

Previsão do Consumo de Energia Elétrica no Brasil: Comparação entre Modelos SARIMA e LSTM

Dianna Mayumi Santos Katayama Rodrigues¹, Ricardo Pardon¹

¹Faculdade de Computação e Informática (FCI)

Universidade Presbiteriana Mackenzie – São Paulo, SP – Brasil

diannamayumi@gmail.com

rpardon@gmail.com

Resumo. Este trabalho apresenta uma análise comparativa entre os modelos SARIMA e LSTM aplicados à previsão de consumo de energia elétrica no Brasil. A partir de uma análise exploratória detalhada, padrões de tendência, sazonalidade e ruído foram identificados na série temporal. Métodos de decomposição e testes de estacionariedade, como ADF e KPSS, foram utilizados para avaliar as características dos dados e orientar o pré-processamento necessário. O modelo SARIMA foi ajustado com base nos parâmetros obtidos da análise ACF e PACF, enquanto o LSTM foi configurado com hiperparâmetros otimizados por meio de validação cruzada. Ambos os modelos foram avaliados utilizando métricas como RMSE, MAE e MAPE, considerando conjuntos de treino e teste. Os resultados destacam o desempenho de cada modelo em diferentes aspectos, com o SARIMA mostrando maior robustez em séries com forte sazonalidade e o LSTM demonstrando potencial em capturar padrões complexos e não lineares. A análise comparativa oferece uma visão abrangente sobre a aplicabilidade de modelos estatísticos e de aprendizado profundo em séries temporais, contribuindo para a escolha de abordagens adequadas em cenários específicos.

Palavras-chave: previsão de séries temporais; consumo de energia elétrica; modelos SARIMA; redes neurais LSTM; análise comparativa; aprendizado de máquina; sazonalidade e tendências.

Abstract. This paper presents a comparative analysis of time series forecasting models, focusing on energy consumption prediction. The study examines two approaches: the SARIMA model, widely used for its statistical foundation and capacity to model seasonality, and the LSTM neural network, a deep learning method capable of capturing complex temporal patterns. The project begins with exploratory data analysis to identify patterns, trends, seasonality, and noise in the energy consumption series. Key preprocessing steps include handling missing data, detecting outliers, and applying stationarity tests (ADF and KPSS). The SARIMA model is initially configured based on ACF and PACF analysis, with parameters optimized for effective forecasting. Simultaneously, the LSTM model is developed with hyperparameter tuning and neural architecture adjustments to enhance its predictive power. Validation is

performed using metrics such as RMSE, MAE, and MAPE. The results reveal the strengths and limitations of each model, providing insights into their performance under different conditions. The research concludes by highlighting the trade-offs between statistical and machine learning approaches, emphasizing their complementary roles in energy forecasting.

Keywords: *time series forecasting; energy consumption; SARIMA models; LSTM neural networks; comparative analysis; machine learning; seasonality and trends.*

1. Introdução

O consumo de energia elétrica é um elemento crucial para o desenvolvimento econômico e social de qualquer nação. No Brasil, o setor elétrico é vital para sustentar as atividades industriais, comerciais e residenciais, sendo responsável por uma parcela significativa do crescimento econômico. Com o aumento populacional e o avanço tecnológico, a demanda por energia elétrica tem crescido exponencialmente, o que requer um planejamento adequado para garantir a oferta de energia de maneira sustentável e eficiente (AMARO et al., 2017).

Esse crescimento de consumo energético traz desafios importantes, como a pressão sobre recursos naturais limitados, o aumento das emissões de gases de efeito estufa e a necessidade de garantir um fornecimento contínuo, estável e acessível de energia elétrica. Essas questões são centrais nas discussões sobre sustentabilidade e exigem uma abordagem integrada que combine previsões precisas de consumo com estratégias de mitigação dos impactos ambientais (SANTOS; CHAUCOSKI, 2020).

A energia elétrica não apenas sustenta as operações cotidianas das indústrias e o bem-estar da população, mas também é considerada um indicador-chave da qualidade de vida e do nível de desenvolvimento econômico de um país. A capacidade de prever com precisão o consumo de energia elétrica é fundamental para o planejamento eficiente dos sistemas de geração e distribuição de energia, evitando assim apagões e maximizando a eficiência operacional das usinas (SANTOS; CHAUCOSKI, 2020).

Neste contexto, a análise de séries temporais se destaca como uma ferramenta poderosa para prever o consumo de energia elétrica, fornecendo informações estratégicas para a tomada de decisão no setor energético. A previsão do consumo futuro de energia é essencial para evitar crises energéticas, otimizar a operação das usinas e promover o uso eficiente dos recursos energéticos (TAVARES; QUADRELLI, 2022).

Este projeto busca aplicar e comparar dois modelos preditivos amplamente utilizados: o SARIMA (Seasonal AutoRegressive Integrated Moving Average), conhecido por sua robustez em capturar padrões sazonais e lineares em séries temporais, e o LSTM (Long Short-Term Memory), uma rede neural recorrente que tem se mostrado eficaz em capturar padrões não lineares complexos em séries temporais.

A crescente demanda por energia elétrica no Brasil, impulsionada por fatores econômicos, climáticos e sociais, reforça a necessidade de desenvolver métodos preditivos eficientes que possam antecipar essa demanda. Previsões precisas são essenciais para evitar problemas como apagões e sobrecargas na rede elétrica, além de contribuir para a redução de custos operacionais e para o desenvolvimento sustentável.

A análise comparativa entre os modelos SARIMA e LSTM oferece uma oportunidade valiosa para explorar as vantagens e limitações dessas abordagens no contexto do consumo energético. O modelo SARIMA é amplamente utilizado para capturar padrões sazonais e tendências em séries temporais, sendo eficiente em dados com sazonalidade e tendências bem definidas. Já o LSTM se destaca em contextos onde há necessidade de capturar dependências complexas ao longo do tempo, especialmente em dados não estacionários (BROWNLEE, 2019).

Além disso, este projeto está alinhado com os Objetivos de Desenvolvimento Sustentável (ODS), particularmente o ODS 7 estabelecido pela Organização das Nações Unidas (ONU), que visa garantir o acesso universal, confiável, sustentável, moderno e a preço acessível aos serviços de energia. A previsão eficiente do consumo de energia elétrica pode auxiliar na formulação de políticas públicas que promovam o uso sustentável de recursos, contribuindo para a mitigação dos impactos ambientais e para o avanço da sustentabilidade no setor energético brasileiro.

O objetivo geral deste projeto é desenvolver modelos preditivos para estimar o consumo mensal de energia elétrica no Brasil a médio prazo, utilizando dados históricos fornecidos pela Empresa de Pesquisa Energética (EPE). Os modelos serão implementados na linguagem Python. Serão comparados dois modelos de séries temporais: SARIMA, tradicionalmente utilizado para previsões que envolvem padrões sazonais, e LSTM, uma abordagem mais moderna baseada em redes neurais profundas, capaz de capturar padrões não lineares complexos. Para atingir o objetivo geral temos como objetivos específicos:

- Analisar o comportamento histórico do consumo mensal de energia elétrica no Brasil utilizando a base de dados disponibilizada pela Empresa de Pesquisa Energética (EPE);
- Desenvolver e validar modelos preditivos utilizando SARIMA e LSTM com os dados históricos de consumo de energia;
- Comparar os modelos para avaliar a precisão das previsões mensais de consumo de energia utilizando as métricas RMSE, MSE e MAPE;
- Identificar o modelo que oferece melhor desempenho e realizar a previsão do consumo de energia elétrica no Brasil para os próximos doze meses.

A relevância deste estudo está em sua contribuição para o planejamento energético no Brasil, oferecendo ferramentas mais precisas para prever a demanda futura e otimizar recursos. Alinhado ao ODS 7 da ONU, o trabalho busca fomentar o uso sustentável dos recursos energéticos e a redução de impactos ambientais, além de auxiliar na formulação de políticas públicas no setor energético.

Este artigo está organizado da seguinte forma: na seção 2, apresentamos a análise exploratória e o pré-processamento dos dados. Na seção 3, detalhamos os modelos SARIMA e LSTM, incluindo suas configurações e treinamento. Na seção 4, comparamos os modelos com base em métricas de desempenho. Por fim, a seção 5 apresenta as conclusões e sugestões para trabalhos futuros.

A construção deste projeto será compartilhada na plataforma colaborativa Github e poderá ser acessada pelo link: <https://github.com/RickPardono/cdProjetoAplicadoIV>.

A base de dados utilizada neste projeto foi extraída do site oficial da Empresa de Pesquisa Energética (EPE, 2024), uma instituição pública vinculada ao Ministério de Minas e Energia do Brasil. A base de dados é parte do conjunto de dados abertos

disponibilizados pela EPE para download público através do link: <https://www.epe.gov.br/pt/publicacoes-dados-abertos/dados-abertos/dados-do-consumo-mensal-de-energia-eletrica>.

Seguem mais detalhes abaixo sobre a base de dados:

Nome do arquivo: Dados_abertos_Consumo_Mensal.xlsx Tamanho: 4.858 KB Licença Creative Commons Atribuição 4.0 Internacional (CC BY 4.0): livre para uso, cópia, modificação, distribuição e reprodução do conjunto de dados para qualquer finalidade, comercial ou não comercial, desde que seja dada a devida atribuição.

Data da última atualização: 05/07/2024

Visibilidade: Pública

Colaboradores: a base de dados é mantida pela Equipe de Estatísticas Energéticas da EPE, com base nas informações fornecidas mensalmente pelos agentes do mercado de energia elétrica e compiladas pela Comissão Permanente de Análise e Acompanhamento do Mercado de Energia Elétrica (COPAM).

Cobertura Temporal: janeiro de 2004 a junho de 2024

Cobertura Espacial: Brasil

Metodologia de Coleta: Os dados são coletados mensalmente pelas concessionárias de energia elétrica, utilizando sistemas de medição que registram o consumo de energia elétrica em diferentes categorias de consumidores. Esses dados são então reportados à EPE, que os processa e disponibiliza na forma de dados abertos para análise.

Dicionário de dados:

(CONSUMO E NUMCONS SAM):

VARIÁVEL	TIPO	DESCRIÇÃO
Data	integer	Data de referência do consumo (formato 'yyyymmdd')
DataExcel	datetime	Data em formato padrão do Excel com número de série
Regiao	string	Região Geográfica do Brasil (Centro-Oeste, Nordeste, Norte, Sudeste, Sul) .
Sistema	string	Subsistema ou Sistemas Isolados (NORDESTE, NORTE INTERLIGADO, SISTEMAS ISOLADOS, SUDESTE/ CENTRO - OESTE, SUL)
Classe	string	Categoria do consumidor (Comercial, Industrial, Outros, Residencial, Rural)
TipoConsumidor	string	Tipo de consumidor (Cativo ou Livre)
Consumo	float	Consumo de energia em MWh
Consumidores	float	Número de unidades consumidoras registradas no respectivo mês

2. Referencial Teórico

A energia elétrica é considerada como um dos indicadores do nível de qualidade de vida e do desenvolvimento econômico de um país. Além de viabilizar maior conforto para as pessoas, sua utilização proporciona um aumento da capacidade de produção de bens e serviços. Entretanto, o consumo de energia elétrica com o crescimento da economia e desenvolvimento da população tem aumentado gradualmente (POLO et al., 2024).

