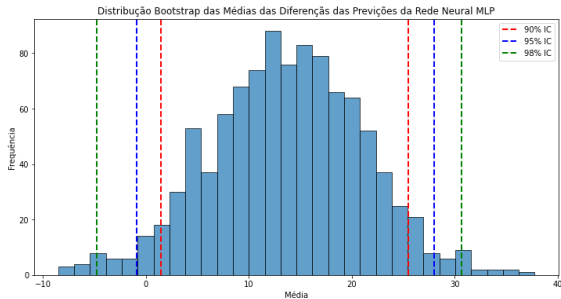
De acordo com a prova de hipóteses da normalidade e os gráficos Q-Q pode-se dizer que a distribuição das diferenças não é normal, portanto, é necessário utilizar amostragem com Bootstrap (Bootstrapping dos resíduos). Para o bootstrap foram realizadas 1000 amostragens com reposição no conjunto de 3023 dados de cada um dos modelos.

Figura 9. Distribuição Bootstrap das Diferenças das Previsões da Regressão Linear



De acordo com a Figura 10, a média das diferenças das amostras apresentam uma distribuição normal, tendo cumprimento do teorema de limite central, onde a distribuição da média amostral tendo a uma distribuição normal para um grande número de amostras. [21]

Figura 10. Distribuição Bootstrap das Diferenças das Previsões da Rede Neural MLP



Posteriormente, obtemos os Intervalos de Confiança (IC) para 90%, 95% e 98% utilizando os percentis. Os limites dos IC são apresentados na Tabela 2.

TABELA 2. INTERVALOS DE CONFIANÇA

Modelo	Intervalo de Confiança	Limite Inferior	Limite Superior
Rede Neural MLP	90%	1.48	25.49
	95%	-0.89	28.01
	98%	-4.76	30.69
Regressão Linear	90%	-24.84	25.74
	95%	-30.21	30.93
	98%	-34.06	35.66

De acordo a teorema do limite central, onde a média da média das amostras converge para uma normal e o tamanho da amostra ($n=3023$) garante a aplicação do t-score no cálculo de Intervalos de Predição (IP). Previamente se devem calcular os valores críticos e após as margens de erros para cada nível de confiança (90%, 95% e 98%). Os limites dos IP são apresentados na Tabela 3.

TABELA 3. INTERVALOS DE PREVISÃO

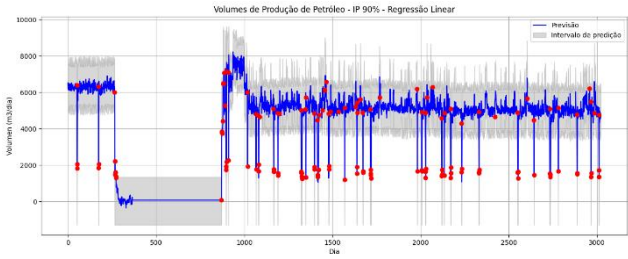
Modelo	Intervalo de Previsão	Limite Inferior	Limite Superior
Rede Neural MLP	90%	-673.72	700.22
	95%	-805.41	831.90
	98%	-958.57	985.06
Regressão Linear	90%	-1356.31	1358.77
	95%	-1616.53	1618.99
	98%	-1919.19	1921.65

REGRESSÃO LINEAR

Foram construídos gráficos da série temporal para os intervalos de predição (IP) de 90%, 95% e 98%.

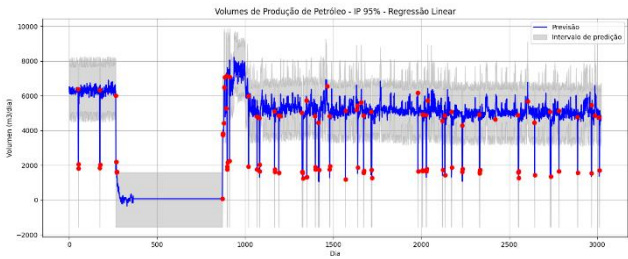
A Figura 11 apresenta os dados para de 90% de IP. Pontos vermelhos indicam previsões pontuais que ficam fora do IP estabelecido.

