

Previsão de Séries Temporais com de Redes Neurais Recorrentes

Rafael V. Curiel
ex165318

Reginaldo C. de Souza
ex165442

Marcus V. S. Monteiro
ex165833

Lucas W. Molin
ex165314

Camila Veggi
ex165308

Abstract—Este estudo aborda a previsão de consumo de energia elétrica no Brasil, empregando técnicas de aprendizado de máquina em uma extensa base de dados pública abrangendo cerca de duas décadas. Para a previsão, são utilizadas redes neurais recorrentes do tipo LSTM, capazes de modelar sequências temporais com dependências de longo prazo. Quatro abordagens de previsão são aplicadas, envolvendo a série original, a série em escala logarítmica, a série diferenciada e a série logarítmica diferenciada. Uma análise exploratória dos dados é realizada, incluindo testes de normalidade e estacionariedade, revelando que os dados não seguem uma distribuição normal e apresentam não estacionariedade. Portanto, técnicas de transformação são aplicadas para tornar os dados estacionários, reduzindo tendência e sazonalidade. A avaliação dos modelos é feita por meio de validação cruzada, com divisão dos dados em conjuntos de treinamento e teste, e as previsões para os próximos 12 meses são comparadas com os valores reais observados. As métricas de avaliação incluem o erro quadrático médio (MSE), a raiz do erro quadrático médio (RMSE), o erro médio absoluto (MAE) e o erro percentual absoluto médio (MAPE).

Index Terms—previsão, redes neurais, modelos, energia

I. INTRODUÇÃO

O consumo de energia elétrica no Brasil tem aumentado ao longo das décadas devido ao crescimento populacional, urbanização e atividade econômica, mas o setor enfrenta desafios como expansão da capacidade e eficiência energética. Este estudo empregou aprendizado de máquina, especificamente redes neurais recorrentes (LSTM), em dados de consumo elétrico de duas décadas. Os resultados podem aprimorar previsões, beneficiando o setor elétrico ao aprimorar eficiência, reduzir custos e garantir a confiabilidade do sistema.

O estudo utilizou quatro abordagens de previsão, incluindo a série original, transformação em escala logarítmica, diferenciação da série original e diferenciação da escala logarítmica, demonstrando potencial para aprimorar modelos de previsão de redes neurais em séries temporais univariadas. Neste estudo, entende-se como diferenciação ou série diferenciada como o cálculo da diferença entre os valores consecutivos da série.

II. METODOLOGIA

A base de dados utilizada foi adquirida do portal da Empresa de Pesquisa Energética (EPE), abrangendo o período de janeiro/04 até julho/23, representando o consumo consolidado de energia elétrica no Brasil em diferentes setores da economia: industrial, comercial e residencial. Os dados passaram

por um pré-processamento que incluiu ajustes estruturais e normalização para uma escala de 1:100000.

A análise dos dados empregou redes neurais recorrentes LSTM (Long Short-Term Memory), conhecidas por modelar sequências temporais com dependências de longo prazo. Os parâmetros padrões dos algoritmos foram mantidos, com algumas modificações, como estrutura de quatro camadas, 64 neurônios na camada principal e funções de ativação 'tanh', de perda 'mse' e com otimizador 'adam'. Os modelos foram treinados com dados de 24 meses para prever os próximos 12 meses, durante 100 épocas.

Antes da aplicação dos modelos, uma análise exploratória dos dados revelou a falta de normalidade, tendências e sazonalidade. Foi possível também identificar o caráter não estacionário da série, isto é, suas propriedades estatísticas, como média e variância, não são constantes ao longo do horizonte de tempo. Desta forma, foram justificadas a aplicação de técnicas de diferenciação e escala logarítmica, ambas para tornar os dados mais próximos da distribuição normal [6]. A avaliação da capacidade de previsão envolveu lotes de treinamento e teste, com previsões de 12 meses comparadas aos valores reais. Métricas como MSE, RMSE, MAE e MAPE foram empregadas na avaliação.

III. RESULTADOS E DISCUSSÃO

A. Estatísticas Descritivas

As estatísticas descritivas, para identificação de padrões e tendências nas variáveis [4], abrangeu um conjunto de 235 observações. A previsão da série "Original" revelou uma média de 364,43, que representa o valor central dessa distribuição. Porém, o desvio padrão de 46,81 foi alto, mostrando uma grande dispersão dos dados em torno da média. As variáveis transformadas, "Log" e "Log_Diff", tiveram médias mais próximas e desvios padrão menores, enquanto a variável "Diff" teve um desvio padrão alto, de 8,81, indicando variações importantes nas diferenças entre as observações. Além disso, os quartis mostraram uma assimetria nos dados, com a mediana da variável "Original" em 378,42 e uma faixa ampla de valores entre o primeiro quartil de 321,43 e o terceiro quartil de 399,87.