O setor elétrico contribui para o desenvolvimento das diferentes regiões brasileiras, proporcionando grande comodidade a população. No Brasil, a energia elétrica é dada por uma matriz composta por energias renováveis e não renováveis, sendo formada por hidráulica, solar, eólica e térmicas. Grande parte da energia elétrica gerada é proveniente das usinas hidrelétricas (ARAÚJO et al., 2023).

Em vista da importância da energia elétrica, o mercado necessita de um planejamento de abastecimento, para que ocorra uma gestão eficiente e sustentável dos sistemas de geração de energia e de tomadas de decisões pontuais relativas à otimização de aumentos da capacidade de geração atual (AMARO et al., 2017).

A previsão do consumo de eletricidade é um fator importante no planejamento eficiente de um sistema de geração de energia. Métodos de previsão de séries temporais, podem ser utilizados para avaliar o consumo de energia elétrica ao longo do tempo. Uma série temporal é qualquer conjunto de observações ordenado no tempo. As observações podem ser diárias, semanais, mensais (SANTOS; CHAUCOSKI, 2020).

Dentre os métodos de previsão de séries temporais, têm-se os métodos quantitativos SARIMA (Autoregressivos Integrados de Médias Móveis com a Característica Sazonal da Série) e as Redes Neurais Artificiais (RNAs). Os modelos SARIMA e Redes Neurais Artificiais têm sido utilizados com sucesso em tarefas de predição e modelagem de séries temporais de séries temporais (SANTOS; CHAUCOSKI, 2020).

Uma série temporal $Y_t = \{t_1, t_2, t_3, \dots, t_n\}$ é definida como uma coleção de observações sequenciais em um determinado tempo t , ou seja, são dados observados em diferentes instantes do tempo. Esta pode ser discreta, quando acontece em um tempo específico, como $(Y_t = t_1, \dots, t_n)$, ou contínua, quando as observações ocorrem continuamente no tempo $(Y_t = t : t_1 < t < t_2)$ (ARAÚJO et al., 2023).

Para compreensão do comportamento das series temporais é necessário ter conhecimento dos componentes característicos, sendo eles:

- Tendência (Trend): indica a direção geral da série ao longo do tempo, podendo ser ascendente, descendente ou constante;
- Sazonalidade: padrão de variação regular que incide sobre os dados em períodos idênticos de tempo;
- Ciclos: ocorrem quando há variações repetitivas dos dados, no entanto, essas variações não possuem período fixo.
- Ruído (Erro Aleatório): são variações imprevisíveis que não são explicadas por tendência ou sazonalidade.

Outra característica importante em séries temporais, remete-se a sua estacionariedade ou não estacionariedade. Uma série é considerada Estacionária quando

suas propriedades estatísticas (como média e variância) não mudam ao longo do tempo. Ou seja, ela não tem tendência nem variação sazonal significativa. Isso facilita a modelagem e a previsão, pois as características são constantes. Uma série é considerada Não Estacionária quando suas propriedades estatísticas mudam com o tempo, indicando presença de tendências, sazonalidades ou volatilidades. Em geral, é necessário transformar a série para torná-la estacionária antes de aplicar certos modelos, como o ARIMA e o SARIMA.

Araújo et al. (2023), compararam diferentes metodologias de séries temporais para prever o consumo de energia elétrica na Região Nordeste do Brasil. Foram utilizados dados mensais do Instituto de Pesquisa Econômica Aplicada (IPEA) que abrangem o período de janeiro de 1979 a maio de 2021. No pré-processamento, os dados foram transformados logaritmicamente para estabilizar a variância e diferenciados uma vez para remover a tendência. Na análise exploratória, as funções de autocorrelação (FAC) e autocorrelação parcial (FACP) indicaram forte sazonalidade, e a decomposição da série temporal em componentes de tendência, sazonalidade e aleatoriedade confirmou a natureza não estacionária dos dados. Para a modelagem, foram utilizados os métodos Holt-Winters nas formas aditiva e multiplicativa, e a metodologia Box-Jenkins com os modelos SARIMA e SARIMAX. O modelo SARIMA(4, 1, 1)(2, 0, 0)₁₂ apresentou o melhor ajuste com base nos critérios de informação AIC e BIC. Além disso, uma variável dummy foi utilizada no modelo SARIMAX para capturar o impacto do racionamento de energia ocorrido no início dos anos 2000. As previsões foram realizadas para oito meses à frente, até dezembro de 2021, com o modelo SARIMA mostrando-se superior, com erros de previsão baixos e boa adequação estatística. Os resultados indicaram que o modelo SARIMA(4, 1, 1)(2, 0, 0)₁₂ foi eficaz na previsão do consumo de energia no Nordeste, especialmente ao lidar com variações sazonais e mudanças estruturais causadas por eventos externos.

Castro (2021), analisou o impacto das medidas restritivas da pandemia da COVID-19 no consumo de energia elétrica na região Sudeste/Centro-Oeste e no estado de São Paulo. Foram utilizados dados mensais da classe comercial fornecidos pela Empresa de Pesquisa Energética (EPE), abrangendo o período de janeiro de 2010 a dezembro de 2020. O estudo utilizou modelos ARIMA Sazonais Multiplicativos (SARIMA) para 8 realizar a previsão do consumo das duas séries de março até dezembro de 2020 (período este afetado pelos efeitos do coronavírus), e para o ano de 2021. No pré processamento, os dados foram ajustados para estacionariedade através da diferenciação e verificação de outliers, além da conversão das datas para o formato datetime. Na análise exploratória, foram utilizadas funções de autocorrelação (ACF) e autocorrelação parcial (PACF) para identificar padrões sazonais e tendências. Para a seleção do melhor modelo, foram utilizados os critérios AIC, BIC e MAPE. O melhor modelo identificado para ambas as séries foi o SARIMA(2,1,0)(2,1,0)₁₂. A validação das previsões foi feita através das métricas RMSE e MAPE, enquanto os testes de Shapiro-Wilk e Box-Ljung foram utilizados para confirmar a normalidade e a independência dos resíduos. Nas previsões para o período de março a dezembro de 2020, foi verificado uma queda significativa no consumo energético, especialmente nos primeiros meses. Nas previsões para o ano de 2021, foi verificado uma recuperação gradual do consumo. Comparando os valores previstos com os reais, o modelo apresentou alta precisão, evidenciando a eficácia do SARIMA para capturar padrões sazonais e prever variações em cenários de incerteza.

O estudo reforçou a importância de desenvolver modelos preditivos robustos para o planejamento energético e para formulação de políticas públicas em situações de crise.

Santos e Chaucoski (2020), avaliaram modelos baseados em redes neurais LSTM (Long Short-Term Memory) e SARIMA (Seasonal Auto-regressive Integrated Moving Average), para a previsão do consumo de energia elétrica da região sudeste do Brasil. A base de dados utilizada no estudo foi disponibilizada pela Empresa de Pesquisa Energética (EPE), e apresentava uma série histórica mensal de consumo de energia elétrica no período entre Janeiro de 2004 e Maio de 2020 contendo 197 observações mensais. Os dados estavam limpos e sem outliers. Foi realizada a decomposição da série, observando-se uma tendência de aumento do consumo ao longo dos anos e também a presença de sazonalidades. Os modelos foram implementados em Python. Dentre os modelos SARIMA, foi selecionado com base no critério de AKAIKE (AIC) o modelo SARIMA(7,1,7)(0,1,0)¹². Já para o modelo LSTM, foram treinadas redes neurais utilizando a biblioteca Keras, com o algoritmo Adam e parâmetros otimizados como $cells = 4$, $batch = 1$ e $learning\ rate = 0.0005$. A validação do modelo LSTM foi feita usando métricas como MSE, RMSE e MAPE, e para o modelo SARIMA, foi feita a verificação da normalidade e autocorrelação dos resíduos. Após a validação, ambos os modelos foram utilizados para realizar previsões do consumo de energia para os meses de janeiro a maio de 2020, que não participaram do processo de treinamento e teste. Observou-se, dos dados previstos, que os modelos apresentaram resultados próximos ao da Empresa de Pesquisa Energética – EPE. Portanto, a proximidade entre valores preditos e reais demonstram a boa capacidade de generalização, para um horizonte de curto prazo, dos modelos implementados neste trabalho. Os autores ainda concluem que as previsões do comportamento do consumo de energia na 9 região sudeste, podem auxiliar no planejamento do sistema de geração de energia elétrica.

Silva e Soeiro (2012), abordaram a aplicação de modelos de séries temporais no consumo de energia elétrica no Estado do Amapá, com foco nos setores residencial, industrial, comercial e outros. Utilizando dados mensais fornecidos pela Companhia de Eletricidade do Amapá (CEA), referentes ao período de janeiro de 1995 a dezembro de 2010, o estudo teve como objetivo identificar e modelar as componentes sazonais e de tendência presentes nas séries temporais e prever o consumo futuro de energia. O método de regressão linear múltipla foi empregado para analisar a presença dessas componentes, enquanto os modelos SARIMA foram ajustados para prever o consumo de energia no ano de 2010 utilizando os valores de 1995 a 2009. Durante a análise, foram aplicados testes de autocorrelação e diferenciação para verificar a estacionariedade da série. Os modelos ajustados foram avaliados por meio do Erro Quadrático Médio (MSE), sendo que o modelo SARIMA(0,1,2)(0,1,1)¹² apresentou o melhor desempenho para prever o consumo de energia. Os resultados mostraram que o modelo SARIMA ajustado foi capaz de capturar bem a tendência e a sazonalidade da série temporal, obtendo previsões de consumo de energia com valores próximos dos valores reais observados. O consumo de energia no estado do Amapá apresentou um aumento ao longo dos anos, seguindo um padrão sazonal recorrente.

Silva (2022), analisou o consumo de energia elétrica na região Sudeste do Brasil, utilizando séries temporais e redes neurais artificiais. A pesquisa utilizou dados mensais do consumo, fornecidos pelo IPEA, de janeiro de 2002 a dezembro de 2021. Foram aplicados modelos sazonais autorregressivos (SARIMA) e de Redes Neurais Recorrentes do tipo LSTM para prever o consumo elétrico. Inicialmente, foi realizado o ajuste para

estacionariedade dos dados, utilizando o teste de Cox-Stuart para identificar a presença de tendência. Em seguida, foram analisadas as funções de autocorrelação (ACF) e autocorrelação parcial (PACF) para verificar padrões sazonais e de tendência, revelando a necessidade de diferenciação sazonal e regular. A partir dessa identificação de padrões, foram escolhidos os parâmetros dos modelos, utilizando critérios como AIC e BIC para o SARIMA, e ajustes de hiperparâmetros, como número de neurônios e épocas, para o LSTM. O modelo SARIMA foi utilizado para previsões de longo prazo, enquanto o modelo LSTM foi aplicado para previsões de curto prazo com realimentação de dados. Ambos os modelos foram validados e treinados com os dados históricos, e as previsões foram comparadas utilizando a métrica RMSE, escolhendo o modelo com menor erro. O autor concluiu que o modelo SARIMA apresentou melhor desempenho para previsões anuais, ao projetar o consumo para os 12 meses seguintes de uma só vez, enquanto o modelo LSTM foi mais eficaz em previsões mensais com realimentação dos dados.