Figura 11. Previsões com IP de 90%



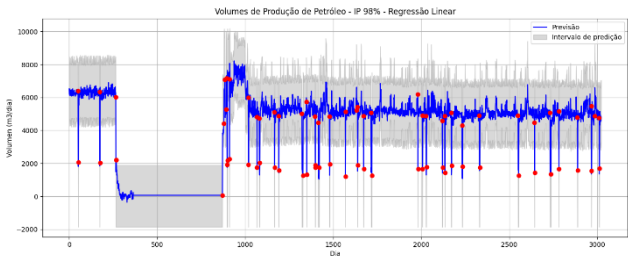
Enquanto, na Figura 12, com um 95% de IP pode-se observar que menos previsões ficam fora do intervalo de previsão.

Figura 12. Previsões com IP de 95%



Já na Figura 13, aumenta-se ainda mais o intervalo de previsão a custo de diminuir a precisão das predições, portanto menos pontos ficam de fora do intervalo de precisão. O IP é determinado pelo quanto podemos cometer um erro do tipo 1, isto é, ter um falso positivo para uma predição pontual.

Figura 13. Previsões com IP de 98%



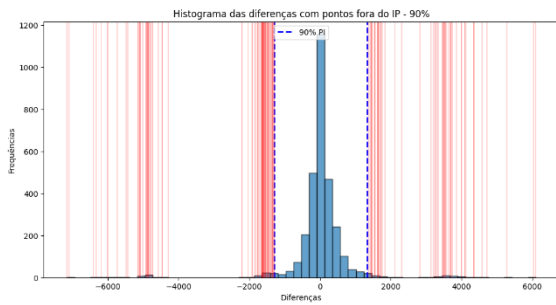
Partindo ao teste de precisão pontual do modelo, foi-se elaborado o seguinte teste de hipótese para cada nível de significância estatística que determinam cada intervalo de previsão.

H0: Os valores previstos **não** estão distribuídos entre os limites do intervalo de predição.

H1: Os valores previstos **estão** distribuídos entre os limites do intervalo de predição.

Para o IP de 90%, a Figura 14 apresenta a distribuição dos valores previstos que estão entre e fora o IP.

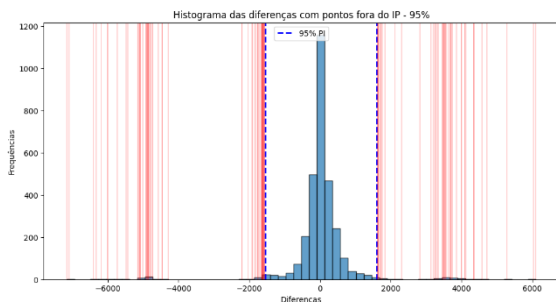
Figura 14. Distribuição das diferenças no IP 90 %



Para o IP de 90% obteve-se 4.7 % de previsões fora do intervalo, que é menor que o nível de significância de 10%, portanto, rejeita-se a hipótese nula em favor da hipótese alternativa, ou seja, há evidência suficiente para indicar que os valores previstos estão distribuídos entre os limites de predição, indicando que a precisão do modelo para esse IP é adequada.

Para o IP de 95%, a Figura 15 apresenta a distribuição dos valores previstos que estão entre e fora do IP.