B. Testes de Normalidade e estacionariedade

No que se refere ao teste de Shapiro, que permite avaliar se a distribuição de uma amostra de dados vem de uma distribuição normal [4], tanto a versão com valores originais

TABLE I: Estatísticas Descritivas

Estatística	Original	Log	Diff	Log_Diff
Contagem	235.00	235.00	235.00	235.00
Média	364.43	5.89	0.64	0.00
Desvio Padrão	46.81	0.13	8.81	0.02
Mínimo	265.08	5.58	-38.75	-0.10
25%	321.43	5.77	-4.48	-0.01
50%	378.42	5.94	0.45	0.00

quanto com valores em escala logarítmica indicam valores p muito baixos ($p \leq 0,05$), sugerindo que os dados não seguem uma distribuição normal. Isso é uma informação importante, visto que muitos modelos de previsão baseiam-se na suposição de normalidade dos dados. Quanto aos testes KPSS, cuja finalidade é avaliar a estacionariedade da série [5], a estatística original sugere a presença de não estacionariedade ($p < 0,05$), enquanto a estatística na série após a diferenciação e diferenciação do logaritmo apresenta valores p maiores ($p > 0,05$), indicando que essas transformações foram eficazes em tornar os dados estacionários.

Os resultados obtidos agregam implicações significativas na modelagem dos algoritmos, uma vez que muitos modelos de previsão, como as redes neurais recorrentes (LSTM) usadas no estudo, pressupõem que os dados sejam normalmente distribuídos e estacionários.

TABLE II: Testes de Normalidade e estacionariedade

Teste	Original	Log	Diff	Log_Diff
Shapiro Estatística	0.9433	0.9266	0.9809	0.9868
Shapiro P-value	0.0000	0.0000	0.0029	0.0286
KPSS Estatística	2.2720	2.2370	0.0900	0.2390
KPSS P-value	0.0100	0.0100	0.1000	0.1000
KPSS Valor Crítico (5%)	0.4630	0.4630	0.4630	0.4630

C. Previsão

Para avaliar a qualidade das previsões, foram utilizadas métricas tradicionais que medem o distanciamento das previsões em relação a um valor real [7], são elas: Erro Quadrático Médio (MSE), Raiz do Erro Quadrático Médio (RMSE), Erro Absoluto Médio (MAE) e o Erro Percentual Absoluto Médio (MAPE). Observa-se que todas as métricas diminuíram à medida que se aplicaram técnicas de pré-processamento e normalização dos dados, sugerindo que tais técnicas contribuíram para a melhoria da precisão das previsões. Desta forma, é importante notar que os valores finais na categoria "Log_Diff" são os mais baixos, indicando um desempenho superior em termos de previsão, com um RMSE de apenas 6,989 e um MAPE de 0,013. Esses resultados reforçam a eficácia das redes neurais recorrentes (LSTM) na análise de séries temporais univariadas.

As figuras a seguir revelam os ajustes das previsões aos valores reais:

TABLE III: Resultados

Métrica	Original	Log	Diff	Log_Diff
MSE	166097.544	3540.045	55.090	48.850
RMSE	407.551	59.498	7.417	6.989
MAE	407.419	58.645	6.040	5.690
MAPE	17.551	0.158	0.014	0.013

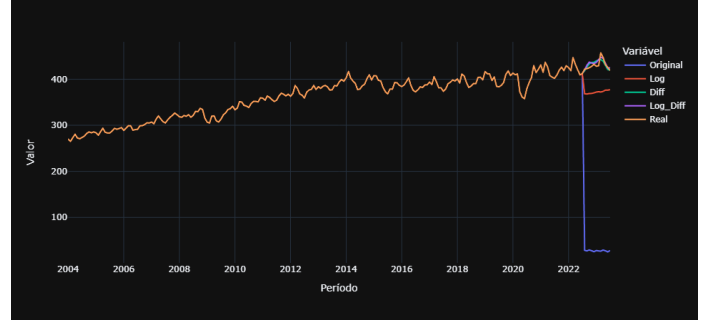


Fig. 1: Visão global dos resultados de previsão.