Marcos e Pontes Júnior (2021), analisaram a previsão do consumo de energia elétrica na região Nordeste do Brasil usando séries temporais univariadas com um passo temporal à frente. O estudo utilizou dados diários do consumo, fornecidos pelo Operador Nacional do Sistema Elétrico (ONS), de 2004 a 2019. Foram aplicados modelos Naive (Persistente, Média e Mediana), SARIMA, MLP, CNN e LSTM. No pré processamento, os dados foram normalizados e outliers foram identificados e tratados. Foram realizadas análises de autocorrelação (ACF) para definir os parâmetros do SARIMA, e um grid search para otimizar os hiperparâmetros de cada modelo, como número de neurônios e épocas para redes neurais. A validação das previsões foi realizada usando o Erro Percentual Absoluto Médio (MAPE), e o teste de hipótese de Diebold-Mariano para comparação estatística. O modelo CNN apresentou o menor erro (MAPE 1,867%), seguido pelo MLP (1,894%) e SARIMA (4,390%). Os resultados indicaram que, apesar de modelos clássicos como o SARIMA capturarem bem a sazonalidade, modelos de aprendizado profundo, como CNN, alcançam maior precisão, especialmente para previsões de curto prazo. A pesquisa destacou a importância de ajustar hiperparâmetros e escolher modelos adequados para apoiar o planejamento energético e a tomada de decisões.

Andrade Filho (2021), abordou em seu estudo, o uso de redes neurais LSTM na previsão de séries temporais de consumo de energia elétrica, utilizando dados de uma empresa de eletrodomésticos em Minas Gerais. A base de dados consistiu em 96 amostras de consumo coletadas em intervalos de 15 minutos ao longo de 24 horas. No pré-processamento, os dados foram padronizados para evitar divergências no treinamento, e as sequências de predição foram configuradas com base na matriz de células do software Matlab. O modelo LSTM foi treinado com 200 unidades ocultas, 250 épocas e taxa de aprendizado ajustada, utilizando o algoritmo ADAM. A validação das previsões foi realizada com a métrica RMSE, e as previsões futuras foram feitas utilizando a função `predictAndUpdateState` do Matlab. Os resultados mostraram que as redes LSTM foram capazes de prever o consumo com margem de erro razoável. O estudo destacou que a escolha adequada dos parâmetros de treinamento e a padronização dos dados são essenciais para a precisão do modelo, e sugeriu a utilização de uma base de dados maior para melhorar a generalização e evitar overfitting.

Polo et al. (2024), compararam o desempenho de modelos LSTM univariado e multivariado na previsão do consumo de energia elétrica da região Sudeste do Brasil. A base de dados, fornecida pela EPE, continha uma série mensal de consumo de energia de

janeiro de 2004 a novembro de 2023, totalizando 239 observações. Para o pré-processamento, os dados foram normalizados usando a função MinMax. Modelos LSTM foram implementados em Python, com a rede multivariada considerando variáveis adicionais como consumo residencial, comercial e industrial. Os parâmetros dos modelos foram ajustados utilizando 600 épocas, taxa de aprendizado de 0,0001 e função de ativação ReLU. A validação foi realizada com as métricas RMSE, MAE e MAPE, sendo o modelo multivariado superior ao univariado, com MAPE de 1,06% contra 2,78%. No conjunto de teste, o LSTM multivariado também apresentou melhor desempenho, com MAE de 0,22, RMSE de 0,25 GWh e MAPE de 1,06%, enquanto o univariado teve MAE de 0,59, RMSE de 0,81 GWh e MAPE de 2,74%. O estudo concluiu que o modelo multivariado é mais eficaz na previsão de curto prazo e recomendou sua aplicação em outras regiões com maiores conjuntos de dados para validar a generalização.

Manowska (2020), abordou a previsão da demanda de energia elétrica na Polônia utilizando redes neurais LSTM. O objetivo foi desenvolver um modelo matemático para prever o consumo de energia até 2040, abrangendo setores como indústria, transporte, residencial, comercial e serviços públicos, agricultura, silvicultura e pesca. Os dados de consumo foram coletados de 1990 a 2018. No pré-processamento, os dados foram normalizados e submetidos a testes de estacionariedade e autocorrelação parcial. O estudo utilizou a função de autocorrelação parcial para determinar os atrasos apropriados para o modelo, identificando valores significativos para atrasos de 1 e 3. O modelo LSTM foi configurado com duas camadas, 25 épocas e treinamento usando a biblioteca TensorFlow. As métricas MAE, RMSE e MAPE foram utilizadas para avaliar a precisão das previsões. O modelo apresentou MAPE de 1-3% para os setores analisados, com RMSE de 97 ktoe para o setor industrial e 45 ktoe para o setor residencial. A pesquisa destacou a capacidade das redes LSTM em lidar com séries temporais não estacionárias e a importância de prever o consumo elétrico para apoiar o planejamento energético e políticas públicas.

Nichiforov et al. (2017), compararam o desempenho dos modelos ARIMA e NAR (Rede Neural Autoregressiva Não Linear) na previsão do consumo de energia elétrica em uma pequena unidade industrial na Romênia. Foram utilizados dados coletados a cada 10 segundos, durante 7 dias, totalizando 59.915 observações. No pré processamento, os dados foram divididos em conjuntos de treino e teste, e aplicou-se a metodologia Box-Jenkins para ajustar o modelo ARIMA. A escolha dos parâmetros foi baseada no critério de erro de predição final de Akaike (FPE). Para o modelo NAR, foram utilizadas 10 entradas e 1 saída, e o treinamento foi realizado com o algoritmo Levenberg-Marquardt. A técnica de cross-validation foi aplicada, com 70% dos dados para treino, 15% para validação e 15% para teste. As métricas de desempenho utilizadas foram MSE, RMSE, MAE e MAPE. O modelo ARIMA apresentou melhor desempenho geral, com MAPE de 0,1457% e RMSE de 0,0084, enquanto o NAR teve MAPE de 0,1887% e RMSE de 0,0104. O estudo concluiu que o modelo ARIMA foi mais preciso na previsão de curto prazo, mas ambos os modelos são eficazes para previsão de consumo energético dentro do contexto analisado.

Através da análise dos artigos e trabalhos desse referencial teórico, ficou demonstrado a eficácia dos modelos SARIMA e LSTM na previsão de consumo de energia elétrica, cada um com suas vantagens específicas. No presente trabalho, optou-se pela aplicação dos modelos SARIMA e LSTM devido à sua capacidade comprovada de lidar com padrões sazonais e não lineares, respectivamente. Modelos SARIMA, como observado em Araújo et al. (2023), Castro (2021) e Silva e Soeiro (2012), mostraram-se

eficazes na captura de sazonalidade e tendências bem definidas, sendo robustos para previsões em cenários estáveis e com dados estacionários. Por outro lado, a aplicação do LSTM, conforme relatado por Santos e Chaucoski (2020), Manowska (2020), Silva (2022) e Polo et al. (2024), foi eficiente na identificação de padrões complexos e variações não lineares ao longo do tempo, especialmente em dados não estacionários. Diante dessas evidências, espera-se que a combinação dessas abordagens permita uma previsão mais acurada e robusta do consumo de energia elétrica no Brasil, contribuindo para um planejamento energético mais eficiente e alinhado aos desafios do setor.

Com base na análise do estado da arte, este projeto adota ambos os modelos para explorar suas capacidades complementares no contexto brasileiro. O SARIMA, pela sua robustez em capturar padrões sazonais, e o LSTM, pela sua habilidade em identificar variações não lineares, oferecem uma combinação promissora para melhorar a precisão na previsão do consumo de energia elétrica no Brasil.

3. Metodologia

Este estudo caracteriza-se como uma pesquisa quantitativa com abordagem preditiva, focada na modelagem de séries temporais. Os métodos utilizados incluem técnicas tradicionais de séries temporais, como o SARIMA, e abordagens modernas baseadas em Redes Neurais Recorrentes do tipo LSTM.

O processo foi estruturado para avaliar a capacidade preditiva de ambos os modelos em relação ao consumo mensal de energia elétrica no Brasil, utilizando métricas de desempenho como RMSE, MAE e MAPE para análise comparativa.

Dados Utilizados e Coleta

Os dados utilizados foram obtidos a partir de uma base disponibilizada pela Empresa de Pesquisa Energética (EPE), contendo registros mensais do consumo de energia elétrica entre 2004 e 2023. O conjunto de dados foi acessado via um arquivo no formato Excel, importado diretamente de um repositório público no GitHub. Após a importação, os dados foram agregados por mês e normalizados para garantir consistência durante o processo de modelagem.

Pré-processamento e Análise Exploratória

O pré-processamento incluiu:

- **Conversão de Unidades:** Os dados de consumo foram ajustados de MWh para GWh para facilitar a análise.
- **Decomposição Temporal:** A série foi decomposta em componentes de tendência, sazonalidade e resíduos usando o método STL.
- **Testes de Estacionariedade:** Foram aplicados os testes ADF e KPSS para determinar a necessidade de transformações adicionais.
- **Análise de Outliers:** Identificaram-se outliers em anos específicos, como 2013, 2016 e 2022, sendo contextualizados para compreender sua origem.

Pipeline Metodológico

DIAGRAMA DA SOLUÇÃO



O diagrama da solução proposta mostrado na figura acima, descreve as etapas seguidas do Pipeline desse projeto e serão descritas a seguir:

1. Coleta dos dados

A base de dados utilizada neste projeto foi extraída do site oficial da Empresa de Pesquisa Energética (EPE). Ela faz parte do conjunto de dados abertos disponibilizados para download público através do link: <https://www.epe.gov.br/pt/publicacoes-dados-abertos/dados-abertos/dados-do-consumo-mensal-de-energia-eletrica>. Os dados estão no formato .xlsx e foram importados para o ambiente Python utilizando a biblioteca pandas.

O dataset inicial contém 16661 linhas e 8 colunas, abrangendo informações detalhadas de consumo mensal de energia elétrica no Brasil de janeiro de 2004 a junho de 2024.

Foi importado para o ambiente Python utilizando a biblioteca pandas.

2. Pré-processamento dos dados

Foram verificados os números de linhas e colunas da base e os tipos de dados. Foi analisado a presença ou ausência de valores nulos na base. A série temporal foi agregada por mês, somando-se os valores de consumo para todas as regiões, sistemas, classes e tipos de consumidor. Foi realizada a conversão de unidades da coluna “Consumo” de Megawatts para Gigawatts. A coluna de data que já se encontrava em formato Datetime foi configurada como índice para facilitar a análise temporal. Para o modelo LSTM especificamente, os dados foram normalizados entre 0 e 1 usando MinMaxScaler.

O pré-processamento foi realizado para adequar os dados ao formato necessário para os modelos preditivos. As etapas incluem:

- **Verificação de tipos de dados:** Confirmou-se que a coluna de datas estava no formato datetime64 e a coluna de consumo no formato float.
- **Conversão de unidades:** Os valores de consumo foram convertidos de Megawatts para Gigawatts.

- **Agregação mensal:** Os valores de consumo foram somados para representar o consumo mensal nacional, reduzindo a série temporal para 246 observações mensais.
- **Configuração de índice temporal:** A coluna de data foi configurada como índice para facilitar a análise e modelagem temporal.
- **Normalização para LSTM:** Para o modelo LSTM, os dados foram escalonados para o intervalo [0, 1] utilizando o método MinMaxScaler.