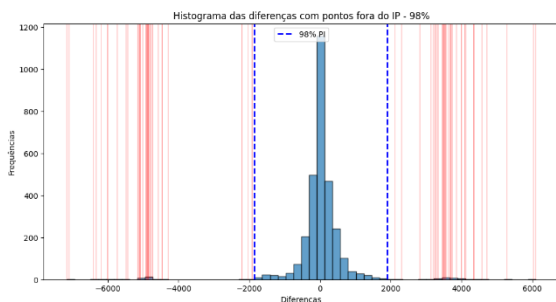
Figura 15. Distribuição das diferenças no IP 95 %



Para o IP de 95% obteve-se 3.6 % de previsões fora do intervalo, que é menor que o nível de significância de 5%, portanto, rejeita-se a hipótese nula em favor da hipótese alternativa, ou seja, há evidência suficiente para indicar que os valores previstos estão distribuídos entre os limites de predição, indicando que a precisão do modelo para este IP é adequada.

Para o IP de 98%, a Figura 16 apresenta a distribuição dos valores previstos que estão entre e fora do IP.

Figura 16. Distribuição das diferenças no IP 98 %



Para o IP de 98% obteve-se 2.5 % de previsões fora do intervalo, que é maior que o nível de significância de 2%, portanto, falha-se em rejeitar a hipótese nula em favor da hipótese alternativa, ou seja, não evidência o suficiente para indicar que os valores previstos estão distribuídos entre os

limites de predição, indicando que a precisão do modelo para este IP é inadequada.

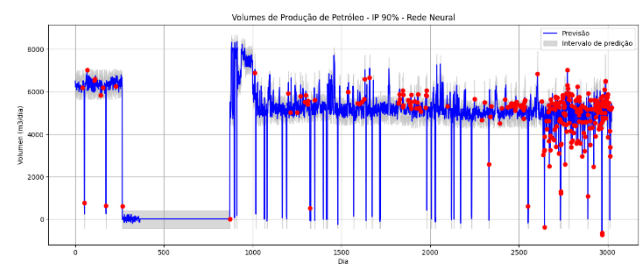
Ao tomarmos este modelo com IP de 98 %, assumimos o risco de termos mais falso-positivos quanto estabelecemos como critério.

No caso, do modelo com rede neural também foram feitos as mesmas apresentações para cada um dos intervalos.

REDE NEURAL MLP

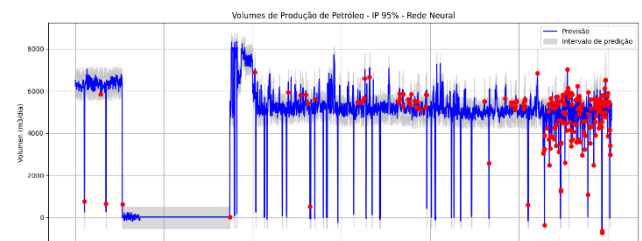
No caso do intervalo de predição de 90%, na Figura 17 se pode ver que o número de pontos vermelhos é maior do que modelo de regressão linear.

Figura 17. Previsões com IP de 90%



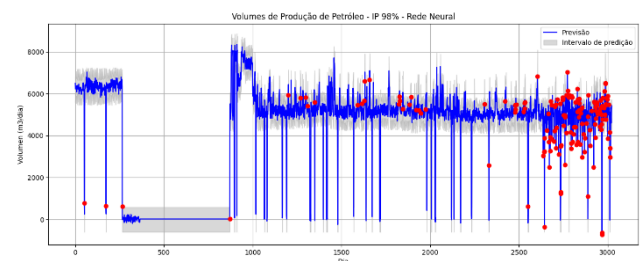
Para o caso do 95% do intervalo de predição os pontos que ficam fora do IP diminuem.

Figura 18. Previsões com IP de 95%



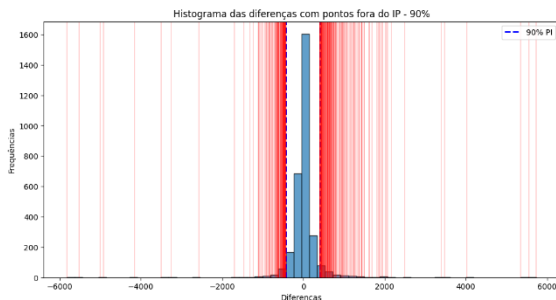
De acordo com a Figura 19, para o 98% de IP, diminui-se ainda mais os pontos que ficam fora do IP.