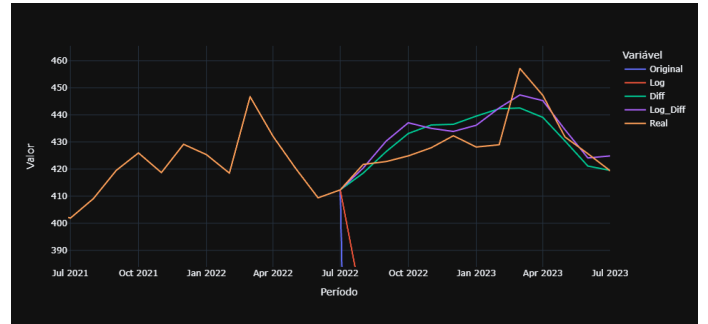


Fig. 2: Visão aproximada dos resultados de previsão.

IV. CONCLUSÃO

Este estudo revelou a eficácia das Redes Neurais Recorrentes (LSTM) na previsão de séries temporais univariadas, notavelmente aplicadas a um conjunto de dados de consumo de energia elétrica no Brasil. A análise exploratória inicial dos dados destacou a não normalidade e a presença de não estacionariedade. Como resultado, procedeu-se à aplicação de técnicas de diferenciação para tornar os dados estacionários, uma abordagem eficaz que viabilizou a implementação de modelos de previsão. Quatro abordagens de previsão foram posteriormente avaliadas, considerando a série original, a série em escala logarítmica, a série diferenciada e a série logarítmica diferenciada. Os resultados da validação cruzada indicaram que a série logarítmica diferenciada apresentou as previsões mais acuradas.

Os resultados alcançados ressaltam a habilidade das LSTMs em modelar efetivamente séries temporais univariadas, mesmo na presença de tendências e sazonalidade, com a diferenciação e a normalização desempenhando um papel fundamental na melhoria da precisão das previsões. Este estudo acrescenta uma contribuição significativa à literatura, ao demonstrar a aplicação prática das LSTMs na previsão de séries temporais

univariadas, oferecendo perspectivas valiosas para aprimorar a eficiência energética e a confiabilidade do sistema elétrico. Contudo, vale notar a limitação da aplicabilidade geral deste estudo devido à especificidade do conjunto de dados utilizado, incentivando pesquisas futuras que explorem a utilidade dessas técnicas em diversos contextos e conjuntos de dados. Adicionalmente, apesar dos ganhos observados na precisão das previsões, existe ainda espaço para aprimoramentos, com a otimização de hiperparâmetros representando uma área promissora a ser investigada. Ademais, esta pesquisa valida a viabilidade e eficácia das LSTMs na previsão de séries temporais univariadas e sinaliza implicações substanciais para o setor elétrico, ao mesmo tempo que oferece um ponto de partida para futuras investigações nesse campo acadêmico.

REFERENCES

- [1] Curiel, R. (2023). "Unicamp Machine Learning Repository." GitHub. Disponível em: https://github.com/r-curiel/unicamp_machine_learning. Acessado em 11 de outubro de 2023.
- [2] Python Software Foundation (2023). "Python 3.10.0 documentation." Disponível em: <https://docs.python.org/3.10/>. Acessado em 11 de outubro de 2023.
- [3] Empresa de Pesquisa Energética (2023). "Consumo de Energia Elétrica." Disponível em: <https://www.epe.gov.br/pt/publicacoes-dados-abertos/publicacoes/consumo-de-energia-eletrica>. Acessado em 11 de outubro de 2023.
- [4] Belfiore, P. & Fávero, L. P. (2021). "Manual de Análise de Dados - Estatística e Modelagem Multivariada com Excel®, SPSS® e Stata®." 1ª edição. GEN LTC. Rio de Janeiro, RJ, Brasil.
- [5] Sieben, E. (2017). "Modelos para previsão em Séries Temporais: A Metodologia Box e Jenkins." Trabalho de Conclusão de Curso (Graduação em Estatística), Fundação Universidade Federal de Rondônia, Campus de Ji-Paraná.
- [6] Box, G. E. P., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). "Time Series Analysis: Forecasting and Control," 5th Edition.
- [7] Oliveira Júnior, C. (2021). Métricas para Regressão: Entendendo as métricas R², MAE, MAPE, MSE e RMSE. Medium. <https://medium.com/data-hackers/prevendo-n%C3%BAmeros-entendendo-m%C3%A9tricas-de-regress%C3%A3o-35545e011e70>.