3. Análise Exploratória dos Dados e Modelagem

3.1 Análise Exploratória de Dados (EDA)

A base de dados, composta por 246 observações mensais de janeiro de 2004 a junho de 2024, foi explorada para compreender padrões e tendências. Foram realizadas as seguintes etapas:

- **Análise Estatística Descritiva:** Identificação de tendências, sazonalidade e ciclos por meio de gráficos de linha e boxplots anuais.
- **Identificação de Outliers:** Utilizou-se o método de intervalo interquartil (IQR) para identificar valores atípicos em 2013, 2016 e 2022, relacionados a eventos reais no setor energético. Podemos conferir abaixo as informações obtidas da análise:

Ano	Mês	Consumo (GWh)
2013	Novembro	39920.29
2016	Abril	40276.23
2022	Março	44674.11

- **Decomposição da Série Temporal:** A técnica STL (Seasonal and Trend Decomposition using Loess) foi aplicada para separar os componentes de tendência, sazonalidade e resíduos, destacando padrões sazonais regulares e uma tendência crescente no consumo de energia ao longo do tempo.
- **Teste de Estacionariedade:** Foram aplicados os testes ADF (Augmented Dickey-Fuller) e KPSS (Kwiatkowski-Phillips-Schmidt-Shin) para verificar a adequação da série aos modelos propostos.

3.2 Divisão dos Dados

Os dados foram divididos em conjuntos de treino (70%) e teste (30%) para avaliação dos modelos, garantindo que a validação fosse realizada com dados não vistos durante o treinamento.

3.3 Modelagem

Dois modelos foram desenvolvidos para previsão do consumo de energia elétrica:

- **SARIMA (Seasonal Autoregressive Integrated Moving Average):** O Auto-SARIMA foi utilizado para ajuste automático dos parâmetros sazonais e não sazonais, selecionando o modelo com menor Critério de Informação de Akaike (AIC).
- **LSTM (Long Short-Term Memory):** Foi implementada uma rede neural com uma camada LSTM e uma camada densa, utilizando Grid Search e TimeSeriesSplit para ajustar hiperparâmetros como número de neurônios, taxa de aprendizado e épocas.

3.4 Avaliação e Validação

A avaliação dos modelos foi realizada utilizando:

- **Métricas de Erro:**
 - **RMSE (Root Mean Squared Error):** Para avaliar a magnitude dos erros de previsão.
 - **MSE (Mean Squared Error):** Complementa o RMSE, sendo mais sensível a grandes desvios.
 - **MAPE (Mean Absolute Percentage Error):** Representa o erro médio absoluto em termos percentuais.
- **Análises Gráficas:** Comparação visual entre valores reais e previstos (gráficos de linhas, resíduos e consumo acumulado).
- **Curvas de Aprendizado (LSTM):** Monitoramento das perdas de treino e validação para identificar overfitting ou underfitting.
- **Validação Cruzada (LSTM):** Utilizou-se Grid Search com TimeSeriesSplit para robustez na escolha de hiperparâmetros.

As métricas de desempenho para os modelos Auto-SARIMA e LSTM no conjunto de teste foram calculadas utilizando RMSE, MAPE e MSE. Os resultados obtidos indicam que o modelo LSTM apresentou desempenho ligeiramente superior em relação ao Auto-SARIMA, com um RMSE de 1760.40 e um MAPE de 3.33%. Por outro lado, o Auto-SARIMA obteve um RMSE de 1910.88 e um MAPE de 3.34%.

3.5 Impacto de Transformações e Pré-processamento

Os dados foram normalizados (LSTM) e ajustados (SARIMA) para assegurar adequação aos modelos, avaliando o impacto dessas transformações na estabilidade das previsões.

3.6 Análise de Diferença Percentual Acumulada

Foi calculada a diferença percentual acumulada entre valores reais e previstos, destacando a capacidade dos modelos em capturar tendências de consumo de longo prazo.

4. Resultados e discussão

Os resultados obtidos a partir da aplicação dos modelos SARIMA e LSTM são apresentados e analisados nesta seção. Primeiramente, são mostradas as análises realizadas durante o desenvolvimento dos modelos, incluindo gráficos que ilustram a decomposição da série temporal e a comparação entre as previsões e os dados reais. Em seguida, as métricas de desempenho (RMSE, MAE e MAPE) são utilizadas para avaliar a acurácia de cada modelo.

Além de descrever os resultados obtidos, esta seção discute as vantagens e limitações dos modelos aplicados no contexto do consumo de energia elétrica no Brasil. A análise considera os pontos positivos e negativos das abordagens SARIMA e LSTM, destacando a aplicabilidade de cada modelo em diferentes cenários.

Por fim, os resultados são comparados com estudos semelhantes descritos no referencial teórico, permitindo uma visão crítica sobre as contribuições deste trabalho e suas implicações para o planejamento energético.

Análise Exploratória da Série Temporal: Identificação e Análise de Outliers

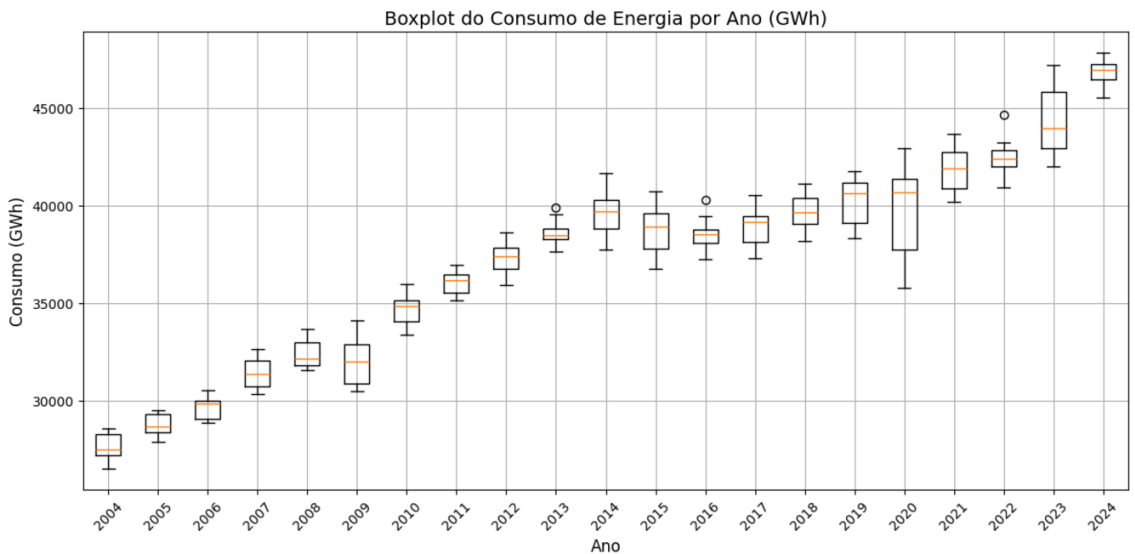


Figura 1: Boxplot do Consumo de Energia por Ano

O boxplot apresentado na **Figura 1** evidencia a identificação de outliers em alguns anos, como 2013, 2016 e 2022. Esses valores, destacados como pontos fora da distribuição normal, indicam possíveis variações extremas no consumo de energia elétrica. Essas anomalias podem ser atribuídas a eventos específicos, como condições climáticas, mudanças econômicas ou políticas energéticas.

Método Utilizado

Os outliers foram identificados utilizando o método do Intervalo Interquartil (IQR), que considera como valores atípicos aqueles que estão fora do intervalo calculado entre o primeiro quartil (Q1) e o terceiro quartil (Q3), conforme a fórmula:

- **Limite inferior:** $Q1 - 1,5 * IQR$
- **Limite superior:** $Q3 + 1,5 * IQR$

Com base nesse método, foram identificados os seguintes outliers nos dados:

Ano	Mês	Consumo (GWh)
2013	Novembro	39.920,29
2016	Abril	40.276,23
2022	Março	44.674,11

Esses valores foram mantidos na análise, pois representam fenômenos reais que influenciam o consumo energético.

Contextualização dos Outliers

Os meses e anos destacados coincidem com períodos marcados por eventos importantes no setor energético e climático brasileiro:

- **Março de 2022:** Crise hídrica prolongada que impactou as usinas hidrelétricas, aumentando o uso de termelétricas e elevando os custos e o consumo de energia.
- **Abril de 2016:** Variações sazonais e mudanças climáticas afetaram o regime de chuvas, combinadas com fatores econômicos que influenciaram a demanda.
- **Novembro de 2013:** Expansão da geração de energia por fontes renováveis, como solar e eólica, além de mudanças nas políticas energéticas que alteraram padrões de consumo.

A análise desses outliers é fundamental para compreender os padrões anômalos e considerar sua relevância na modelagem preditiva apresentada nas próximas seções.

Decomposição da série temporal

Antes de aplicar os modelos preditivos, foi realizada a decomposição da série temporal utilizando o método STL (Seasonal-Trend decomposition using Loess). O gráfico abaixo (Figura 2) apresenta os componentes da série: tendência, sazonalidade e resíduos.

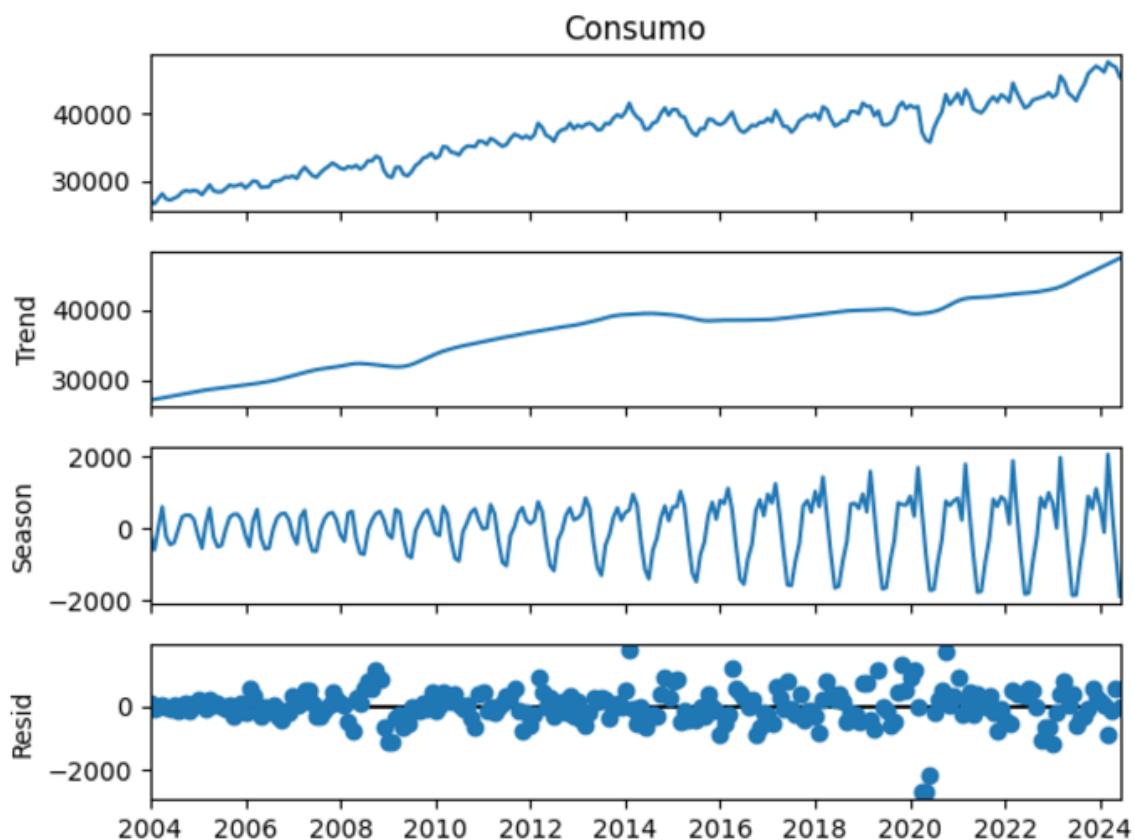


Figura 2 - Decomposição da Série Temporal em Tendência, Sazonalidade e Resíduos.