Figura 19. Previsões com IP de 98%



Para o IP de 90%, a Figura 20 apresenta a distribuição dos valores previstos que estão entre e fora do IP.

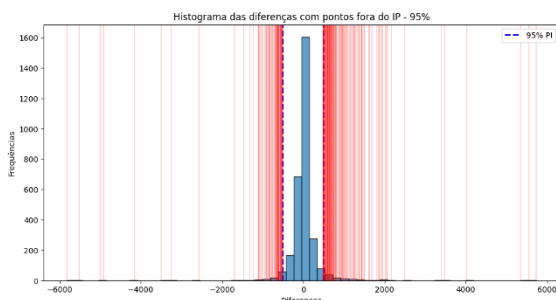
Figura 20. Distribuição das diferenças no IP 90 %



Para o IP de 90% obteve-se 8.6 % de previsões fora do intervalo, que é menor que o nível de significância de 10%, portanto, rejeita-se a hipótese nula em favor da hipótese alternativa, ou seja, há evidência suficiente para indicar que os valores previstos estão distribuídos entre os limites de predição, indicando que a precisão do modelo para esse IP é adequada.

Para o IP de 95%, a Figura 21 apresenta a distribuição dos valores previstos que estão entre e fora o IP.

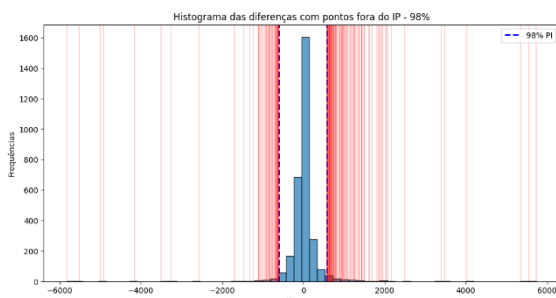
Figura 21. Distribuição das diferenças no IP 95 %



Para o IP de 95% obteve-se 6.4 % de previsões fora do intervalo, que é maior que o nível de significância de 5%, portanto, falha-se em rejeitar a hipótese nula em favor da hipótese alternativa, ou seja, não há evidência suficiente para indicar que os valores previstos estão distribuídos entre os limites de predição, indicando que a precisão do modelo para este IP é inadequada.

Para o IP de 98%, a Figura 22 apresenta a distribuição dos valores previstos que estão entre e fora o IP.

Figura 22. Distribuição das diferenças no IP 98 %



Para o IP de 98% obteve-se 5.1 % de previsões fora do intervalo, que é maior que o nível de significância de 2%, portanto, falha-se rejeitar a hipótese nula em favor da hipótese alternativa, ou seja, não há evidência suficiente para indicar que os valores previstos estão distribuídos

entre os limites de predição, indicando que a precisão do modelo para este IP é inadequada.

IV. RESULTADOS E DISCUÇÃO

No caso das distribuições das Diferenças, Figuras 5 e 6, pode-se dizer, que no caso do modelo com a rede neural a distribuição é mais centrada na média, apresentando maior precisão do que o modelo de regressão.

No caso dos gráficos de avaliação do modelo de regressão linear, Figuras 1 e 2, se demonstrou a existência de não linearidades que não foram capturadas pelo modelo de regressão linear, mas que foram capturadas ao utilizar o modelo com a rede neural com função de ativação não-linear. É provável que a existência de não linearidades no modelo de regressão linear seja consequência do que o modelo foi construído com dados com distribuição não normal.

A não normalidade da distribuição dos dados de Diferenças foi demonstrada tanto com Estatística Descritiva utilizando Q-Q Plot, assim como por Inferência Estatística por meio do teste de hipótese de normalidade de Shapiro-Wilk e Kolmogorov-Smirnov, cujos resultados em ambos os casos foi de p-valor < 0.05 .