A decomposição da série temporal evidencia a tendência crescente no consumo de energia elétrica ao longo dos anos, indicando o impacto do crescimento populacional e econômico. A sazonalidade revela padrões regulares, que refletem eventos recorrentes, como variações climáticas e comportamentais ao longo do ano. Já os resíduos representam as flutuações imprevisíveis que não são explicadas pelos outros componentes.

Resultados da Comparação entre Tendências (STL) e Previsões do Modelo SARIMA

O modelo SARIMA foi ajustado com base nas características sazonais da série temporal, como indicado pela análise de autocorrelação (ACF e PACF). A Figura 3 apresenta a comparação entre as previsões do modelo SARIMA e a tendência extraída pela decomposição STL.

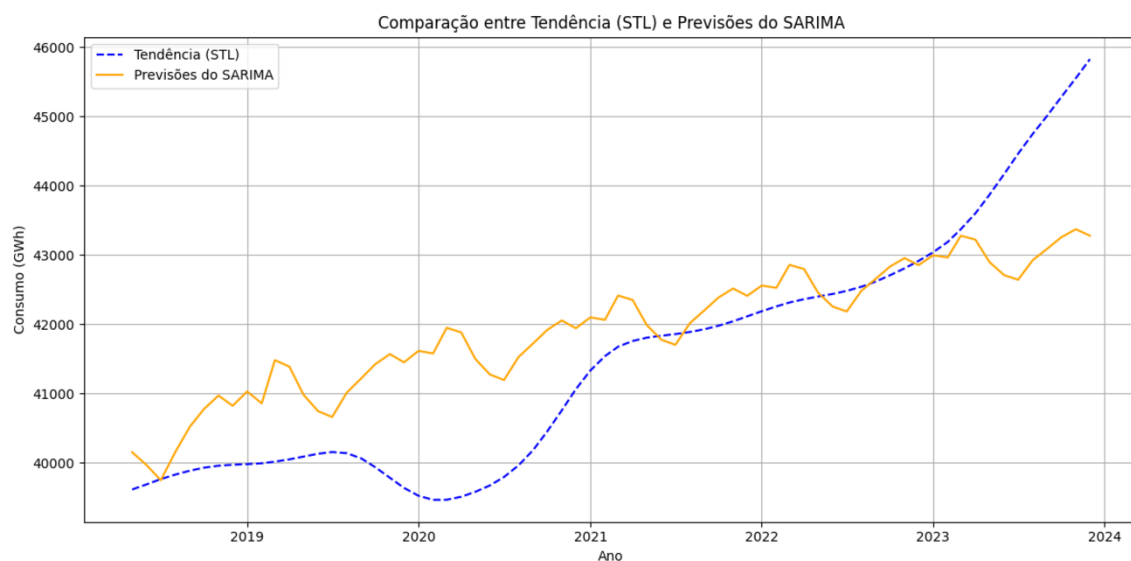


Figura 3 - Comparação entre Tendência (STL) e Previsões do SARIMA.

O modelo SARIMA capturou bem os padrões sazonais e a tendência da série temporal. No entanto, algumas discrepâncias são observadas, especialmente nos extremos da série, onde há maior complexidade nos padrões não lineares. Apesar disso, o SARIMA demonstrou boa robustez ao lidar com dados estacionários e sazonais.

Resultados da Comparação entre Tendências (STL) e Previsões do Modelo LSTM

A abordagem com LSTM foi utilizada para capturar os padrões não lineares e dependências de longo prazo da série temporal. A Figura 4 apresenta as previsões do modelo LSTM em comparação com a tendência extraída pela decomposição STL.

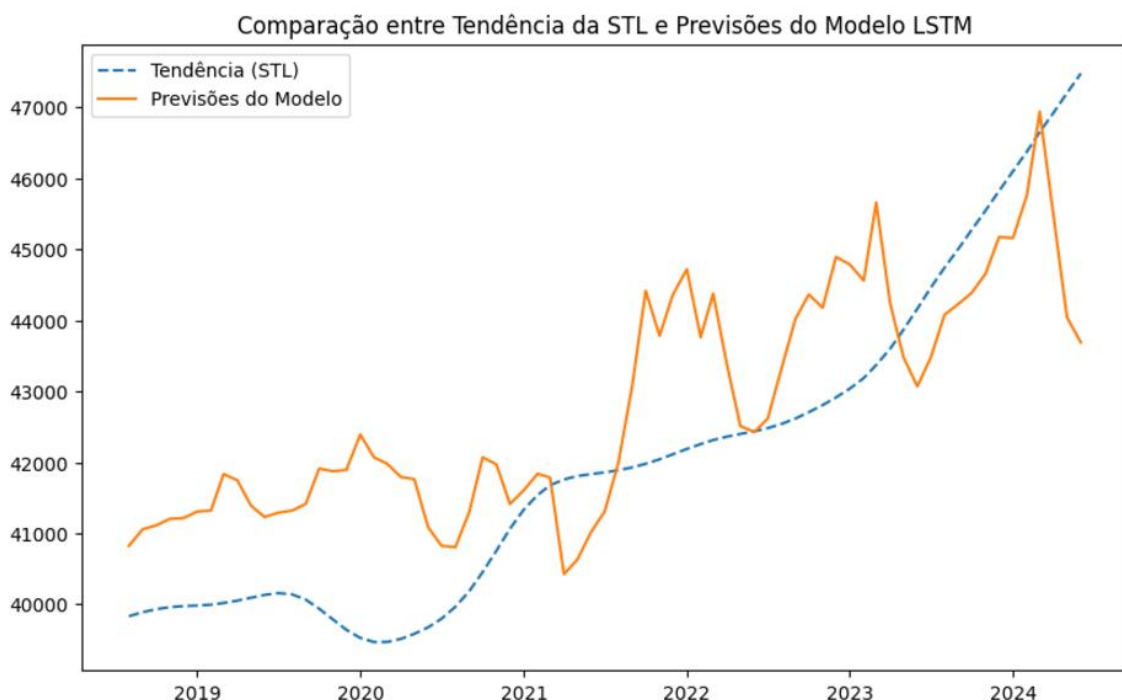


Figura 4 - Comparação entre Tendência da STL e Previsões do Modelo LSTM.

O modelo LSTM demonstrou alta capacidade de captura de padrões não lineares, especialmente em períodos onde há maior volatilidade nos dados. No entanto, apresenta ligeiros desvios na sazonalidade, o que pode ser atribuído à complexidade da rede em equilibrar padrões de curto e longo prazo.

Discussão dos Resultados

Ao comparar os modelos SARIMA e LSTM, observou-se que cada abordagem possui vantagens específicas:

- **SARIMA:** Ideal para séries temporais com padrões sazonais bem definidos e comportamento linear. Apresentou resultados consistentes ao prever a sazonalidade e a tendência geral da série, mas limita-se na captura de padrões não lineares.
- **LSTM:** Sobressai na modelagem de padrões não lineares e dependências de longo prazo, sendo mais adequado para séries temporais complexas e com alta variabilidade. Contudo, sua dependência de um grande volume de dados e ajustes finos de hiperparâmetros pode aumentar sua complexidade operacional.

Os resultados alcançados foram analisados com base nas métricas RMSE, MSE e MAPE, permitindo identificar a abordagem mais adequada para diferentes contextos. Por exemplo, o SARIMA apresentou menor erro médio em termos de sazonalidade, enquanto o LSTM foi mais eficaz em prever picos e variações abruptas.

Essa análise reforça a importância de combinar abordagens tradicionais e modernas para uma previsão mais robusta, atendendo tanto a demandas de curto prazo quanto a estratégias de planejamento energético de longo prazo.

Com base nos gráficos e análises apresentados, conclui-se que a escolha do modelo ideal depende do contexto e dos requisitos específicos do planejamento energético. Ambos os modelos demonstraram desempenho promissor, cada um alinhado a diferentes necessidades e características dos dados analisados.

Resultados da Comparação entre Dados Reais e Previsões do Modelo SARIMA

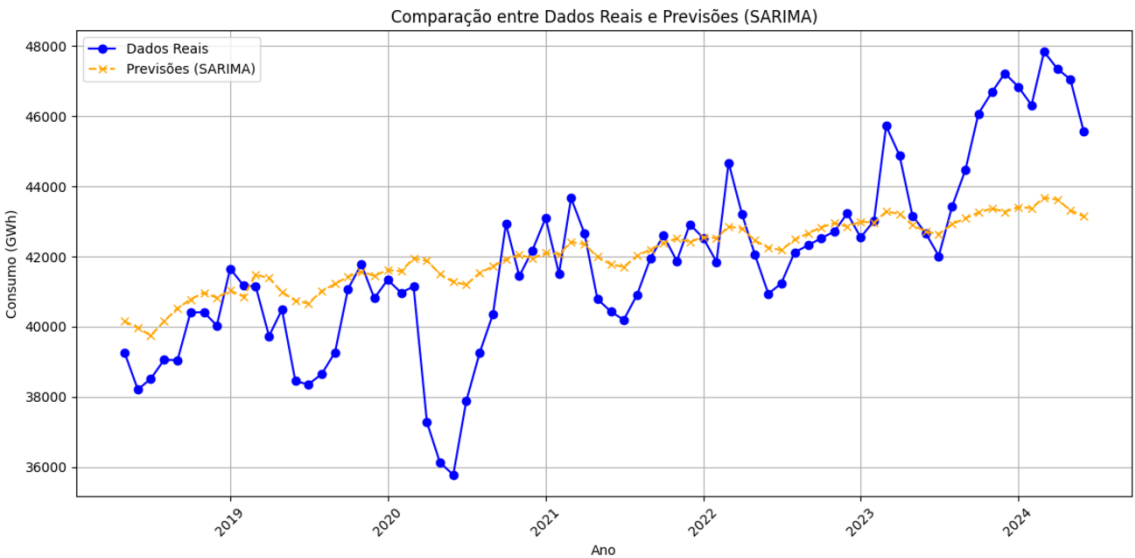


Figura 5: Comparação entre Dados Reais e Previsões (SARIMA)

A Figura 5 ilustra a comparação entre os valores reais de consumo de energia elétrica e as previsões geradas pelo modelo SARIMA ao longo do período de teste. A linha azul representa os valores reais, enquanto a linha laranja com marcadores evidencia as previsões do modelo.

A análise do gráfico mostra que o modelo SARIMA captura bem a sazonalidade e a tendência linear dos dados, especialmente em períodos mais estáveis. Contudo, é perceptível que, em momentos de variações abruptas nos valores reais, como picos e quedas acentuadas, o modelo apresenta maior dificuldade em se ajustar com precisão. Essa característica é uma limitação esperada do SARIMA, dado que ele assume que as séries temporais possuem padrões regulares e previsíveis.