De acordo à Tabela 2, pode-se ver que os limites de Intervalos de Confiança do modelo com a rede neural são mais amplos, com relação do modelo de regressão linear. De acordo com a Tabela 3, a amplitude dos limites dos Intervalos de Predição (IP) dos dois modelos são pequenas, sendo os da regressão linear menores. Comparando as amplitudes dos intervalos entre IC e IP, os IP são mais amplos do que IC.

De acordo aos resultados da Tabela 4, pode-se dizer o modelo de Rede Neural apresenta menor erro médio, porém menor capacidade de lidar com outliers, sendo que para os IP de 95% e 98% o modelo falha no teste de significância estatística que avalia sua capacidade de previsão pontual. Para o modelo de regressão linear, o IP no qual o modelo falha é apenas o de 98 %, mostrando que para os IP de 90% e 95% o modelo é mais confiável em suas previsões. Ao confrontarmos estes resultados, é possível que: o modelo de Rede Neural apresente maior capacidade de generalização a novos dados, porém ao custo de menor precisão no tratamento de outliers, isto é, o mapeamento do modelo de regressão linear se contorce mais no espaço a fim de acomodar os dados utilizados neste trabalho, obtendo maior precisão a custo de uma menor generalização (carece de novas pesquisas).

TABELA 4. INTERVALOS DE PREDIÇÕES

Modelo	MSE	RMSE	Intervalo de Previsão	Predições fora do IP[%]
Regressão Linear	0.009	0.099	90%	4.7
			95%	3.6
			98%	2.5
Rede Neural MLP	0.003	0.050	90%	8.6
			95%	6.4
			98%	5.1

Foi observado que o número de predições fora dos intervalos obviamente diminui conforme o intervalo aumenta seu limite, o que é esperado, visto que se aumenta o grau de confiança nas previsões a custo de uma menor precisão. Entretanto, um IP maior não implica em maior facilidade de se aceitar a hipótese nula, pelo contrário, foi observado que tanto para a regressão linear quanto para a rede neural o intervalo de predição de 98 % falhou em rejeitar a hipótese nula, isto indica uma grande concentração de pontos nas extremidades da distribuição.

Mesmo que uma predição possua um erro, ela pode ter uma margem de aceitação, logo, é importante conhecer quais são essas previsões dentro de um intervalo de predição (IP) e assim determinar quais riscos de uma previsão errada o modelo apresenta. Portanto, dado as distinções de resultados obtidos para ambos os modelos, o uso de IP pode ser considerado como uma ferramenta complementar aos indicadores globais de MSE e RMSE.

Finalmente, os intervalos de predição, por sua utilidade, parecem que ficam mais perto do lado da camada de negócio pontual para a tomada de decisões em vez do lado da construção do modelo preditivo.

V. CONCLUSÃO

Se conclui que o método de Intervalo de Predição é eficaz porque permite detectar exatamente quais predições podem ser consideradas críticas.