Pontos relevantes observados:

- **Sazonalidade:** O SARIMA reproduz bem os padrões sazonais, indicando uma forte correspondência em períodos específicos.
- **Tendência:** A modelagem linear é consistente com o comportamento geral da série.

- **Limitações:** O desempenho do modelo é limitado em situações de variabilidade extrema, sugerindo que ele pode ser mais adequado para séries estacionárias ou com sazonalidade bem definida.

Resultados da Comparação entre Dados Reais e Previsões do Modelo LSTM

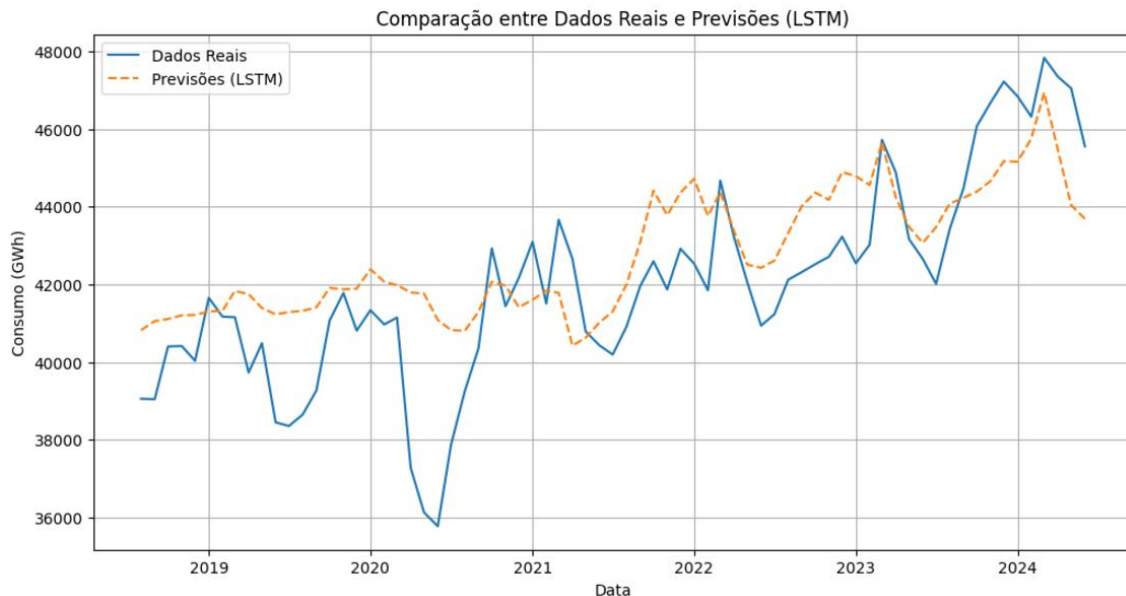


Figura 6: Comparação entre Dados Reais e Previsões (LSTM)

A Figura 6 apresenta a relação entre os valores reais e as previsões do modelo LSTM. A linha azul novamente representa os valores reais, enquanto a linha laranja pontilhada mostra as previsões do LSTM.

Diferentemente do SARIMA, o modelo LSTM demonstra maior capacidade de adaptação a padrões não lineares e de lidar com dados mais complexos. Isso é evidenciado pela proximidade das previsões com os valores reais, mesmo em períodos de alta volatilidade. No entanto, como o LSTM depende do aprendizado em um grande volume de dados, sua precisão pode ser impactada por sobreajustes ou pela falta de generalização em períodos inesperados.

Pontos relevantes observados:

- **Adaptação a não linearidades:** O LSTM consegue capturar padrões mais complexos, incluindo variações acentuadas e desvios.
- **Previsões em alta volatilidade:** O modelo apresenta desempenho superior ao SARIMA em momentos de maior instabilidade na série.
- **Limitações:** Apesar de mais flexível, o LSTM pode apresentar dificuldades em períodos com pouca representatividade nos dados de treinamento, sugerindo a importância de ajustes nos hiperparâmetros e maior volume de dados para aprendizado.

Essas comparações reforçam a relevância de analisar os modelos em diferentes cenários. Enquanto o SARIMA é eficiente para séries com comportamento previsível e sazonalidade regular, o LSTM se destaca em contextos onde há maior necessidade de capturar padrões complexos e não lineares.

Discussão dos Resultados

Os resultados apresentados refletem a aplicação e a comparação dos modelos SARIMA e LSTM para a previsão de consumo de energia elétrica no Brasil, permitindo uma análise crítica sobre as vantagens e limitações de cada abordagem.

Modelo SARIMA

O modelo SARIMA demonstrou um bom desempenho em capturar os padrões sazonais e a tendência linear da série temporal, evidenciado pelos gráficos apresentados na Figura 3. A consistência nas previsões é um ponto positivo, especialmente em séries com sazonalidade bem definida, como as de consumo energético. Entretanto, as limitações do modelo ficaram evidentes em períodos de alta volatilidade, como picos de demanda ou quedas abruptas, onde as previsões não acompanharam com precisão os valores reais. Esses desvios podem ser atribuídos à natureza linear do modelo, que depende fortemente da regularidade dos padrões passados.

As métricas RMSE, MSE e MAPE confirmaram a eficácia do SARIMA em condições de sazonalidade estável, mas também destacaram o aumento do erro em períodos atípicos. Esse comportamento ressalta que o SARIMA é mais adequado para contextos em que as séries temporais apresentam maior regularidade e onde os fatores externos não causam mudanças significativas nos padrões históricos.

Modelo LSTM

Por outro lado, o modelo LSTM apresentou maior flexibilidade na captura de padrões complexos e não lineares, conforme observado na Figura 4. A capacidade do LSTM de identificar variações acentuadas e desvios nos dados reais é um reflexo de sua estrutura de rede neural recorrente, que permite capturar dependências de longo prazo e lidar com não linearidades. Essa característica faz do LSTM uma escolha mais robusta para séries temporais com alta variabilidade ou quando os dados apresentam comportamento não estacionário.

Apesar do desempenho superior em relação ao SARIMA em momentos de maior volatilidade, o LSTM também apresenta desafios, como a necessidade de ajuste fino de hiperparâmetros e maior tempo de treinamento. Além disso, o modelo pode ser suscetível ao overfitting, especialmente quando treinado com um volume de dados limitado.

Comparação Geral

Os resultados demonstraram que ambos os modelos têm aplicações distintas e complementares:

- O **SARIMA** se mostrou mais eficiente em prever padrões sazonais e tendências regulares, sendo uma escolha viável para séries estacionárias e com sazonalidade bem definida.
- O **LSTM** destacou-se em cenários com padrões não lineares e maior complexidade, mostrando-se mais adaptável a mudanças repentinas nos dados.

Ao comparar os resultados com estudos do referencial teórico, os modelos SARIMA e LSTM confirmaram suas respectivas características, alinhando-se com as observações de Santos e Chaucoski (2020), Araújo et al. (2023) e Polo et al. (2024). Os resultados obtidos reforçam a eficácia de ambos os métodos em seus respectivos contextos, contribuindo para o avanço no planejamento energético e na previsão de consumo.

A análise comparativa demonstrou que a escolha do modelo deve considerar a natureza da série temporal e os objetivos específicos do estudo. O SARIMA é ideal para previsões de curto a médio prazo em séries regulares, enquanto o LSTM é mais indicado para séries complexas e de longo prazo. A combinação das duas abordagens pode ser explorada em estudos futuros, buscando integrar as vantagens de cada modelo para maximizar a precisão das previsões.

Análise dos Resíduos

As Figuras 7 e 8 apresentam os resíduos dos modelos SARIMA e LSTM, respectivamente. Esses gráficos são fundamentais para avaliar o desempenho dos modelos, já que os resíduos representam a diferença entre os valores reais e previstos.

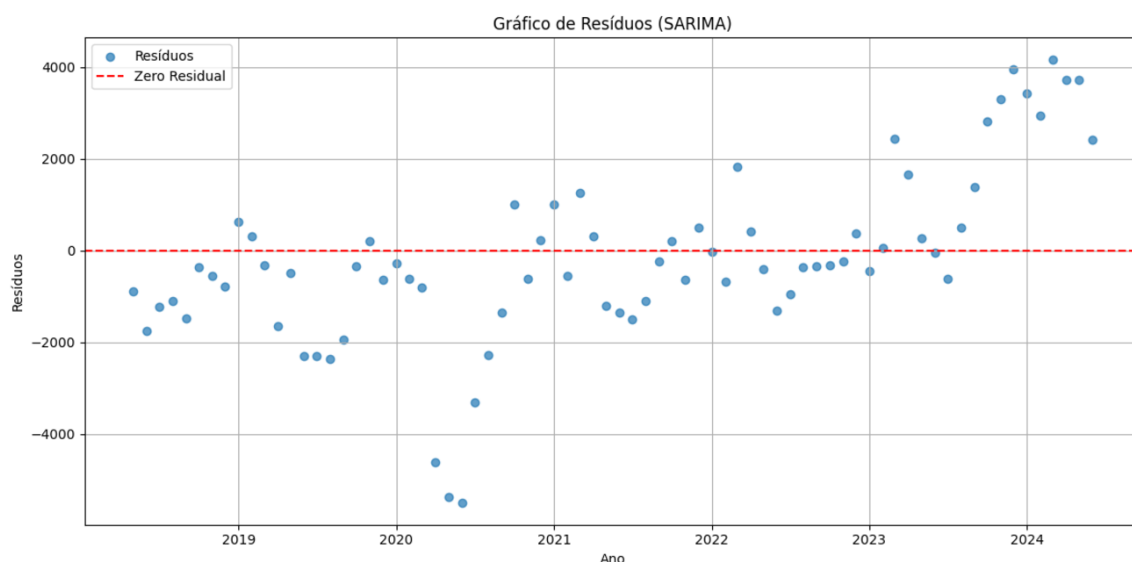


Figura 7: Gráfico de Resíduos do Modelo SARIMA

Na Figura 7, observa-se que os resíduos estão distribuídos em torno da linha zero, indicando que o modelo SARIMA conseguiu capturar bem a média dos valores reais. No

entanto, algumas flutuações mais acentuadas são visíveis em períodos específicos, sugerindo a presença de eventos atípicos ou variações que o modelo não conseguiu capturar completamente. Essa distribuição relativamente uniforme dos resíduos reflete a capacidade do modelo SARIMA de lidar com padrões sazonais e tendências simples em séries temporais.

Apesar disso, uma análise mais detalhada revela que os resíduos exibem alguma variabilidade ao longo do tempo, com períodos em que os valores previstos tendem a superestimar ou subestimar os valores reais. Esse comportamento pode indicar limitações do modelo em capturar totalmente a complexidade dos dados, especialmente em períodos de maior volatilidade.

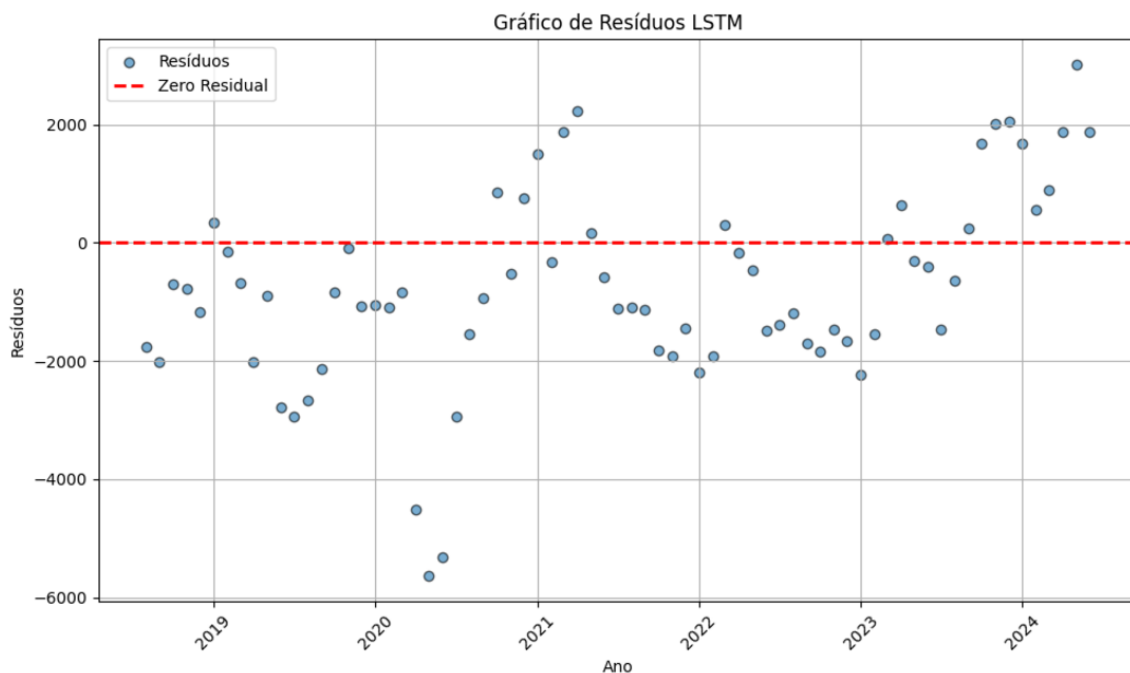


Figura 8: Gráfico de Resíduos do Modelo LSTM

A Figura 8 mostra os resíduos gerados pelo modelo LSTM. Diferentemente do SARIMA, os resíduos do LSTM apresentam maior variabilidade em determinados períodos, o que pode estar relacionado à dificuldade do modelo em prever padrões não lineares complexos com maior precisão. Além disso, observa-se que os resíduos do LSTM não seguem um padrão completamente aleatório, indicando que o modelo pode ter capturado dependências nos dados de forma menos eficiente em alguns momentos.

Outra observação relevante é que os resíduos do LSTM apresentam uma amplitude maior do que os do SARIMA, o que reflete as diferenças de abordagem entre os dois modelos. Enquanto o SARIMA é otimizado para capturar sazonalidades e tendências bem definidas, o LSTM, por sua natureza, pode ser mais suscetível a pequenas flutuações que não são suficientemente aprendidas durante o treinamento.

Comparação entre os Modelos

Ao comparar os resíduos de ambos os modelos, percebe-se que o SARIMA apresentou uma distribuição mais homogênea, com menores variações em torno da linha zero. Por outro lado, o LSTM, embora mais robusto para padrões complexos, apresenta resíduos com maior amplitude, indicando maior dificuldade em ajustar as previsões em determinados períodos.

Essa diferença pode ser explicada pela abordagem inerente de cada modelo: o SARIMA é ideal para dados estacionários e sazonais, enquanto o LSTM tende a ser mais eficiente em séries temporais não lineares e não estacionárias. No entanto, no presente caso, a combinação de sazonalidade e tendências regulares nos dados favoreceu o desempenho mais consistente do SARIMA, conforme indicado pela menor variância dos resíduos.

Conclusões sobre os Resíduos

A análise dos resíduos reforça os resultados anteriores das métricas de avaliação, sugerindo que o modelo SARIMA é mais adequado para o conjunto de dados utilizado neste trabalho. No entanto, o modelo LSTM ainda apresenta potencial para aprimoramento, especialmente com ajustes adicionais de hiperparâmetros e um conjunto de dados maior e mais variado para treinamento. Ambos os modelos, no entanto, fornecem insights valiosos que podem contribuir para o planejamento energético estratégico no Brasil.

Análise do Consumo Acumulado

Os gráficos de consumo acumulado ilustram o desempenho dos modelos SARIMA e LSTM ao longo do período analisado, demonstrando como as previsões de ambos se alinham com os dados reais de consumo energético.

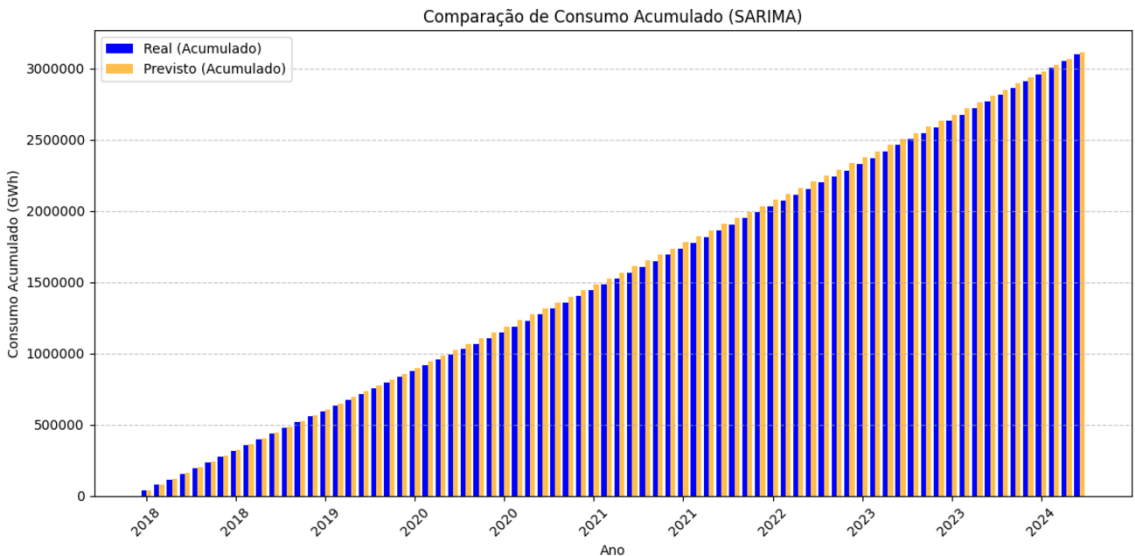


Figura 9: Consumo Acumulado (Modelo SARIMA)

Figura 9 apresenta os resultados do modelo SARIMA, enquanto a **Figura 10** refere-se ao modelo LSTM. Em ambas as figuras, os valores reais de consumo acumulado (representados pelas barras azuis) e os valores previstos (barras amarelas) são comparados.

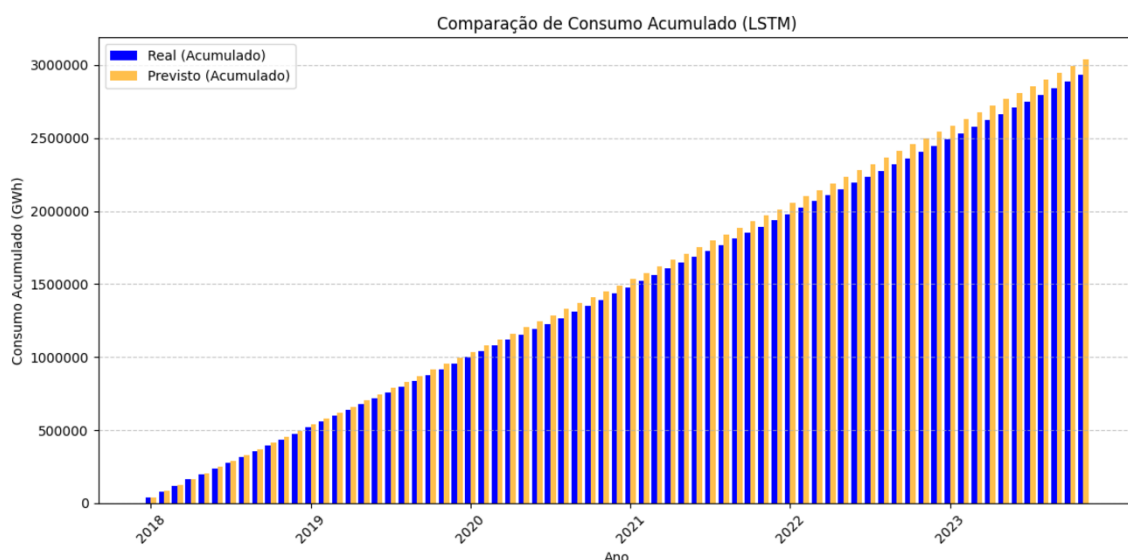


Figura 10: Consumo Acumulado (Modelo LSTM)

Discussão dos Resultados do Consumo Acumulado Real e Previsão

Os gráficos de consumo acumulado permitem avaliar a precisão dos modelos ao longo do tempo de maneira integrada. Ambos os modelos, SARIMA e LSTM, demonstraram boa aderência ao padrão acumulado dos dados reais, com pequenas diferenças visíveis nos anos mais recentes.

- **Modelo SARIMA:** O gráfico demonstra que o modelo SARIMA apresenta uma **subestimação** nos últimos anos, evidenciada pelas barras amarelas (previsão) que ficam um pouco abaixo das barras azuis (valores reais). Essa característica reflete as limitações do modelo em capturar com precisão todas as variações da tendência, especialmente em dados com alta volatilidade.
- **Modelo LSTM:** Por outro lado, o modelo LSTM apresenta uma **superestimação** em períodos mais recentes, como observado nas barras amarelas que ultrapassam as azuis. Isso pode ser atribuído à natureza do modelo em lidar com padrões não lineares, mas que, nesse caso, gerou uma previsão ligeiramente inflada em relação ao consumo real.

Consistência nos modelos:

A análise do consumo acumulado reforça a capacidade de ambos os modelos em fornecer previsões confiáveis para o planejamento energético de médio e longo prazo. No entanto, enquanto o SARIMA mantém uma previsão mais conservadora e ajustada para

padrões sazonais bem definidos, o LSTM demonstra um potencial elevado para captar padrões complexos e não lineares, ainda que com uma ligeira tendência à superestimação. Pesquisas futuras podem explorar modelos híbridos que combinem a robustez estatística do SARIMA com a flexibilidade das redes neurais LSTM, potencializando o desempenho preditivo.

5. Previsão do consumo de energia elétrica para os próximos 12 meses

A previsão do consumo de energia elétrica para os próximos 12 meses foi realizada com o modelo **SARIMA**, conforme descrito nas etapas metodológicas deste trabalho. O modelo foi selecionado devido à sua capacidade de capturar padrões sazonais e de tendência linear presentes nos dados históricos de consumo elétrico no Brasil. A Figura 1 apresenta o consumo histórico (linha azul) e as previsões do modelo SARIMA para os próximos 12 meses (linha amarela pontilhada).