REFERÊNCIAS

- [1] IBM, “AI versus machine learning versus deep learning versus neural networks: What’s the difference?,” 2023. <https://www.ibm.com/think/topics/ai-vs-machine-learning-vs-deep-learning-vs-neural-networks>.
- [2] IBM, “Deep learning vs. machine learning,” *www.ibm.com*. <https://www.ibm.com/br-pt/topics/deep-learning#:~:text=O deep learning elimina parte,da dependência de especialistas humanos>.
- [3] Artificial Intelligence, “Does data preprocessing is necessary and important in deep learning?,” *ai.stackexchange.com*, 2023. <https://ai.stackexchange.com/questions/41272/does-data-preprocessing-is-necessary-and-important-in-deep-learning>.
- [4] Bookdown, “Prueba de Shapiro-Wilks,” *bookdown.org*. <https://bookdown.org/dietrichson/metodos-cuantitativos/test-de-normalidad.html>.
- [5] Á. Roco-Videla, R. Landabur-Ayala, N. Maureira-Carsalade, and M. Olguin-Barraza, “¿Cómo determinar efectivamente si una serie de datos sigue una distribución normal cuando el tamaño muestral es pequeño?,” *Nutr. Hosp.*, vol. 40, 2023, doi: <https://dx.doi.org/10.20960/nh.04519>.
- [6] SlideShare, “Prueba de Bondad de Ajuste de Kolmogorov-Smirnov (KS),” *www2.ulpgc.es*, 2023. https://www2.ulpgc.es/hege/almacen/download/5/5015/Complemento_3_Prueba_de_Bondad_de_Ajuste_de_Kolmogorov_Smirnov.pdf.
- [7] K. M. Domingues, F. L. Pereira de Oliveira, F. Cruz, and L. França Bessegato, “Estimação de Intervalos de Confiança via Reamostragem Bootstrap,” *Encontro Interestad. Eng. Produção*, 2015, [Online]. Available: <https://www.est.ufmg.br/ftp/fcruz/pub/eindep-icboot.pdf>.
- [8] F. P. Mayer, “Métodos de reamostragem,” <http://cursos.leg.ufpr.br/>. http://cursos.leg.ufpr.br/ce089/10_bootstrap.html.
- [9] K. Y. X. Wang, “Confidence and prediction intervals,” *kevinwangstats.com*, 2021. <https://www.kevinwangstats.com/post/2021-05-30-confidence-and-prediction-intervals/>.
- [10] S. Glen, “Statistics How To,” *StatisticsHowTo.com*. <https://www.statisticshowto.com/probability-and-statistics/confidence-interval/>.
- [11] B. Artley, “Time Series Forecasting: Prediction Intervals,” *towardsdatascience.com*, 2022. <https://towardsdatascience.com/time-series-forecasting-prediction-intervals-360b1bf4b085>.
- [12] GeeksforGeeks, “How to Find the T Critical Value in Python?,” *geeksforgeeks.org*, 2022. <https://www.geeksforgeeks.org/how-to-find-the-t-critical-value-in-python/>.
- [13] R. Colman, “Prediction Interval, the wider sister of Confidence Interval,” *datascienceplus.com*, 2018. <https://datascienceplus.com/prediction-interval-the-wider-sister-of-confidence-interval/>.
- [14] R. Hyndman and G. Athanasopoulos, “FORECASTING: PRINCIPLES AND PRACTICE,” *otexts.com*, 2018. <https://otexts.com/fpp2/prediction-intervals.html>.
- [15] J. Perktold, S. Seabold, and J. Taylor, “Linear regression diagnostics,” *statsmodels.org*, 2024. https://www.statsmodels.org/dev/examples/notebooks/generated/linear_regression_diagnostics_plots.html.
- [16] B. Pujala, “Creating Diagnostic Plots in Python,” *medium.com*, 2023. <https://medium.com/@pujalabhanuprakash/creating-diagnostic-plots-in-python-6e71f53ffe9a>.
- [17] PennStateScience, “Residuals vs. Fits Plot,” *online.stat.psu.edu*, 2018. <https://online.stat.psu.edu/stat462/node/117/>.
- [18] Aprender Estatística Fácil, “Teste de Normalidade,” *estatisticafacil.org*, 2023. <https://estatisticafacil.org/teste-de-normalidade/>.
- [19] Psicometria.online, “O que é o teste de Shapiro-Wilk?,” *blog.psicometriaonline.com.br*. <https://www.blog.psicometriaonline.com.br/o-que-e-o-teste-de-shapiro-wilk/>.
- [20] Scipy, “kstest,” *docs.scipy.org*. <https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.kstest.html>.
- [21] Wikipedia, “Teorema del límite central,” *es.wikipedia.org*, 2024. https://es.wikipedia.org/wiki/Teorema_del_límite_central.