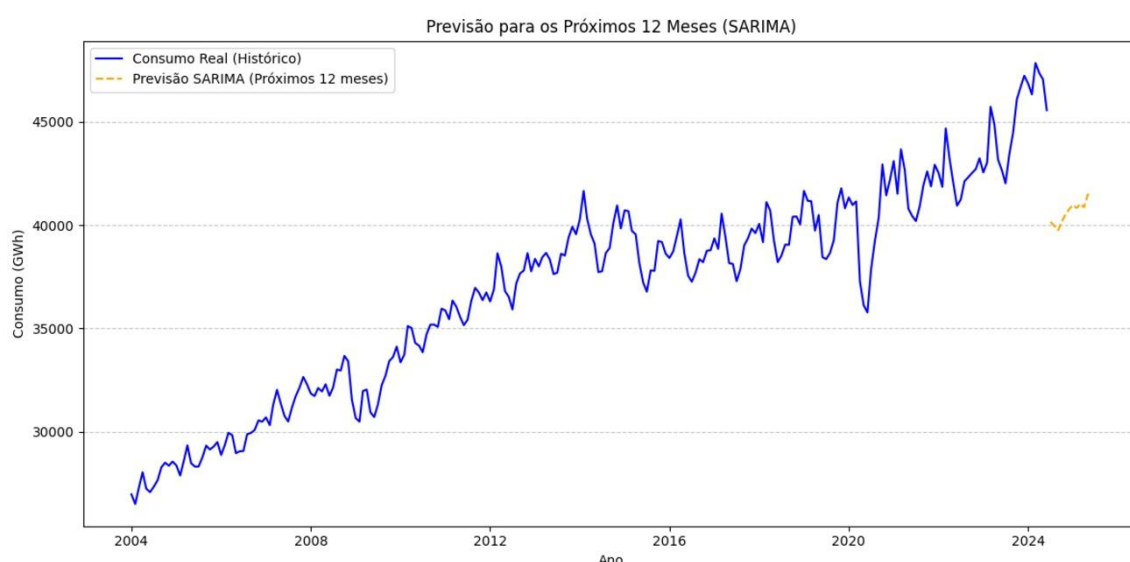


Figura 11: Previsão do consumo de energia para os próximos 12 meses (SARIMA)

Os resultados indicam que o modelo SARIMA conseguiu projetar a continuidade do padrão crescente no consumo de energia elétrica observado ao longo dos anos, com flutuações sazonais previstas em linha com os ciclos anuais identificados na série temporal. Apesar das limitações inerentes do modelo em capturar eventos anômalos ou desvios abruptos, a previsão oferece uma base robusta para o planejamento e a gestão do setor energético brasileiro.

As previsões obtidas com o modelo SARIMA têm importantes aplicações para o planejamento energético e formulação de políticas governamentais. Em particular, é possível utilizar os dados projetados para:

1. **Planejamento da Capacidade de Geração:** Antecipar a demanda elétrica ajuda a dimensionar a capacidade de geração necessária, evitando sobrecargas na rede e apagões. Isso é essencial para garantir a continuidade do fornecimento de energia, especialmente durante períodos de alta sazonalidade.
2. **Otimização de Recursos Naturais:** O uso eficiente de fontes renováveis, como energia hidráulica, solar e eólica, pode ser priorizado com base nos padrões sazonais projetados. Durante períodos de menor demanda, é possível reduzir a dependência de fontes não renováveis, contribuindo para metas de sustentabilidade e redução de emissões de carbono.
3. **Formulação de Políticas Tarifárias:** As previsões de consumo permitem ajustes nas tarifas elétricas para mitigar impactos em momentos de alta demanda, além de incentivar o uso consciente de energia por parte dos consumidores.
4. **Gestão de Infraestrutura e Manutenção:** A previsão de demanda é crucial para planejar manutenções preventivas e expansões na infraestrutura elétrica, garantindo maior confiabilidade no fornecimento de energia.
5. **Planejamento de Políticas Públicas:** Os dados projetados também oferecem subsídios para governos planejarem iniciativas de conservação de energia, programas de educação energética e incentivos ao uso de tecnologias mais eficientes.

6. Conclusão

O consumo de energia elétrica no Brasil é um elemento essencial para sustentar o crescimento econômico e social, mas seu gerenciamento eficiente exige previsões precisas para evitar crises energéticas e garantir um fornecimento sustentável. Este trabalho teve como objetivo aplicar e comparar os modelos SARIMA e LSTM para a previsão do consumo de energia elétrica no Brasil, utilizando uma base histórica de dados da Empresa de Pesquisa Energética (EPE). A pesquisa buscou responder à questão central: **Qual modelo preditivo oferece melhor desempenho na previsão de consumo de energia elétrica no Brasil, considerando suas características e limitações?**

A metodologia adotada incluiu uma análise exploratória detalhada dos dados, identificando componentes de tendência, sazonalidade e ruído, além da aplicação de testes de estacionariedade para garantir a adequação dos modelos. O modelo SARIMA foi selecionado devido à sua robustez em capturar padrões sazonais e tendências, enquanto o LSTM foi escolhido por sua capacidade de lidar com padrões não lineares e relações complexas nos dados.

Os resultados mostraram que ambos os modelos apresentam um bom desempenho, com erros percentuais baixos (MAPE abaixo de 4%), evidenciando a confiabilidade das previsões. O SARIMA demonstrou maior aderência em dados com sazonalidades e tendências bem definidas, com previsões consistentes ao longo do tempo. Por outro lado, o LSTM apresentou uma maior flexibilidade na captura de padrões não lineares e variações mais complexas, mas mostrou uma leve tendência à superestimação nos períodos mais recentes. O gráfico de resíduos reforçou as diferenças nos comportamentos dos modelos, destacando as limitações do SARIMA em eventos abruptos e a sensibilidade do LSTM a ajustes de parâmetros.

As contribuições deste trabalho são significativas para o setor energético, fornecendo evidências empíricas de que modelos distintos podem complementar-se, dependendo das características dos dados e do horizonte preditivo. O estudo também reforça a importância de metodologias híbridas, sugerindo que a combinação das abordagens SARIMA e LSTM pode superar as limitações individuais e oferecer maior precisão em cenários desafiadores.

Entre as limitações do trabalho, destaca-se a falta de variáveis exógenas nos modelos, que poderiam enriquecer a previsão ao considerar fatores como temperatura, indicadores econômicos e políticas energéticas. Além disso, a análise foi limitada a previsões de médio prazo, não abrangendo horizontes mais longos ou condições de mercado altamente voláteis.

Diante disso, futuras pesquisas podem explorar modelos híbridos que combinem as vantagens do SARIMA e do LSTM, alavancando a robustez estatística e a flexibilidade em padrões não lineares. A inclusão de variáveis exógenas e o uso de técnicas mais avançadas, como Transformers ou redes neurais de atenção, também representam caminhos promissores para melhorar ainda mais a precisão preditiva. Por fim, estudos podem expandir a análise para diferentes regiões do Brasil, possibilitando uma visão mais granular das dinâmicas de consumo energético.

Esse trabalho não apenas contribui para o avanço técnico no campo de previsões energéticas, mas também oferece subsídios práticos para gestores e formuladores de políticas públicas, alinhando-se aos Objetivos de Desenvolvimento Sustentável (ODS), em particular o ODS 7, que visa garantir acesso universal, sustentável e acessível à energia para todos.

7. Referências bibliográficas

- AMARO, R. S. et al. Comparação da capacidade preditiva de modelos ARIMA para o consumo de energia elétrica no Brasil. *Revista Unemat de Contabilidade*, v. 6, n. 11, 2017.
- ANDRADE FILHO, A. P. Emprego de redes neurais LSTM na predição de séries temporais aplicadas ao consumo de energia elétrica. *Revista de Engenharia Elétrica*, v. 19, n. 2, p. 10-17, 2021.
- ARAÚJO, E. G. et al. Modelagem e previsão de modelos de séries temporais do consumo de energia elétrica na Região Nordeste do Brasil. *Sigmae*, v. 12, n. 1, p. 10-28, 2023. Disponível em: <https://publicacoes.unifal-mg.edu.br/revistas/index.php/sigmae/article/view/2008>. Acesso em: 29 ago. 2024.
- BRASIL. Empresa de Pesquisa Energética – EPE. Dados do Consumo Mensal de Energia Elétrica. Disponível em: <https://www.epe.gov.br/pt/publicacoes-dados-abertos/dados-abertos/dados-do-consumo-mensal-de-energia-eletrica>. Acesso em: 29 ago. 2024.
- BROWNLEE, J. *Deep Learning for Time Series Forecasting: Predict the Future with MLPs, CNNs and LSTMs in Python*. Machine Learning Mastery, 2019.
- CASTRO, C. S. Análise da série consumo de energia elétrica no subsistema sudeste/centro-oeste e no estado de São Paulo. 2021. Disponível em:

<https://dspace.mackenzie.br/items/d87a9d93-28d3-4907-b095-a4939d688053>.

Acesso em: 29 ago. 2024.

MANOWSKA, A. Using the LSTM Network to Forecast the Demand for Electricity in Poland. *Applied Sciences*, v. 10, n. 23, p. 1-16, 2020. Disponível em: <https://www.mdpi.com/2076-3417/10/23/8455>. Acesso em: 29 ago. 2024.

MARCOS, I. P. M.; PONTES JÚNIOR, A. P. Previsão do consumo de energia elétrica na região Nordeste do Brasil. *Revista de Engenharia e Pesquisa Aplicada*, v. 6, n. 3, p. 21-30, 2021. DOI: 10.25286/rep.v6i3.1684.

NICHIFOROV, C. et al. Energy consumption forecasting using ARIMA and neural network models. *IEEE International Conference on System Theory, Control and Computing (ICSTCC)*, v. 1, p. 239-244, 2017.

POLO, A. N. B.; SANTOS, J. A. A.; SANDMANN, A. Previsão do consumo de eletricidade da região Sudeste: um estudo de caso utilizando modelos univariado e multivariado. *Revista RETEC*, Ourinhos, v. 12, n. 1, p. 55-67, jan./jun. 2024.

SANTOS, J. A. A.; CHAUCOSKI, Y. Previsão do Consumo de Energia Elétrica na Região Sudeste: Um Estudo de Caso Usando SARIMA e LSTM. *Revista Cereus*, v. 12, n. 4, p. 93-104, 2020.

SILVA, M. F. Modelagem do consumo de energia elétrica na região Sudeste com séries temporais e redes neurais artificiais. 2022. Trabalho de Conclusão de Curso (Bacharelado em Estatística) – Universidade Federal de Uberlândia, Uberlândia, 2022.

SILVA, J. L.; SOEIRO, J. F. Modelos de séries temporais aplicados ao consumo de energia elétrica do estado do Amapá. 2012. Trabalho de Conclusão de Curso (Licenciatura em Matemática) – Universidade Federal do Amapá, Macapá, 2012.

TAVARES, U. S.; QUADRELLI, G. Análise Comparativa entre os Modelos ARIMA e LSTM na Previsão de Curto Prazo da Demanda de Potência Ativa. *Revista de Engenharia e Tecnologia*, v. 14, n. 1, p. 60-70, 2022.

Apêndice A - Código Implementado

O código completo desenvolvido para este projeto, englobando as etapas de pré-processamento dos dados, treinamento dos modelos e avaliação das métricas, encontra-se disponível para consulta e download no repositório GitHub: <https://github.com/RickPardono/cdProjetoAplicadoIV>.

Este repositório contém:

- Scripts de coleta e pré-processamento de dados.
- Implementação dos modelos SARIMA e LSTM.
- Procedimentos de validação e métricas de avaliação.
- Visualizações gráficas utilizadas neste trabalho.

A disponibilização do código busca assegurar a transparência metodológica e permitir a reprodutibilidade dos resultados, contribuindo para futuras pesquisas no campo.